Random Sets: Unification and Computation for Information Fusion—A Retrospective Assessment

Ronald Mahler
Lockheed Martin MS2 Tactical Systems
3333 Pilot Knob Road
Eagan MN 55121
USA
ronald.p.mahler@lmco.com

Abstract – *The author introduced the key ideas of finite-set statistics (FISST) between 1994 and 1996, as a means of theoretically unifying the major aspects of information fusion under a single probabilistic umbrella. FISST has been considerably extended and refined since that time, especially in regard to the development of principled approximation strategies. For the last several years FISST has been and is being applied to a number of practical applied-research problems. Several research teams around the world are currently investigating FISST techniques. The purpose of this keynote paper is to provide a retrospective assessment of the last decade of random set information fusion research: its antecedents; its techniques, tools, conceptual evolution, and current state of the art; its applications; its critics and imitators; and its possible future directions.*

Keywords: random sets, point processes, information fusion, expert systems, nonlinear filtering, Bayesian.

1 Introduction

The tenth anniversary of the introduction of finite-set statistics (FISST) is an opportune time for a retrospective assessment. In a recent invited article in the *IEEE Aerospace and Electronics Systems Magazine* [56], we described FISST at an elementary, tutorial level. This invited keynote paper builds on that paper by trying to provide a "road map" through FISST-related papers and techniques. The following points will be central to the discussion that follows:

- *The fact that multisensor-multitarget problems can be formulated using random sets—or any other formalism—is, in and of itself, of limited engineering interest.*
- *If a formulation requires increased mathematical complexity, this should result in increased engineering advantage, especially computationally.*
- *Conversely, if a formulation is so heedlessly or vaguely formulated that it repeatedly leads to engineering blunders, it is simplistic—not simple.*
- *The point of approximation is to tractably preserve as much application realism as possible—not to achieve tractability under simplistic assumptions.*

Given this, we will provide a comprehensive, high-level assessment of FISST: of its antecedents; its beginnings and motivations; its techniques, tools, conceptual evolution, and current state of the art; its applications; its critics and imitators; and its possible future directions.

In the remainder of this Introduction we summarize these issues at a broad-brush level, deferring greater detail until later sections. We describe the major motivations underlying FISST in section 1.1; the three phases of its conceptual evolution—foundations, refinement, and computation—in sections 1.2, 1.3 and 1.4, respectively; its applications in section 1.5; its imitator-critics in section 1.6; and the organization of the paper in section 1.7.

1.1 Motivations

Three primary obstacles have inhibited the maturation of general information fusion as a unified, systematic, scientifically founded engineering discipline.

- The first is *the highly disparate and ambiguous forms that information can have*. Many kinds of data, such as that supplied by tracking radars, can be described in statistical form. On the other hand, statistically uncharacterizable real-world variations make other kinds of data (e.g., synthetic aperture radar (SAR) images) difficult to model. It has been even more unclear how still other forms of data—natural-language statements, features extracted from signatures, rules drawn from knowledge-bases—might be mathematically modeled and processed. Numerous expert systems approaches—fuzzy set theory, the Dempster-Shafer theory, etc.—have been proposed to address such problems. But their sheer number and variety—not to mention their perceived lack of defensible probabilistic foundations—has probably generated more confusion than enlightenment.
- The second obstacle is the fact that multisource-multitarget systems introduce a major complication. Such systems are comprised of *randomly varying numbers of randomly varying objects of various kinds*: randomly varying collections of targets, randomly varying collections of sensors and sensor-carrying platforms, and randomly varying observation-scans collected by those sensors. A rigorous mathematical foundation for stochastic multi-object problems—*point process theory* [9, 32, 83, 86]—has been in existence for decades. However, this theory has traditionally been formulated

with the requirements of mathematicians rather than engineers in mind.

• The third obstacle is the most crucial of all. The daunting combinatorial complexity of multisensor-multitarget systems (see section 3.1) guarantees that no systematic unification of information fusion will be useful unless it incorporates an equally systematic calculus for devising principled algorithmic approximation strategies.

1.2 1994-1996: Foundations

FISST was developed to address such issues—see sections 3 and 4. Beginning in the late 1970s, many researchers realized that random set theory might provide a way out of the increasing fragmentation and confusion in expert systems theory (see section 2.2). Inspired by the pioneering work of I.R. Goodman and H.T. Nguyen [19], in 1994 we proposed random set theory as a theoretically rigorous way of integrating point process theory with expert systems theory—thereby producing a unified and purely probabilistic foundation for much of multisensor-multitarget information fusion [59]. Most importantly, we noted that this synthesis could be formulated as a generalization of the "Statistics 101" formalism that most signal processing engineers learn as undergraduates. This initial three years of research was codified in Chapters 2, 4-8 of the 1997 book *Mathematics of Data Fusion* [20].

• The following familiar single-sensor, single-target concepts and techniques can be directly generalized—though not necessarily in a cookbook fashion—to the multisensor-multitarget realm: probability-mass functions, motion and measurement models, density functions, integrals, derivatives, Markov densities and likelihood functions, optimal state estimators and filters, Kullback-Leibler discrimination and information theory, etc. Thus multisensor-multitarget problems can be addressed using the same formal Bayesian modeling methods accepted as standard in single-sensor, single-target problems (see [43] and sections 3.4, 3.5, and 4.3).

The mathematical core of this aspect of FISST is a novel distillation and reformulation of those aspects of point process theory most pertinent to information fusion. The core of this reformulation consists of the following concepts: belief-mass functions (belief measures) and their set derivatives; and multi-object density functions and their set integrals (see section 4.2).

• We proposed (Chapter 7 of [20]) that formal modeling methods could be extended to encompass fuzzy logic, Dempster-Shafer theory, and rule-based evidence. They could thereby also encompass evidence that is ambiguous either in itself or because of uncertainty in its generation (see section 4.4). In particular, unification of the multi-object and the expert-systems aspects of information fusion results in part because the belief functions of Dempster-Shafer theory are special cases of belief-mass functions (the probability laws of multi-object systems).

1.3 1997-1999: Refinement

The work reported in [20] provided a solid theoretical basis for multisensor-multitarget problems, along with a proposed foundation for ambiguous evidence. The following three years of research were devoted primarily to devising a more systematic and practical approach for dealing with ambiguous evidence. We also fleshed out the engineering details of the formal Bayes modeling methodology proposed in 1994-1996. Most of this research was summarized in the 2000 monograph, *An Introduction to Multisource-Multitarget Statistics and Its Applications* [45]; and in Chapter 14 of the book *Handbook of Multisensor Data Fusion* [52].

• In Chapter 7 of [20] we proposed two basic concepts: (1) modeling ambiguous evidence as random closed subsets of some measurement space; and (2) likelihood functions for ambiguous evidence, as well as three specific approaches for defining them. These are special cases of what we now call "generalized likelihood functions." We showed that generalized likelihood functions can be formulated as ordinary conditional probabilities and used to good advantage in generalized Bayes recursive filters (see section 4.4).

• We addressed a type of evidence that we had previously left unexplored—data (such as SAR signatures) that is itself unambiguous but which, because of uncharacterizable real-world statistical variability, has imprecisely specified likelihoods (see section 4.4).

• Once this had all been integrated, what resulted was a generalization of formal Bayesian statistical modeling to both multisensor-multitarget problems and to ambiguous forms of evidence (see section 4.5 and [45, 58]).

1.4 2000-2004: Computation

FISST is in part a systems-level generalization of formal Bayesian modeling procedures (see [43, 56], and section 3.4) to the multisource-multitarget realm (section 3.5). As with any top-down approach it generates an obvious challenge: *computability*. The combinatorial complexity of multisource-multitarget problems guarantees that system-level solutions will be impractical without drastic but principled approximation techniques. One of our emphases since 1999 has been to generalize familiar single-sensor, single-target approximation methods to the multisource-multitarget realm.

• We began by devising a multitarget analog of the Gaussian approximation (see [49] and pp. 49-52 of [45]). If multitarget likelihoods, multitarget Markov densities, and multitarget prior densities are assumed to have a "para-Gaussian" form, then we showed that the multitarget Bayes filter is recursively closed in the sense that all multitarget posterior distributions must also be approximately para-Gaussian. This potentially allowed multitarget posteriors to be described by a relatively small number of parameters, which can be propagated instead of the actual multitarget posteriors. However, this approach also had two significant limitations: it provided no means for modeling target appearance; and it is algebraically complex for more than a few targets.

• More recently, and using a different analogy, we have devised a multitarget statistical generalization of the concept of a first-order moment filter (i.e., a constant-gain Kalman filter). The resulting approximate multitarget filter, called the *probability hypothesis density* (PHD) filter, has attracted a great deal of interest since its introduction in 2000 (see section 5.3).

- Our PHD filter research was originally based on brute-force methods of proof. However, it soon led us to extend our original core mathematical approach. In ordinary statistics probability distributions must sometimes be augmented by concepts such as characteristic function, moment generating function, etc. Likewise, we have found it useful to introduce *probability generating functionals* (p.g.fl.'s) and their *functional derivatives* (section 4.6), which are natural generalizations of belief-mass functions and set derivatives, respectively. Under certain circumstances, the multitarget Bayes filter (section 4.5) can be simplified when rewritten in terms of p.g.fl.'s. This fact has produced a systematic approach for constructing principled approximations (section 4.7) that are applicable to multitarget detection and tracking (section 5.3); group-target detection and tracking (section 5.4); and sensor management (section 5.5).

1.5 Applications

Since its introduction in 1994 FISST has subsumed an increasingly comprehensive swath of information fusion that includes multisource-multisensor integration (Level 1 fusion); expert systems theory; sensor management for Level 1 fusion, including management of dispersed mobile sensors; group-target detection, tracking, and classification; robust automatic target recognition; and scientific performance evaluation. FISST-based algorithms are being or have been investigated under basic and applied R&D contracts from U.S. Department of Defense agencies such as the Army Research Office, the Air Force Office of Scientific Research, SPAWAR Systems Center, the Missile Defense Agency, the Army Missile Research and Defense Command, and three sites of the Air Force Research Laboratory. Several FISST-based research efforts in Australia, Canada, Sweden, and the United States are currently in progress.

Selected applications will be summarized in section 5.

1.6 Imitator-critics of FISST

FISST has attracted much international interest in a relatively short time. The lesson that some seem to have drawn is that the royal road to recognition is impressive-looking mathematical complexity—specifically, more complicated point process formulations supposedly more general than FISST. Others have drawn the opposite conclusion: that renown will follow if one skims a few insights from FISST; strips off the mathematics that make them rigorous, general, and useful; and then proclaims the resulting "plain-vanilla Bayesian approach" to be a great advance over FISST because, being "straightforward," it avoids point process theory altogether.

As we argue in [38, 43] and in section 6, all such imitator-critics have not only completely missed the point but have embraced the same fallacy: the belief that *mere changes of notation add technical substance*. Most ironically, some who have disparaged FISST as "obfuscated" have unwittingly re-invented random set concepts in highly obfuscated notation. Moreover and as we argue in section 3.1, the "plain-vanilla Bayesians" have manufactured a spurious appearance of simplicity by promoting a succession of algorithms that are certainly "straightforward" but also afflicted by inherent—but less than candidly acknowledged—computational "logjams."

The point of FISST is not that multitarget problems can be formulated in terms of random sets. It is, rather, that random set techniques provide a systematic toolbox of explicit, rigorous, and general procedures that address many difficulties—those involving ambiguous evidence, unification, and computation, especially. A major purpose of this paper is to describe this toolbox.

1.7 Organization of the paper

We sketch the historical antecedents of FISST in section 2. Section 3 explains why FISST is necessary not only for unification of information fusion within a formal Bayes modeling framework, but is crucial if the computational "logjams" of multisensor-multitarget filtering and sensor management are to be surmounted. The main elements of FISST are summarized in section 4. Recent and current applied-research applications of FISST are summarized in section 5. The criticisms and claims of the imitator-critics of FISST are addressed in section 6. We list potential future research directions in section 7 and Errata in section 8. Conclusions may be found in section 9.

2 Random sets and information fusion: A brief bibliographical history

This section sketches the antecedents of FISST. At a purely mathematical level, FISST is a synthesis of two separate though intertwined strands of research: *random measure theory*, employed as a foundation for the study of randomly varying populations; and *random set theory*, employed as a foundation for the study of randomly varying geometrical shapes. We summarize these strands in sections 2.1 and 2.2, respectively. We conclude in section 2.3 by citing related lines of research.

2.1 Statistics of multi-object systems

Point process theory arose historically as the mathematical theory of randomly-varying time-arrival phenomena such as queues. For example, the number of customers waiting in line for a bank teller varies randomly with time. The Poisson distribution $\pi(n) = e^{-\lambda} \lambda^n/n!$ is the simplest model of random processes of this type. It states that the probability that there will be n customers in a queue at any given time is $\pi(n)$ where λ is the average length of the queue. But what if the customers could be anywhere in the building? Their randomness would include not only number but position. Such a dispersed queue is an example of a *population process* or *multidimensional point process* [9, 32, 83, 86]. Moyal [65] introduced the first theory of population processes over forty years ago. Noting that population processes could be regarded as randomly varying sets, he instead chose a different mathematical formulation: random counting measures (see section 6.1). By the time of Fisher's survey in 1972

[16], random measure theory had become the usual language for point processes among mathematicians.

Other mathematicians, meanwhile, were concerned about a completely different problem. In ordinary signal processing one models physical phenomena as "signals" obscured by "noise," the former to be extracted from the latter by optimal filters. But how can we construct such filters if the signal is some geometric shape and the noise is a geometrically structured clutter process? Mathéron devised the first systematic theory of random (closed) subsets in 1975 to address such problems [62]. It was subsequently shown that random locally finite subsets provide an essentially equivalent mathematical foundation for point processes: for example, Ripley in 1976 [76] and Baudin in 1986 [5]. This point is widely understood (see pp. 100-102 of [86] and section 6.1).

2.2 Statistics of expert systems

Beginning in the 1970s, a number of researchers throughout the world began to uncover close connections between random set theory and many aspects of expert system theory. Orlov [69, 70] and Höhle [27] demonstrated relationships between random sets and fuzzy set theory, though the most systematic work in this direction is due to Goodman [18]. In 1978, Nguyen connected random set theory with the Dempster-Shafer theory of evidence [68]. Dubois and Prade have published a number of papers relating random sets with expert system theory [12]. In the late 1990s, the author proposed random sets as a means of probabilistically modeling rule-based evidence [53, 54], as well as a basis for versions of crisp and fuzzy Dempster-Shafer theory that are consistent with Bayes' rule [17, 39, 40].

By the early 1990s, some researchers were proposing random set theory as a unifying foundation for expert systems theory [24, 75]. Such publications included books by Goodman and Nguyen in 1985 [19] and Kruse, Schwecke and Heinsohn in 1991 [35]. The author was the primary organizer of a 1996 international workshop, the purpose of which was to bring together the different communities of random set researchers [21].

Some of the potentially most profound research in recent years has begun to place possibility (Sugeno) measures on a firm theoretical basis in real analysis [22, 23]. We will return to this in section 7.3.

2.3 Related research

In 1986 Mori et. al. proposed random sets as a foundation for multitarget filtering, though in the context of multiple-hypothesis techniques [64]. In 1987 Washburn proposed a multitarget Bayes filter based on point processes, though with the number of targets presumed known [90]. More recently, Portenko et. al. have used branching-process concepts to model target appearance and disappearance in 1997 [73]. Brown and Liu have proposed the use of point processes in modeling for threat assessment [7].

The earliest work on the multitarget Bayes filter, when the number n of targets is not known and must be determined along with the states of the targets, appears to be due to Miller, O'Sullivan, Srivistava, Lanterman, et. al. [36]. Their "jump diffusion" approach solves stochastic diffusion equations on non-Euclidean manifolds. It is also apparently the only approach to systematically deal with continuous evolution of the multitarget state. Mahler was apparently the first to systematically deal with the general discrete state-evolution case (Bethel and Paras [6] assume discrete observation and state variables). In recent years several researchers have implemented the multitarget Bayes filter using particle-systems, Markov chain, and other approximations. Representative instances include Agate et al. [1], Ballantyne et. al. [3], Doucet et. al. [11], Everett et. al. [15], Hue, Le Cadre et. al., [29], Isard and MacCormick [30], and Orton and Fitzgerald [71].

The approach of Stone et. al. [85] is best described as heuristic (see pp. 42, 91-93 of [45] and p. 24 of Chapter 14 of [52]). Kastella's "JMP [joint multitarget probabilities], and the conceptual apparatus surrounding it, are elements of…finite-set statistics (FISST)" (see p. 27 of [66]. Kastella's "multitarget microdensity" approach [31] is a re-invention of basic FISST concepts using random density notation rather than random set notation for a simple point process (see [38] and section 6.1).

In recent years several researchers have initiated studies of FISST techniques. Challa, Vo, and Wang [8] have shown that the IPDA tracking approach arises directly from the FISST methodology. Implementations of FISST-based multitarget filters are being investigated by Lin, Kirubarajan, and Bar-Shalom [37]; Moreland and Challa [63]; Panta, Vo, Doucet, and Singh [72]; Punithakumar and Kirubarajan [74]; Sidenbladh, Svenson, and Schubert [81]; Sidenbladh and Wirkander [82]; Shoenfeld [79]; and Tobias and Lanterman [87]. Much of this work has centered on the FISST probability hypothesis density (PHD) filter of section 5.3.

3 Why random sets—or FISST?

We continue the discussion of section 1.1 by describing in greater detail the major motivating factors underlying FISST. We begin, in sections 3.1 and 3.2, by spotlighting the daunting computational complexity of multisensor-multitarget problems, and therefore why "plain-vanilla Bayesian" implementations are not only heuristic but inherently intractable under realistic conditions. Then we describe some of the ways in which single-target and multitarget statistics differ (section 3.3); summarize the elements of formal Bayes modeling (section 3.4); and describe what is required to extend formal modeling to multisensor-multitarget problems (section 3.5).

3.1 Why isn't general multitarget filtering "straightforward"?

We (1) begin with a computational analysis of a "straightforward," "plain-vanilla Bayesian," fixed-grid implementation of the single-sensor, single-target Bayes filter; (2) extend this analysis to the open-loop single-sensor, multitarget filter; and (3) further extend it to the closed-loop multisensor-multitarget filter—i.e., to sensor management. See [43] for a more extensive discussion. [1]

- The *single-sensor, single-target filter* [4]

$$f_{k+1|k}(\mathbf{x} \mid Z^k) = \int f_{k+1|k}(\mathbf{x} \mid \mathbf{w})\, f_{k|k}(\mathbf{w} \mid Z^k) d\mathbf{w} \qquad (1)$$

$$f_{k+1|k+1}(\mathbf{x} \mid Z^{k+1}) \propto f_{k+1}(\mathbf{z}_{k+1} \mid \mathbf{x})\, f_{k+1|k}(\mathbf{x} \mid Z^k) \qquad (2)$$

is described in more detail in section 3.4 below. It is far more computationally demanding than conventional techniques such as the EKF, and much recent research has been devoted to its approximate real-time implementation. Caution is in order, because in some of these efforts inherently intractable fixed-grid implementations have been successively hyped as "developed to the point where they can now be considered for transition to deployable systems," and then quietly abandoned in their own turn.

In fixed-grid approximation, one chooses fixed bounded regions of the measurement space $\mathbb{R}^M$ and the state space $\mathbb{R}^N$ and discretizes these regions into collections of $\mu = \mu_0^M$ measurement-cells and $\nu = \nu_0^N$ state-cells, where μ_0 and ν_0 are the respective numbers of single-dimensional cells. One models target and sensor constraints by heuristically specifying Markov transitions $f_{k+1|k}(\mathbf{y}|\mathbf{x})$ from each state-cell to all others; and sensor noise by heuristically specifying the likelihood $f_{k+1}(\mathbf{z}|\mathbf{x})$ that the sensor will collect an observation-cell if the target is present at a given state-cell. One then implements Eqs. (1) and (2) on a cell-by-cell basis, with the target state estimate at time-step k being that cell which maximizes the discretized version of the posterior $f_{k|k}(\mathbf{x}|Z^k)$.

However, as the statisticians J.C. Naylor and A.F.M. Smith have remarked, "The implementation of Bayesian inference procedures can be made to appear deceptively simple" (p. 214 of [67])—as is indeed the case here.

The most obvious issue is computational tractability. Let a be the number of operations required to compute the Markov density $f_{k+1|k}(\mathbf{y}|\mathbf{x})$ for fixed $\mathbf{y}$, $\mathbf{x}$. Then computation of $f_{k+1|k}(*|Z^k)$ using Eq. (1) requires at least $\nu^2 a$ operations. Let b be the number of operations required to compute the likelihood $f_{k+1}(\mathbf{z}|\mathbf{x})$ for fixed $\mathbf{z}$, $\mathbf{x}$. Then computation of $f_{k+1|k+1}(*|Z^{k+1})$ using Eq. (2) requires at least νb operations. So, each recursive cycle of Eqs. (1) and (2) requires at least $\nu^2 a+\nu b$ operations—which is computationally problematic for real-time application. Computational complexity can be reduced only for "toy" problems. For example, $\nu^2 a+\nu b$ becomes $(1+b)\nu$ if we assume that states transition only to immediately neighboring cells; and to 2ν if we further assume binary observations ($\mathbf{z} = 0$ or $\mathbf{z} = 1$).

Matters can be improved if we use a principled approximation technique such as particle systems [10]. The number ν of state-cells is replaced by the generally smaller number π of particles and $\nu^2 a+\nu b$ becomes $(a+b)\pi$—or $(a+1)\pi$ if binary observations are assumed. However, note that π will be large in those applications for which particle methods are appropriate—e.g., in applications where traditional methods fail, e.g. low SNR.

A more subtle issue is the fact that, in this "plain-vanilla Bayesian approach," *modeling and implementation have been irrevocably intertwined*. The models are *heuristic contrivances that are specific to a particular implementation technique*—fixed-grid or particle discretizations. Indeed, this is the exact opposite of the formal Bayes modeling paradigm described in section 3.4.

- The *single-sensor, multitarget Bayes filter*

$$f_{k+1|k}(X \mid Z^{(k)}) = \int f_{k+1|k}(X \mid W)\, f_{k|k}(W \mid Z^{(k)}) \delta W \qquad (1')$$

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) \propto f_{k+1}(Z_{k+1} \mid X)\, f_{k+1|k}(X \mid Z^{(k)}) \qquad (2')$$

is described in more detail in sections 3.5 and 4.5 below. It is far more computationally challenging than its single-sensor, single-target special case, and so even more powerful approximation strategies are required. Once again, simple but inherently intractable "plain-vanilla Bayesian" fixed-grid implementations have been successively hyped as "powerful and robust computational methods," only to be quietly abandoned in turn.

Simplify by supposing that target number n is known and target motions are independent, so that the multitarget Markov density $f_{k+1|k}(Y|X)$ must have the form

$$f_{k+1|k}(\mathbf{y}_1,\ldots,\mathbf{y}_n|\mathbf{x}_1,\ldots,\mathbf{x}_n) = f_{k+1|k}(\mathbf{y}_1|\mathbf{x}_1) \cdots f_{k+1|k}(\mathbf{y}_n|\mathbf{x}_n)$$

Computation of $f_{k+1|k}(*|Z^{(k)})$ using Eq. (1′) requires at least $n\nu^{2n}a$ operations. Assume no missed detections or false alarms, so that the multitarget likelihood for a single sensor must have the form

$$f_{k+1}(\mathbf{z}_1,\ldots,\mathbf{z}_n \mid \mathbf{x}_1,\ldots,\mathbf{x}_n) = \tfrac{1}{n!}\sum_{\sigma} f_{k+1}(\mathbf{z}_1 \mid \mathbf{x}_{\sigma 1})\cdots f_{k+1}(\mathbf{z}_n \mid \mathbf{x}_{\sigma n})$$

where the summation is over all permutations σ on $1,\ldots,n$. Computation of $f_{k+1}(\mathbf{z}_1,\ldots,\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_n)$ for fixed $\mathbf{x}_1,\ldots,\mathbf{x}_n$ requires at least $n{\cdot}n!b$ operations. So, computation of $f_{k+1|k+1}(*|Z^{(k+1)})$ using Eq. (2′) requires at least $n{\cdot}n!\nu^n b$ operations. Thus at least $n\nu^{2n}a+n{\cdot}n!\nu^n b$ are required for each recursive cycle of Eqs. (1′) and (2′).

We can attempt to increase tractability by stripping off even more application realism than is already the case. We get $n\nu^n(1+n!b)$ if states transition only to immediately neighboring cells; and $n\nu^n(1+b)$ if in addition targets are well-separated; and $2n\nu^n$ if in addition observations are binary. But even with the resulting drastic, cumulative loss of realism, "plain-vanilla Bayesian" implementation is *inherently intractable even when using the simplest possible model of targets moving between discrete locations in one dimension*, i.e. when $N = 1$.

Once again, matters can be improved if we employ particle-systems techniques—but only up to a point. Roughly speaking, the number ν^n of multitarget cells is replaced by the number Π of multitarget particles. In this case at least $n(a+n!b)\Pi$ operations are required, or $n(a+1)\Pi$ if observations are binary and targets are well-separated. Significant increases in computational efficiency can be achieved since Π will often be substantially smaller than ν^n. However, it is disingenuous of the "plain-vanilla Bayesians" to claim tractability and practicality, given their unrealistic assumptions.

In general Π must be made very large under realistic conditions, such as lower SNR, where conventional methods such as MHT fail and therefore particle methods are not inappropriate. So one must ask: In which

---

[1] Note: The discussion of computational complexity in [43] was somewhat careless and is corrected in what follows. Our ultimate conclusions remain unchanged.

applications will particle-system implementation of Eqs. (1′) and (2′) be *both* tractable and appropriate? What does one do when it is appropriate but *intractable*?

A more subtle issue is that the multitarget models are, once again, heuristic contrivances specific to some particular implementation technique. This should be contrasted with the multitarget formal Bayes modeling paradigm described in section 3.5 below.

• The theoretical basis for sensor management, the *closed-loop multisensor-multitarget Bayes filter*, is truly computationally daunting. It consists of Eqs. (1′) and (2′) used with a multisensor-multitarget likelihood function, alternated with the optimization of some objective function. Predictably, inherently intractable "plain-vanilla Bayesian," fixed-grid approximation is being promoted as a practical approach to sensor management.

Let there be n^* identical but conditionally independent sensors with sensor state space $\mathbb{R}^{N^*}$. In this case $f_{k+1|k+1}(X|Z^{(k+1)})$ is a function of the states $\mathbf{x}^{*1}_{k+1},...,\mathbf{x}^{*n^*}_{k+1}$ of the sensors at the next time-step k+1, as well as of the unknown observation-sets $Z^1_{k+1},...,Z^{n^*}_{k+1}$ that the sensors will collect at that time. Denote $\mathbf{x}^*_{k+1} = (\mathbf{x}^{*1}_{k+1},...,\mathbf{x}^{*n^*}_{k+1})$. In closed-loop multitarget filtering, after computation of $f_{k+1|k+1}(X) = f_{k+1|k+1}(X|Z^{(k+1)})$ from $f_{k+1|k}(X) = f_{k+1|k}(X|Z^{(k+1)})$, we are to select $\mathbf{x}^*_{k+1}$ by optimizing some objective function. This could be, for example, the expected value

$$I(\mathbf{x}^*_{k+1}) = \int I(\mathbf{x}^*_{k+1}, Z^1,...,Z^{n^*})\, f_{k+1}(Z^1,...,Z^{n^*})\delta Z^1 \cdots \delta Z^{n^*}$$

of the multitarget Kullback-Leibler discrimination [44]

$$I(\mathbf{x}^*_{k+1}, Z^1,...,Z^{n^*}) = \int f_{k+1|k+1}(X) \log\left(\frac{f_{k+1|k+1}(X)}{f_{k+1|k}(X)}\right)\delta X$$

where

$$f_{k+1}(Z^1,...,Z^{n^*}) = \int f^1_{k+1}(Z^1 \mid X) \cdots f^{n^*}_{k+1}(Z^{n^*} \mid X)\, f_{k+1|k}(X)\delta X$$

and where the $f^i_{k+1}(Z^1 \mid X)$ are the multitarget likelihoods for the sensors. Discretize the state and measurement spaces as before. In addition, chose a fixed region of $\mathbb{R}^{N^*}$ and discretize it into a collection of $\nu^* = (\nu^*_0)^{N^*}$ cells, where ν^*_0 is the number of single-dimensional cells. Once again, assume that the number of targets is known. Then for fixed $\mathbf{x}^*_{k+1}$, computation of $I(\mathbf{x}^*_{k+1})$ requires at least $(n^* nn!b) \cdot \nu^n \cdot (\mu^n)^{n^*}$ operations—since at least $n^* nn!b$ are required to compute the $f^i_{k+1}(Z^i \mid X)$'s; at least ν^n to compute $\int \cdot\, \delta X$; and at least $(\mu^n)^{n^*}$ to compute the $\int \cdot\, \delta Z^i$'s. We must compute $I(\mathbf{x}^*_{k+1})$ for each choice of $\mathbf{x}^*_{k+1}$ before determining which value of $I(\mathbf{x}^*_{k+1})$ is largest. So, at least $(n^* nn!b) \cdot \nu^n \cdot (\mu^n \nu^*)^{n^*}$ operations are required for each recursive cycle.

As before we can try to increase tractability by stripping off additional real-world complexity: $(n^* n!) \cdot \nu^n \cdot (\nu^*)^{n^*}$ if all sensor observations are binary; $n^* \cdot \nu^n \cdot (\nu^*)^{n^*}$ if targets are well-separated; and $\nu^n \cdot \nu^*$ if there is only one sensor. If we replace multitarget cells by multitarget particles, $\int \cdot\, \delta X$ is replaced by a summation over the number Π of multitarget particles and so a minimum of $(n^* nn!b) \cdot \Pi \cdot (\mu^n \nu^*)^{n^*}$ operations are required—or $n^* \cdot \Pi \cdot (\nu^*)^{n^*}$ given all simplifying assumptions. But even then, $(\nu^*)^{n^*}$ will be small only if n^* is small—e.g., if we further assume $n^* = 1$ (single sensor). That is: even with resort to particle methods and to drastic stripping away of application realism, "plain-vanilla Bayesian" fixed-grid discretization (in this case, of the sensor state space) leads to inherent computational "logjams."

3.2 Beyond heuristics and "logjams"

Algorithms can be always cobbled together using catch-as-catch-can, brute-force implementations such as those just described. However, "straightforward" approximation can lead not only to intractability but to numerical instability, poor convergence, and so on. Also, algorithm behavior may be difficult to diagnose because of the hidden assumptions and *ad hoc* design choices inherent to heuristic implementation.

Clearly, deeper insight is required if such difficulties are to be surmounted. *The "plain-vanilla Bayesian approach" is actually an obstacle to such insight.* Because of its obscurantist insistence that modeling and computational implementation must be "straightforward" (i.e., simplistically intertwined), and that anything else is "obfuscated," one dares not investigate the deep structure of multitarget filtering—and thereby possibly devise deeper approximation techniques.

But this investigation is precisely what FISST dares to attempt. Instead of adopting "straightforward" but simplistic approximations and then stripping off realism until they become tractable, as in section 3.1, we need an exactly contrary paradigm. We need to presume as much realism as possible—e.g., lower SNR, target appearance and disappearance—and then try to devise approximations that are potentially tractable *despite* this fact.

Furthermore, we need a formal modeling methodology that is *non-heuristic and implementation-independent*. That is, it results in general mathematical formulas for Markov densities and likelihood functions, constructed from general statistical models. Once one has chosen a specific implementation technique these formulas can be carefully reformulated in terms of this technique. This will ensure that the statistical assumptions underlying the original models are being implemented as faithfully as possible. This in turn helps ensure that the conditions under which the implemented filter will be approximately optimal are explicitly specified. Algorithm behavior should also tend to be more explicable, because assumptions and important design decisions in both the model and the approximations have been carefully parsed into a systematic, disciplined chain of reasoning.

One of the purposes of FISST is to provide a mathematical toolbox that makes these things possible. Tools for formal modeling are described in sections 3.5 and 4.3, and tools for computation in section 4.7.

3.3 How is multitarget statistics different?

FISST is "engineering-friendly" in that it is *geometric* (i.e., treats multitarget systems as visualizable images);

and directly generalizes the Bayes "Statistics 101" formalism that most signal processing engineers already understand—including formal Bayes-statistical modeling methods. However, these methods do not generalize in a straightforward manner [38, 43, 45, 52, 56].

The following are examples of how multisensor-multitarget statistics differs from single-sensor, single-target statistics—differences that FISST specifically addresses. The standard Bayes-optimal state estimators are not defined in general, and neither are such familiar concepts as expected value, least-squares optimization, and Shannon entropy. Other concepts, such as miss distance, require major reworking [25]. Also, without FISST no explicit, general, and systematic techniques exist for modeling multisensor-multitarget problems and then transforming these models into Bayesian form.

A more subtle theoretical issue is the fact that the multitarget analogs of concepts such as integrals and density functions do not exactly follow the usual measure-theoretic "recipe" (see section 4.8).

3.4 What is formal Bayes modeling?

In formal modeling [43, 56], one begins with a careful Bayes-statistical specification (a model) of the problem. Then, from it, one derives an implementation-independent, mathematically optimal solution and theoretically principled approximations of this solution.

This modeling methodology has become the accepted standard in single-sensor, single-target R&D. We begin with a continuous-variable *formal measurement model* that describes sensor behavior and a continuous-variable *formal motion model* that models interim target behavior:

$$\mathbf{z}_k = h_k(\mathbf{x}) + \Delta\mathbf{z}_k, \qquad \mathbf{x}_{k+1} = g_k(\mathbf{x}) + \Delta\mathbf{x}_k \tag{3}$$

where $\Delta\mathbf{z}_k$ and $\Delta\mathbf{x}_k$ are random noise vectors with carefully specified statistical properties. Using undergraduate calculus, one can show that the *true likelihood density* and *true Markov density*—i.e., the densities that faithfully reflect the original models—are:

$$f_k(\mathbf{z} \mid \mathbf{x}) = f_{\Delta\mathbf{z}_k}(\mathbf{z} - h_k(\mathbf{x})), \qquad f_{k+1|k}(\mathbf{y} \mid \mathbf{x}) = f_{\Delta\mathbf{x}_k}(\mathbf{y} - g_k(\mathbf{x})) \tag{4}$$

In practice, these formulas can be looked up in a textbook and so we need never actually bother with their formal derivation. But in the multitarget case no textbook yet exists that allows us such an easy escape from mathematics and careful thinking—see section 4.3.

Given this, we can address single-sensor, single-target tracking, detection, and identification problems using the Bayes recursive filter of Eqs. (1) and (2):

$$f_{k+1|k}(\mathbf{x} \mid Z^k) = \int f_{k+1|k}(\mathbf{x} \mid \mathbf{w}) f_{k|k}(\mathbf{w} \mid Z^k) d\mathbf{w}$$

$$f_{k+1|k+1}(\mathbf{x} \mid Z^{k+1}) \propto f_{k+1}(\mathbf{z}_{k+1} \mid \mathbf{x}) f_{k+1|k}(\mathbf{x} \mid Z^k)$$

Here $f_{k|k}(\mathbf{x}|Z^k)$ is the time-evolving Bayes posterior distribution; Z^k : $\mathbf{z}_1,\ldots, \mathbf{z}_k$ is the time-sequence of measurements; and the Bayes normalization factor is

$$f_{k+1}(\mathbf{z}_{k+1} \mid Z^k) = \int f_{k+1}(\mathbf{z}_{k+1} \mid \mathbf{x}) f_{k+1|k}(\mathbf{x} \mid Z^k) d\mathbf{x} \tag{5}$$

The posterior distribution $f_{k|k}(\mathbf{x}|Z^k)$ encapsulates everything that we know about the target state at time-step k. It is not useful unless we have a "mathematical can opener" that allows us to extract the information that we really want: position, velocity, identity, etc. We do not have a Bayes-optimal solution unless this can opener is a *Bayes-optimal state estimator*—i.e., one that minimizes the Bayes risk (see p. 31 of [88]). The most usual such estimators are the posterior expectation and the maximum *a posteriori* estimator, respectively:

$$\hat{\mathbf{x}}_{k|k} = \int \mathbf{x} \cdot f_{k|k}(\mathbf{x} \mid Z^k) d\mathbf{x}, \qquad \hat{\mathbf{x}}_{k|k} = \arg\sup_{\mathbf{x}} f_{k|k}(\mathbf{x} \mid Z^k) \tag{6}$$

A state estimator should have other desirable properties, e.g. rapid and stable convergence to the actual target state.

3.5 What is multisensor-multitarget formal Bayes modeling?

One of the basic goals of FISST is to extend this accepted methodology to multisensor-multitarget problems.

- We would like to have a general, systematic procedure for constructing a *multitarget measurement model* for any given sensor, i.e. an equation of the form

$$\underbrace{Z}_{\text{all measurements}} = \underbrace{h(X)}_{\text{target-generated measurements}} \cup \underbrace{\Delta Z}_{\text{non-target generated measurements}} \tag{7}$$

Here, $h(X)$ models observations directly generated by targets, but taking things such as missed detections and the sensor field of view into account; whereas ΔZ models observations that are not target-generated (e.g., false alarms, clutter, electronic countermeasures (ECM), etc).

- We would like to have a general, systematic procedure for constructing a a *multitarget motion model*:

$$\underbrace{X}_{\text{all targets}} = \underbrace{g(X_0)}_{\text{tpre-existing targets (including target disappearance)}} \cup \underbrace{\Delta X}_{\text{newly appearing targets}} \tag{8}$$

Here, $g(X_0)$ describes the current states of all targets that previously existed, but taking into account the probability that any given target may disappear. Also, ΔX describes the generation of new targets in the scene.

- We would like to have a general, systematic procedure for transforming Eq. (7) into the corresponding true multitarget likelihood function $f_k(Z|X)$—i.e., the one that faithfully reflects the multitarget sensor model. Likewise we would like to have a general, systematic procedure for transforming Eq. (8) into the corresponding true multitarget Markov density $f_{k+1|k}(Y|X)$—i.e., the one that faithfully reflects the multitarget motion model. Such a procedure should be inherently *implementation-independent*, in that it produces general formulas for $f_k(Z|X)$ and $f_{k+1|k}(Y|X)$. Implementations that preserve the statistical assumptions underlying the original models can then be derived from these formulas.

If we shirk such issues we fail to grasp that there is an inherent problem—how do we know that $f_k(Z|X)$ is not a heuristic contrivance or not erroneously constructed? Any boast of "optimality" is hollow if $f_k(Z|X)$ models the wrong sensors. Similar comments apply to $f_{k+1|k}(Y|X)$.

- Once in possession of these basics we can address multisensor-multitarget tracking, detection, and identification problems using the multitarget Bayes filter of Eqs. (1′) and (2′)

$$f_{k+1|k}(X \mid Z^{(k)}) = \int f_{k+1|k}(X \mid W) f_{k|k}(W \mid Z^{(k)}) \delta W$$

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) \propto f_{k+1}(Z_{k+1} \mid X) f_{k+1|k}(X \mid Z^{(k)})$$

where

$$f_{k+1}(Z_{k+1} \mid Z^{(k)}) = \int f_{k+1}(Z_{k+1} \mid X) f_{k+1|k}(X \mid Z^{(k)}) \delta X \tag{5′}$$

(see Eqs. (29), (30), and (31) below). But even here we encounter complications because the integrals in any multitarget generalization of those equations must sum over not only all possible target states but also over all possible numbers of targets (see Eq. (11) below).

- We would like to have multitarget Bayes-optimal state estimators that allow us to determine the number and states of the targets. Here, however, we encounter an unexpected difficulty: *the naïve generalizations of the single-target Bayes-optimal estimators of Eq. (6) do not exist in general* (section 4.5). We must devise new estimators and prove that they are well-behaved.

4 A "road map" of FISST

In this section we summarize the major elements of FISST. The problem of accurately modeling multitarget state spaces and multisensor-multitarget measurement spaces is described in section 4.1. Belief-mass functions, set derivatives, and set integrals are introduced in section 4.2; and their application to multisensor-multitarget formal Bayesian modeling in section 4.3. The extension of formal modeling to ambiguous data sources is summarized in section 4.4. The multisource-multitarget Bayes filter is described in section 4.5. Probability generating functionals and their functional derivatives are introduced in section 4.6; and their use in systematic computational approximation explored in section 4.7. The relationship between FISST and measure-theoretic probability theory is addressed in section 4.8.

4.1 Random state- and measurement-sets

The complete description of the state of a multitarget system requires a *unified state representation:* a finite set of the form $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ where n is the number of targets and $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are the state vectors of the individual targets (in general, $\mathbf{x}$ is assumed to include a discrete identity/label state variable). This description must include the possibility $n = 0$—i.e., no target is present, in which case we write $X = \emptyset$. Such a unified representation accounts for the fact that n is variable and that targets have no physically inherent order. Thus $\{\mathbf{x}_1,\mathbf{x}_2\} = \{\mathbf{x}_2,\mathbf{x}_1\}$ is a single unified state-model of two targets with state-vectors $\mathbf{x}_1$, $\mathbf{x}_2$. On the other hand, vectors $(\mathbf{x}_1,\mathbf{x}_2)$ and $(\mathbf{x}_2,\mathbf{x}_1)$ do not correctly represent the physical multitarget state since they do so redundantly and cannot model its inherent permutation symmetry. Likewise, alternative point process representations add complexity with no new substance (see section 6.1).

In a careful Bayesian approach [38, 43] the unknown state must be a random quantity. Consequently, the unknown state-set at time-step k must be a randomly varying finite set $\Xi_{k|k}$. One cannot define a random variable of any kind without, typically, first defining a topology on the space of objects to be randomized and then defining random elements of that space in terms of the Borel subsets [38, 43]. The space of state-sets is topologized using the Mathéron "hit-or-miss" topology [20, 62]. Once this is done, the probability law of a finite random state-set Ξ is its probability-mass function (a.k.a. probability measure) $p_\Xi(O) = \Pr(\Xi\in O)$ where O is any Borel-measurable subset of the Mathéron topology.

Similar considerations apply to observations. We must begin with a *unified observation representation:* a finite set of the form $Z = \{\mathbf{z}_1,\ldots,\mathbf{z}_m\}$ where m is the number of observations and $\mathbf{z}_1,\ldots,\mathbf{z}_m$ are observation-vectors generated by all of the sensors from all of targets (in general, $\mathbf{z}$ is assumed to include a discrete sensor tag describing the originating sensor). When no observations have been collected we write $Z = \emptyset$.

Note that, in particular (and contrary to the assertions of the "plain-vanilla Bayesians"), the $\mathbf{z}_i$ *need not be post-detection reports*. It is routine practice to represent signature data as vectors: vectors of SAR pixels, of HRRR range-bins, of passive-acoustic frequency-bins, etc.

It should also be noted that sensors have their own state-vectors $\mathbf{x}^*$. So, in general sensor management applications (such as those involving UAVs), one must define a joint multisensor-multitarget state space in which the unified state is a finite set whose elements are state-vectors of both targets and sensors [50, 55].

4.2 Belief-mass functions and multitarget integro-differential calculus

Suppose that we have a random finite subset Ψ of some space $\mathbf{Y}$ (e.g., the space of target states or the space of measurements from any sensor). The statistical behavior of Ψ is described by its probability-mass function (a.k.a. probability measure) $\Pr(\Psi\in O)$. For engineering purposes it is inconvenient to deal with Borel sets O which are continuously infinite sets whose elements are finite sets. The *Choquet-Mathéron theorem* (see p. 30 of [62] or p. 96 of [20]) states that the additive probability measure $p_\Psi(O) = \Pr(\Psi\in O)$ is equivalent to the *non-additive* measure (a.k.a. "capacity" or "Choquet functional")

$$\pi_\Psi(S) = \Pr(\Psi\cap S\neq\emptyset)$$

where S is a subset of *ordinary* single-target state space. Therefore, $p_\Psi(O)$ is also equivalent to

$$\beta_\Psi(S) = 1-\pi_\Psi(S^c) = 1-\Pr(\Psi\cap S^c\neq\emptyset) = \Pr(\Psi\subseteq S).$$

So for engineering purposes we can dispense with $p_\Psi(O)$ entirely and use $\beta_\Psi(S)$ instead. By analogy with $p_\Psi(O)$ we call $\beta_\Psi(S)$ the *belief-mass function* (a.k.a. *belief measure*) of the random finite set Ψ.[2]

In single-target problems we usually prefer the density function $f_\mathbf{Y}(\mathbf{y})$ of $p_\mathbf{Y}(S)$ to $p_\mathbf{Y}(S)$ itself. The defining relationship between the two is

$$p_\mathbf{Y}(S) = \int_S f_\mathbf{Y}(\mathbf{y})d\mathbf{y} \tag{9}$$

in which case $f_\mathbf{Y}(\mathbf{y})$ is called the *Radon-Nikodým derivative* of $p_\mathbf{Y}(S)$.

In multitarget engineering we would likewise prefer a multitarget density function $f_\Psi(Y)$ to $\beta_\Psi(S)$. By analogy with Eq. (9) it should be defined by

$$\beta_\Psi(S) = \int_S f_\Psi(Y)\delta Y \tag{10}$$

---

[2] We use this terminology for the following reason. Suppose that $\mathbf{Y}$ is finite and that $\Psi\neq\emptyset$. Then $\mathrm{Bl}(S) = \Pr(\Psi\subseteq S)$ is a Dempster-Shafer belief function.

This equation does not make sense unless we define the indicated integral. Let $f(Y)$ be any real-valued function of a finite-set variable $Y \subseteq \boldsymbol{Y}$ which has the following property. For each $n \geq 0$, use the convention $f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\}) = 0$ whenever $\mathbf{y}_i = \mathbf{y}_j$ for some $i \neq j$, and also assume that $\int f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\})d\mathbf{y}_1\cdots d\mathbf{y}_n$ is finite and has no units of measurement. Then the *set integral* of $f(Y)$ in a region $S \subseteq \boldsymbol{Y}$ is defined as

$$\int_S f(Y)\delta Y = f(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{S^n} f(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\}) d\mathbf{y}_1 \cdots d\mathbf{y}_n \quad (11)$$

Given any belief-mass function $\beta_\Psi(S)$, how can we construct its corresponding density function $f_\Psi(Y)$ so that Eq. (10) is satisfied? This requires the inverse operation of the set integral, the *set derivative*. For arbitrary functions $F(S)$ of a finite-set variable S and for $\mathbf{y}_1,\ldots,\mathbf{y}_n$ distinct, it is defined by

$$\frac{\delta F}{\delta \mathbf{y}}(S) = \lim_{v(E_\mathbf{y})\to 0} \frac{F(S \cup E_\mathbf{y}) - F(S)}{v(E_\mathbf{y})} \quad (12)$$

$$\frac{\delta \beta}{\delta Y}(S) = \frac{\delta^n \beta}{\delta \mathbf{y}_n \cdots \delta \mathbf{y}_1}(S) = \frac{\delta}{\delta \mathbf{y}_n} \frac{\delta^{n-1}\beta}{\delta \mathbf{y}_{n-1} \cdots \delta \mathbf{y}_1}(S)$$

where $E_\mathbf{y}$ is a small neighborhood of $\mathbf{y}$; and where $v(S)$ is the hypervolume (i.e., Lebesgue measure) of set S.[3]

The set derivative is the continuous-variable analog of the Möbius transform of Dempster-Shafer theory (see p. 149 of [20]). It can be computed using "turn the crank" rules such as the following (see pp. 31-32 of [45], pp. 143, 146, 151 of [20]):

- *Sum Rule*:

$$\frac{\delta}{\delta Y}\big(a_1\beta_1(S) + a_2\beta_2(S)\big) = a_1 \frac{\delta \beta_1}{\delta Y}(S) + a_2 \frac{\delta \beta_2}{\delta Y}(S); \quad (13)$$

- *Product Rule*:

$$\frac{\delta}{\delta Y}\big(\beta_1(S)\beta_2(S)\big) = \sum_{W \subseteq Y} \frac{\delta \beta_1}{\delta W}(S) \frac{\delta \beta_1}{\delta (Y-W)}(S); \quad (14)$$

- *Chain Rule*:

$$\frac{\delta}{\delta \mathbf{y}}\big(f(\beta_1(S),\ldots,\beta_n(S))\big) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\beta_1(S),\ldots,\beta_n(S)) \frac{\delta \beta_i}{\delta \mathbf{y}}(S); \quad (15)$$

- *Constant Rule*: If $Y \neq \emptyset$ then

$$\frac{\delta}{\delta Y} K = 0 \quad (16)$$

- *Power Rule*: If $p(S)$ is a probability mass function with density function $f_p(\mathbf{y})$ then

$$\frac{\delta}{\delta Y} p(S)^n = \begin{cases} \frac{n!}{(n-k)!} p(S)^{n-k} f_p(\mathbf{y}_1) \cdots f_p(\mathbf{y}_k) & \text{if} \quad k \leq n \\ 0 & \text{if} \quad k > n \end{cases} \quad (17)$$

Given this it can be shown that

$$\beta_\Psi(S) = \int_S \frac{\delta \beta_\Psi}{\delta Y}(\emptyset)\delta Y \quad (18)$$

That is,

$$f_\Psi(Y) = \frac{\delta \beta_\Psi}{\delta Y}(\emptyset) \quad (19)$$

is the multi-object density function of $\beta_\Psi(S)$.

Note that because of Eqs. (10) and (18), $f_\Psi(Y)$ and $\beta_\Psi(S)$ contain exactly the same information, and so are *equivalent representations of the probability law of* Ψ.

[3] *Warning*: the first equation of (12) has been simplified for the sake of clarity. For further details, see pp. 144-151, 157-162 of [20]. The notation $\delta/\delta\mathbf{x}$ is a simplified version of a common notation used in physics (see pp. 173-174 of [78]).

As an example, let $\Psi = \{\mathbf{Y}\}$ where $\mathbf{Y}$ is a random vector in $\boldsymbol{Y}$. Then $\beta_\Psi(S) = \Pr(\{\mathbf{Y}\}\subseteq S) = \Pr(\mathbf{Y}\in S) = p_\mathbf{Y}(S)$. Likewise, the set integral becomes an ordinary single-target integral and Eq. (10) becomes Eq. (9). In other words: FISST multi-object statistics reduces to the usual "Statistics 101" formalism in the single-object case.

4.3 Multisensor-multitarget Bayes modeling

Belief-mass functions and their set derivatives provide us with the means for generalizing formal Bayes modeling to multisensor-multitarget problems as in [45]. We construct a formal measurement model $\Sigma_k = T_k(X) \cup C_k(X)$ and motion model $\Xi_{k+1|k} = D_k(X) \cup B_k(X)$. Then we construct their corresponding belief-mass functions:

$$\beta_k(S \mid X) = \Pr(\Sigma_k \subseteq S \mid X) \quad (20)$$

$$\beta_{k+1|k}(T \mid X) = \Pr(\Xi_{k+1|k} \subseteq T \mid X)\,. \quad (21)$$

Finally, from Eq. (19) we can explicitly construct general, implementation-independent formulas for the true multitarget likelihood function and the true multitarget Markov density using:

$$f_k(Z \mid X) = \frac{\delta \beta_k}{\delta Z}(\emptyset \mid X) \quad (22)$$

$$f_{k+1|k}(Y \mid X) = \frac{\delta \beta_{k+1|k}}{\delta Y}(\emptyset \mid X) \quad (23)$$

These multi-object density functions contain the same information as their respective belief-mass functions—and therefore the same information as the models used to construct those belief-mass functions.

4.4 Formal modeling of ambiguous data

We would like to extend the formal modeling process just outlined to data sources that exhibit varying degrees of ambiguity due to ignorance [58].

As previously noted, ambiguity of information sources has two major manifestations. First, the source may be inherently ambiguous in and of itself—e.g., natural language statements, attributes extracted from sensor signatures by human operators, rules drawn from knowledge bases, etc. It is unclear how to represent such evidence as "data" $\mathbf{z}$, let alone how to construct a likelihood function $f(\mathbf{z}|\mathbf{x})$ that models the process by which it is generated. Second, some forms of data, such as SAR images, may not be ambiguous in this sense; but their likelihood functions cannot be constructed with precision because of inherently uncharacterizable real-world statistical variations (e.g., irregular placement of equipment, wet mud, dents, etc., in the case of SAR).

There are three steps in the FISST approach to dealing with data that is ambiguous in and of itself. First, such evidence is modeled as a *random closed subset* Θ *of the underlying measurement space*. Second, some modeling technique—fuzzy logic, Dempster-Shafer theory, rules—is used to construct Θ. Third, a *generalized likelihood function* $\rho(\Theta|\mathbf{x})$, based on an *ambiguous signature model base* and a *data-to-model matching technique*, is used to hedge against uncertainties both in data-modeling and in the modeling of data generation.

As for ambiguously generated unambiguous evidence, generalized likelihood functions $\rho(\mathbf{z}|\mathbf{x})$ are used to model what is known about data generation and to hedge against the uncertainties caused by what is not known.

Once these steps have been taken, we use the same basic procedures to be outlined in later sections, but employing all three types of likelihood functions as necessary: $\rho(\Theta|\mathbf{x})$, $\rho(\mathbf{z}|\mathbf{x})$, or $f(\mathbf{z}|\mathbf{x})$.

We briefly describe each of these steps in more detail:

• *Why random set models of ambiguous data?* Begin with a simple example. Suppose that we are given the measurement model $\mathbf{Z} = C\mathbf{x} + \mathbf{W}$ where $\mathbf{x}$ is the target state, $\mathbf{W}$ is random noise, and C is an invertible matrix. Let B be an observation that is imprecise in the sense that it is a subset of measurement space that merely constrains the possible values of $\mathbf{z}$—i.e., $\mathbf{z} \in B$. If we are uncertain about the validity of the constraint $\mathbf{z} \in B$ then there may be many possible constraints of varying degrees of plausibility. We model this ambiguity as a randomly varying subset Θ of measurements, where $\Pr(\Theta = B)$ represents our degree of belief in the specific constraint B. The random subset of all states that are consistent with Θ is then $\Gamma = \{C^{-1}(\mathbf{z}\text{-}\mathbf{W}) \mid \mathbf{z} \in \Theta\}$.

• *Random set modeling techniques*. But how do we construct useful models Θ? We could use a fuzzy-set modeling process, in which ambiguity in the data $\mathbf{z}$ is modeled as a fuzzy membership function $g(\mathbf{z})$ on measurement space. Let A be a uniformly distributed random number in $[0,1]$. Then the random subset

$$\Theta = \Sigma_A(g) = \{\mathbf{z} \mid A \le g(\mathbf{z})\} \qquad (24)$$

contains the same information as the fuzzy model $g(\mathbf{z})$. Alternatively, we could use a Dempster-Shafer modeling process, in which uncertainty in the data is modeled as a basic mass assignment $m(S)$ on measurement space, with $m(S) = 0$ for all but a finite number of S. In this case we can replace $m(S)$ by a random subset Θ such that $\Pr(\Theta{=}S) = m(S)$. Or, suppose that the evidence takes the form of a rule drawn from a knowledge base. Rules have the form $X \Rightarrow S$ = '*if X then S*' where X,S are subsets of a (finite) universe U with N elements. let Φ be a uniformly distributed random subset of U—that is, $\Pr(\Phi{=}S) = 2^{-N}$ for all S. Then a random set representation of the rule $X \Rightarrow S$ is [53]

$$\Sigma_A(X \Rightarrow S) = (S \cap X) \cup (X^c \cap \Phi) \qquad (25)$$

• *Generalized likelihood functions for ambiguous data*. Given a random set model Θ of an ambiguous observation, the next step in a strict Bayesian formulation would be to specify a likelihood function for this data as a conditional probability density $f(\Theta|\mathbf{x})$. FISST employs an engineering compromise. From knowledge of the data source we construct an "ambiguous model base"—a family $\Sigma_\mathbf{x}$ of random closed subsets of measurement space indexed by $\mathbf{x}$. Then we choose a methodology for comparing observations Θ to models $\Sigma_\mathbf{x}$ and define the generalized likelihood function to be the probability that evidence matches a given model. For example,

$$\rho(\Theta|\mathbf{x}) = \Pr(\Theta \cap \Sigma_\mathbf{x} \neq \emptyset) \qquad (26)$$

measures the degree to which the evidence Θ does not flatly contradict the model $\Sigma_\mathbf{x}$. It can be shown that such generalized likelihood functions are conditional probabilities [46]. They are not conventional likelihood functions, however, since in general $\int \rho(\Theta|\mathbf{x})d\Theta = \infty$ (and thus the term "generalized").

Explicit and useful formulas can be constructed for generalized likelihood functions $\rho(\Theta|\mathbf{x})$ when evidence and models are based on fuzzy set models. For example, if $\Theta = \Sigma_A(g)$ and $\Sigma_\mathbf{x} = \Sigma_A(f_\mathbf{x})$ then Eq. (26) becomes

$$\rho(\Theta|\mathbf{x}) = \Pr(\Theta \cap \Sigma_\mathbf{x} \neq \emptyset) = \sup_\mathbf{z} \min\{f_\mathbf{x}(\mathbf{z}), g(\mathbf{z})\}$$

Bayes-rule filters based on these likelihood functions have been shown to result in good state estimation in spite of the ambiguity inherent in the evidence and the models. See [60, 84] for more details.

• *Generalized likelihood functions for ambiguously-generated unambiguous data*. The conventional approach to this problem would be "robust estimation" [28, 33]. For example, one common model of ambiguity is the "ε-contamination model," in which the unknown density is assumed to have the form $(1\text{-}\varepsilon)f_0 + \varepsilon g$ where f_0 is a nominal choice for the likelihood and g is in some class of probability distributions. The unknown state $\mathbf{x}$ can then be estimated robustly in a manner that is optimal in some rigorously specified sense.

However, there is a fundamental paradox associated with any approach that presumes a "certain representation of uncertainty." *In such approaches, the uncertainty model is chosen for its mathematical tractability rather than its actual pertinence to the unknowable structure of uncertainty*. Stated differently: How can one assume that any such model bears resemblance to the actual structure of uncertainty, since this structure is due to ignorance—and therefore we cannot know what it is?

The FISST approach is based on a different viewpoint: *The purpose of an uncertainty model is to hedge the estimation process against inherently unknowable uncertainties*. We assume that enough is known about the underlying likelihood function that it can be "trapped" in a *random error bar*: $L_\mathbf{z}(\mathbf{x}) \in J_\mathbf{z}(\mathbf{x})$, where for each fixed $\mathbf{z}$ and each fixed $\mathbf{x}$, $J_\mathbf{z}(\mathbf{x})$ is a random closed interval of the positive real numbers. If we have a sequence of independent, identically distributed observations $\mathbf{z}_1, \ldots, \mathbf{z}_m$, then interval arithmetic can be used to construct the random error bar

$$J(\mathbf{x}) = J_{\mathbf{z}_1,\ldots,\mathbf{z}_m}(\mathbf{x}) = J_{\mathbf{z}_1}(\mathbf{x}) \cdots J_{\mathbf{z}_m}(\mathbf{x}) \qquad (27)$$

for the nominal joint likelihood function:

$$L_{\mathbf{z}_1,\ldots,\mathbf{z}_m}(\mathbf{x}) = L_{\mathbf{z}_1}(\mathbf{x}) \cdots L_{\mathbf{z}_m}(\mathbf{x}) \in J(\mathbf{x}) \qquad (28)$$

Now let L be any *plausible likelihood function*—i.e., any density function such that $L(\mathbf{x}) \in J(\mathbf{x})$ for all $\mathbf{x}$. For each such likelihood we can construct a plausible state estimate $\mathbf{x}_L$ using the ML or MAP estimators. The set of all such plausible estimates is a random subset Γ of state space. The generalized likelihood function is defined to be the probability $\rho(\mathbf{z}_1,\ldots, \mathbf{z}_m|\mathbf{x}) = \Pr(\mathbf{x} \in \Gamma)$ that a given state $\mathbf{x}$ is plausible. The most plausible states are those states that maximize this generalized likelihood function. Explicit formulas are available for computing $\rho(\mathbf{z}_1,\ldots,\mathbf{z}_m|\mathbf{x})$. These formulas have an especially usable form if the original random errors bars are based on fuzzy set models: $J_\mathbf{z}(\mathbf{x}) = \Sigma_A(f_\mathbf{z})$. See [26] for more details.

4.5 Unified fusion of multisource-multitarget information

With these preliminaries in place we are in a position to generalize the single-sensor, single-target Bayes filter of Eqs. (1), (2), (5), and (6) to multisource-multitarget problems [45, 58]. They become, respectively,

$$f_{k+1|k}(X \mid Z^{(k)}) = \int f_{k+1|k}(X \mid W)\, f_{k|k}(W \mid Z^{(k)})\delta W \quad (29)$$

$$f_{k+1|k+1}(X \mid Z^{(k+1)}) \propto f_{k+1}(Z_{k+1} \mid X)\, f_{k+1|k}(X \mid Z^{(k)}) \quad (30)$$

$$f_{k+1}(Z_{k+1} \mid Z^{(k)}) = \int f_{k+1}(Z_{k+1} \mid X)\, f_{k+1|k}(X \mid Z^{(k)})\delta X \quad (31)$$

Here $f_{k|k}(X|Z^{(k)})$ is the multitarget posterior distribution; $Z^{(k)}$: $Z_1,\ldots, Z_k$ is the time-sequence of multisource measurement-sets; and the integrals are set integrals. The multitarget posterior distribution has the form

$f_{k|k}(\emptyset|Z^{(k)})$: no targets present

$f_{k|k}(\{\mathbf{x}_1\}|Z^{(k)})$: one target with state $\mathbf{x}_1$

$f_{k|k}(\{\mathbf{x}_1, \mathbf{x}_2\}|Z^{(k)})$: two targets with states $\mathbf{x}_1, \mathbf{x}_2$

⋮

$f_{k|k}(\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}|Z^{(k)})$: n targets with states $\mathbf{x}_1,\ldots,\mathbf{x}_n$

⋮

A subtle point to notice is this: in a careful Bayesian formulation [38, 43], $f_{k|k}(X|Z^{(k)})$ must be a *single function* defined on the unified multitarget state X. It is not correct to partition it by target number into a family of ordinary densities $f_{k|k}(\emptyset|Z^{(k)})$, $f_{k|k}(\mathbf{x}_1|Z^{(k)})$, $f_{k|k}(\mathbf{x}_1,\mathbf{x}_2|Z^{(k)})$,…., $f_{k|k}(\mathbf{x}_1,\ldots,\mathbf{x}_n|Z^{(k)})$,…. defined on vectors $\mathbf{x}_1$, $(\mathbf{x}_1,\mathbf{x}_2)$,…., $(\mathbf{x}_1,\ldots,\mathbf{x}_n)$,… Likewise, if one emphasizes the fact that the densities $f_{k|k}(\mathbf{x}_1,\ldots,\mathbf{x}_n|Z^{(k)})$ are inherently permutation-symmetric, then this means that multitarget states are also inherently permutation-symmetric—and therefore cannot be correctly modeled as vectors $(\mathbf{x}_1,\ldots,\mathbf{x}_n)$.

Care must be exercised with multitarget density functions insofar as units of measurements are concerned. Let u denote the units of measurement of $f_\Psi(\{\mathbf{y}\})$. Then the units of measurement of $f_\Psi(Y)$ are $u^{-|Y|}$. This has several practical consequences. First, the naïve multitarget analogs of various integrals such as entropy and square-error are undefined because of units-mismatch problems:

$$-\int f_\Psi(Y)\log f_\Psi(Y)\delta Y, \qquad \int \| f_{\Psi_1}(Y) - f_{\Psi_2}(Y)\|^2\, \delta Y$$

Second, *the naive multitarget analogs of the standard Bayes-optimal estimators of Eq. (6) are not defined in general.* For example, consider the naïve generalization of the MAP estimator. Simplify by assuming that targets are in the 1-D interval [0,2] and distance is measured in meters. Assume that $f_{k|k}(X|Z^{(k)})$ has the following form: $f_{k|k}(X|Z^{(k)}) = 0.5$ if $X = \emptyset$, $f_{k|k}(X|Z^{(k)}) = 0.25\ \mathrm{m}^{-1}$ for any $X = \{x\}$, and $f_{k|k}(X|Z^{(k)}) = 0$ otherwise. That is: there is a 50-50 chance that no target exists and, if otherwise, it is a single target that is equally likely to be anywhere in [0,2]. Since $f_{k|k}(X|Z^{(k)})$ is largest at $X = \emptyset$ the naïve estimator leads us to conclude that no target is present since $0.5 > 0.25$. Now change units of measurement from meters to kilometers. Then $f_{k|k}(X|Z^{(k)}) = 250\ \mathrm{km}^{-1}$ if $X = \{x\}$ and so a target *is* present! The paradox arises because *the naïve estimator prescribes an impossible procedure*: comparing a unitless quantity $f_{k|k}(X|Z^{(k)})$ (when $X = \emptyset$) to a quantity $f_{k|k}(X|Z^{(k)})$ with units (when $X = \{x\}$).

Consequently, new multitarget state estimators must be devised and shown to be Bayes-optimal, convergent, etc. (see pp. 42-44 of [45]). One such estimator is

$$\hat{X}_{k|k} = \arg\sup_X \frac{f_{k|k}(X \mid Z^{(k)})}{c^{|X|}} \quad (32)$$

where c is a constant that has the same units of measurement as $f_{k|k}(\{\mathbf{x}\}|Z^{(k)})$.

A novel aspect of any such multitarget state estimator is that it *optimally unifies into a single procedure two conflicting procedures that are normally accomplished separately*: *target detection* (determining whether or not targets exist and to what number) and *target estimation* (determining the states of the targets, if they exist).

Given these preliminaries, we have produced a unified approach for fusing disparate forms of evidence.

4.6 Probability generating functionals (p.g.fl.'s) and functional derivatives

What if we wanted to extend the concept of a belief-mass function to fuzzy sets? This is easily done by noting that Eq. (10) can be rewritten as

$$\beta_\Psi(S) = \int_S f_\Psi(Y)\delta Y = \int \mathbf{1}_S^Y \cdot f_\Psi(Y)\delta Y = \sum_{n=0}^{\infty}\frac{1}{n!}\int \mathbf{1}_S(\mathbf{y}_1)\cdots\mathbf{1}_S(\mathbf{y}_n)\, f_\Psi(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\})d\mathbf{y}_1\cdots d\mathbf{y}_n \quad (33)$$

where $\mathbf{1}_S(\mathbf{y})$ is defined by $\mathbf{1}_S(\mathbf{y}) = 1$ if $\mathbf{y} \in S$ and $\mathbf{1}_S(\mathbf{y}) = 0$ otherwise; and where

$$\mathbf{1}_S^Y = \prod_{\mathbf{y}\in Y}\mathbf{1}_S(\mathbf{y})$$

- Now, let $\mu(\mathbf{y})$ be the membership function for a fuzzy set. Then we generalize Eq. (33) as

$$G_\Psi[\mu] = \int \mu^Y \cdot f_\Psi(Y)\delta Y = \sum_{n=0}^{\infty}\frac{1}{n!}\int \mu(\mathbf{y}_1)\cdots\mu(\mathbf{y}_n)\, f_\Psi(\{\mathbf{y}_1,\ldots,\mathbf{y}_n\})d\mathbf{y}_1\cdots d\mathbf{y}_n \quad (34)$$

where

$$\mu^Y = \prod_{\mathbf{y}\in Y}\mu(\mathbf{y})\,. \quad (35)$$

In the point process literature $G_\Psi[\mu]$ is known as the *probability generating functional* (p.g.fl.) of Ψ (see pp. 141, 220 of [9]). Note that $G_\Psi[\mathbf{1}_S] = \beta_\Psi(S)$, so that p.g.fl.'s do indeed generalize belief-mass functions.

The p.g.fl. $G_\Psi[\mu]$ is, like the multi-object density $f_\Psi(X)$ and the belief-mass function $\beta_\Psi(S)$, a fundamental descriptor of the statistics of Ψ. But it is often more useful than $f_\Psi(X)$ or $\beta_\Psi(S)$ because it results in much simpler formulas. For example, suppose that we want to incorporate *target preference*—the fact that some targets are of greater tactical interest than others—into our models. We model target preference as a fuzzy membership function $\rho(\mathbf{x})$ on target states. Let

$$G_{k|k}[h] = \int h^X f_{k|k}(X|Z^{(k)})\delta X$$

be the p.g.fl. of the multitarget system at time-step k. Then it can be shown that the p.g.fl. of the same system, biased to emphasize current or potential targets of interest, is given by $G_{k|k}[h;\rho] = G_{k|k}[1-\rho+\rho h]$ (see [50, 57]).

- The set derivative of a belief-mass function can be generalized to functional derivatives of p.g.fl.'s. Recall that the *gradient derivative* (a.k.a. directional or Frechét derivative) of a real-valued function $G(\mathbf{x})$ in the direction of a vector $\mathbf{w}$ is

$$\frac{\partial G}{\partial \mathbf{w}}(\mathbf{x}) = \lim_{\varepsilon \to 0} \frac{G(\mathbf{x}+\varepsilon \cdot \mathbf{w}) - G(\mathbf{x})}{\varepsilon} \quad (36)$$

where for each $\mathbf{x}$ the function $\mathbf{w} \to \frac{\partial G}{\partial \mathbf{w}}(\mathbf{x})$ is linear and continuous; and so

$$\frac{\partial G}{\partial \mathbf{w}}(\mathbf{x}) = w_1 \frac{\partial G}{\partial w_1}(\mathbf{x}) + ... + w_N \frac{\partial G}{\partial w_N}(\mathbf{x})$$

for all $\mathbf{w} = (w_1,\ldots, w_N)$, where the derivatives on the right are ordinary partial derivatives. Likewise, the gradient derivative of a p.g.fl. $G[h]$ in the direction of the function g is

$$\frac{\partial G}{\partial g}[h] = \lim_{\varepsilon \to 0} \frac{G[h+\varepsilon \cdot g] - G[h]}{\varepsilon} \quad (37)$$

where for each h the functional $g \to \frac{\partial G}{\partial g}[h]$ is linear and continuous. In physics, gradient derivatives with $g = \delta_{\mathbf{x}}$ are called "functional derivatives" (pp. 173-174 of [78] and pp. 140-141 of [47]). Using the simplified version of this physics notation employed in FISST, we define the *functional derivatives* of a p.g.fl. $G[h]$ as:

$$\frac{\delta^0 G}{\delta \mathbf{x}^0}[h] = G[h], \qquad \frac{\delta G}{\delta \mathbf{x}}[h] = \frac{\partial G}{\partial \delta_{\mathbf{x}}}[h] \quad (38)$$

$$\frac{\delta^n G}{\delta \mathbf{x}_1 \cdots \delta \mathbf{x}_n}[h] = \frac{\partial^n G}{\partial \delta_{\mathbf{x}_1} \cdots \partial \delta_{\mathbf{x}_n}}[h] \quad (39)$$

It can be shown (see p. 1162 of [48]) that the set derivative of $\beta_\Xi(S)$ is a functional derivative of $G_\Xi[\mu]$

$$\frac{\delta \beta_\Xi}{\delta \mathbf{x}}(S) = \frac{\partial G_\Xi}{\partial \delta_{\mathbf{x}}}[\mathbf{1}_S] \quad (40)$$

with $g = \delta_{\mathbf{x}}$ and $h = \mathbf{1}_S$: Likewise for the iterated derivatives:

$$\frac{\delta \beta_\Xi}{\delta X}(S) = \frac{\delta^n \beta_\Xi}{\delta \mathbf{x}_1 \cdots \delta \mathbf{x}_n}(S) = \frac{\partial^n G_\Xi}{\partial \delta_{\mathbf{x}_1} \cdots \partial \delta_{\mathbf{x}_n}}[\mathbf{1}_S] \quad (41)$$

for $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ with $\mathbf{x}_1,\ldots,\mathbf{x}_n$ distinct. So, the multitarget probability distribution of a random state-set Ξ is, for $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$:

$$f_\Xi(X) = \frac{\delta^n \beta_\Xi}{\delta \mathbf{x}_1 \cdots \delta \mathbf{x}_n}(\emptyset) = \frac{\delta^n G_\Xi}{\delta \mathbf{x}_1 \cdots \delta \mathbf{x}_n}[0]. \quad (42)$$

4.7 Principled approximation

The generalized FISST calculus outlined in the previous section provides the foundation for a systematic procedure for devising computational approximation strategies. This procedure has been used, for example, to derive the predictor and corrector equations for the PHD filter of section 5.3 [48]. A detailed description of this procedure is beyond the scope of this paper. However, generally speaking it consists of the following steps:

• *Step 1*: Rewrite the multitarget predictor integral, Eq. (29), in p.g.fl. form:

$$G_{k+1|k}[h] = \int G_{k+1|k}[h \mid X] \cdot f_{k|k}(X \mid Z^{(k)}) \delta X \quad (43)$$

where

$$G_{k+1|k}[h \mid X] = \int h^Y \cdot G_{k+1|k}(Y \mid X) \delta Y \quad (44)$$

where h^Y is as defined in Eq. (35).

• *Step 2*: Given a multitarget Markov density based on a specific multitarget motion model as in Eq. (8), derive a formula of the form $G_{k+1|k}[h] = G_{k|k}[\Phi[h]]$ for some functional transformation $h \to \Phi[h]$. This formula can then be used to derive approximate prediction equations—e.g., for the predicted first-order multitarget moment (or PHD, see section 5.3):

$$D_{k+1|k}(\mathbf{x} \mid Z^{(k)}) = \frac{\delta G_{k+1|k}}{\delta \mathbf{x}}[1] \quad (45)$$

• *Step 3*: Rewrite the numerator of multitarget Bayes' rule, Eq. (30), as a p.g.fl.:

$$F_{k+1}[g,h] = \int h^X \cdot G_{k+1}[g \mid X] f_{k+1|k}(X \mid Z^{(k)}) \delta X \quad (46)$$

where

$$G_{k+1}[g \mid X] = \int g^Z \cdot f_{k+1}(Z \mid X) \delta Z \quad (47)$$

• *Step 4*: Rewrite the multitarget Bayes' rule, Eq. (30), in terms of p.g.fl.'s and their functional derivatives:

$$G_{k+1|k+1}[h] = \frac{\dfrac{\delta F_{k+1}}{\delta Z_{m+1}}[0,h]}{\dfrac{\delta F_{k+1}}{\delta Z_{m+1}}[0,1]} \quad (48)$$

• *Step 5*: Assume that the predicted p.g.fl. $G_{k+1|k}[h]$ has a suitably simplified form such as

$$G_{k+1|k}[h] = \exp\left(-\lambda + \lambda \int h(\mathbf{x}) s(\mathbf{x}) d\mathbf{x}\right) \quad (49)$$

(the Poisson approximation) or

$$G_{k+1|k}[h] = \prod_{j=1}^{n} \left(1 - q_j + q_j \int h(\mathbf{x}) f_j(\mathbf{x}) d\mathbf{x}\right) \quad (50)$$

(the multi-hypothesis correlator approximation).

• *Step 6*: Using a multitarget likelihood function constructed from a specific measurement model as in Eq. (7), derive the updated first-order moment (the PHD):

$$D_{k+1|k+1}(\mathbf{x} \mid Z^{(k+1)}) = \frac{\delta G_{k+1|k+1}}{\delta \mathbf{x}}[1] \quad (51)$$

• *Step 7*: Suppose that we are given some objective function for use in sensor management, such as the posterior expected number of targets

$$N_{k+1|k+1} = \frac{\partial}{\partial y} G_{k+1|k+1}[e^y] = \int |X| \cdot f_{k+1|k+1}(X \mid Z^{(k+1)}) \delta X \quad (52)$$

Use the approximations of Step 5 to derive approximate formulas for the objective function.

4.8 FISST and formal probability theory

We elaborate on the statement made in section 3.3, that multitarget integrals and multitarget densities lack a familiar measure-theoretic foundation. Some readers may choose to skip this section and move on to section 5.

Begin by recalling the definition of a general set integral (see pp. 141-143 of [20]). Let O be a Borel-measurable subset of the Mathéron topology on the space of finite subsets X of state space. Let χ_n be the transformation from vectors $(\mathbf{x}_1,\ldots, \mathbf{x}_n)$ of length n into finite subsets of finite subsets of cardinality $\leq n$ defined by $\chi_n(\mathbf{x}_1,\ldots, \mathbf{x}_n) = \{\mathbf{x}_1,\ldots, \mathbf{x}_n\}$. Let $\chi_n^{-1}(\sigma_n \cap O)$ denote the subset of vectors $(\mathbf{x}_1,\ldots, \mathbf{x}_n)$ such that $\chi_n(\mathbf{x}_1,\ldots, \mathbf{x}_n) \in O \cap \sigma_n$ where σ_n denotes the set of finite state-sets of cardinality n. Then the set integral evaluated on O is

$$\int_O f(X)\delta X = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\chi_n^{-1}(\sigma_n \cap O)} f(\{\mathbf{x}_1,..., \mathbf{x}_n\}) d\mathbf{x}_1 \cdots d\mathbf{x}_n \quad (55)$$

Let $\mathbf{1}_O(X)$ be the set membership function of O: $\mathbf{1}_O(X) = 1$ if $X \in O$ and $\mathbf{1}_O(X) = 0$ otherwise. Note that $\mathbf{1}_O(X)$ has no units of measurement. If the set integral were an integral with respect to some measure μ then

$$\mu(O) = \int \mathbf{1}_O(X)\mu(dX) = \int \mathbf{1}_O(X)\delta X = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\chi_n^{-1}(\sigma_n \cap O)} \mathbf{1}_O(\{\mathbf{x}_1,..., \mathbf{x}_n\}) d\mathbf{x}_1 \cdots d\mathbf{x}_n \quad (56)$$

However, in general the infinite sum is undefined since each of its terms has different units of measurement. That is: the *set integral is not a measure-theoretic integral*.

The problem is that Lebesgue measure cannot be generalized to multitarget state space without introducing an arbitrary "fudge factor." Inspection of Eq. (56) reveals that the only simple generalization of Lebesgue measure is

$$\lambda_c(O) = \int_O \frac{1}{c^{|X|}} \delta X \quad (57)$$

where c is some nonnegative constant with the same units of measurement as $\mathbf{x}$. However, the "fudge factor" c not only can have arbitrary magnitude but also has no evident physical interpretation. Likewise, the integral

$$\int_O F(X)\lambda_c(dX) = \sum_{n=0}^{\infty} \frac{1}{n! \cdot c^{|X|}} \int_{\chi_n^{-1}(\sigma_n \cap O)} F(\{\mathbf{x}_1,..., \mathbf{x}_n\}) d\mathbf{x}_1 \cdots d\mathbf{x}_n \quad (58)$$

corresponding to the measure λ_c depends on the same fudge factor; and will be definable only if $F(X)$ has no units of measurement for all X.

Similarly, let $p_\Xi(O) = \Pr(\Xi \in O)$ be the probability measure of the random state-set Ξ. Then its Radon-Nikodým derivative $F_\Xi(X)$ is defined by

$$p_\Xi(O) = \int_O F_\Xi(X)\lambda_c(dX) = \int_O f_\Xi(X)\delta X \quad (59)$$

so that

$$F_\Xi(X) = c^{|X|} \cdot f_\Xi(X) \quad (60)$$

almost everywhere. Consequently, the usual measure-theoretic definition of the probability density function of the random state-set Ξ—as the Radon-Nikodým derivative of the probability measure of Ξ—also unavoidably depends on the arbitrary fudge factor c.

In summary: The concept of a set integral and of a FISST multitarget density function are required to avoid the arbitrariness introduced by usual measure-theoretic integrals and density functions. Stated differently: There are two ways of defining multitarget density functions that are not quite the same: as Frechét functional derivatives $f_\Xi(X)$ of $G_\Xi[h]$; or as Radon-Nikodým derivatives $F_\Xi(X)$ of $p_\Xi(O)$. Set integrals are antiderivatives of Frechét functional derivatives, not Radon-Nikodým derivatives.

5 FISST and selected applications

In this section we briefly illustrate the application of FISST techniques to a few selected practical applications. Most of this work has been conducted by a team consisting of Lockheed Martin MS2 Tactical Systems (LMTS) of Eagan, Minnesota, and Scientific Systems Company Inc. (SSCI) of Woburn, Massachusetts. We summarize the application of FISST to scientific performance assessment (section 5.1); robust automatic target recognition for synthetic aperture radar (section 5.2); the PHD approximate multitarget filter (section 5.3); group target detection and tracking (section 5.4); and sensor management (section 5.5).

5.1 Scientific performance evaluation

Performance metrics for information fusion tend to be cobbled together as *ad hoc* afterthoughts in the development of particular fusion algorithms. *Ad hoc* metrology leads to serious practical difficulties, however. The traditional approach is to use metrics which measure some particular ("local") aspect of algorithm competence, e.g. miss distance, track purity, etc. Such "local metrics" often produce more confusion than clarity. This is because information fusion algorithms are highly nonlinear systems whose parts interact with each other in often unpredictable ways. For example, optimization of an algorithm with respect to one particular local metric (e.g. target localization accuracy) will not infrequently result in algorithm degradation as measured by another local metric (e.g. target ID accuracy).

In response to this conundrum, many attempts have been made to devise *ad hoc* composite metrics—for example, weighted averages of local metrics. In practice, such metrics are arbitrary and difficult to interpret.

The LMTS/SSCI team has used FISST techniques to develop a systematic, scientifically based approach to performance assessment of multisensor-multitarget algorithms. This research has concentrated on the development of *system-level measures of performance*, and has taken two directions. The first has focused on *information-theoretic measures of performance* and the second on *multitarget miss distance*. We have devised efficient implementations of both approaches.

- *Information-based performance assessment.* This approach is based on the concept of measuring the over-all mathematical information produced by a multisensor-multitarget information fusion algorithm (see Chapter 8 of [20]). While it incorporates all aspects of algorithm performance into a single figure of merit, it can be tractably applied only to multi-hypothesis correlator (MHC)-type algorithms. The reason for this is as follows. The output of any information fusion algorithm is some multitarget distribution $f_k(X)$ that approximates the multitarget posterior $f_{k|k}(X|Z^{(k)})$ produced by the ideal multitarget Bayes filter. If the algorithm is of MHC type then it is possible to derive an approximate closed-form formula for $f_k(X)$.

When ground truth is known we define a multitarget probability density $g_k(X)$ that describes ground truth at time-step k. Then the degree to which algorithm output deviates from ground truth in an information-theoretic sense is measured by the following multitarget generalization of the Kullback-Leibler discrimination:

$$I(g_k; f_k) = \int g_k(X) \log\left(\frac{g_k(X)}{f_k(X)}\right)\delta X \cong \text{constant} - \log f_k(G_k) \quad (61)$$

where $G_k = \{\mathbf{g}_1,\ldots, \mathbf{g}_\gamma\}$ in this case represents the set of states of the actual targets at time-step k.

When ground truth is unknown, the best that we can do is measure the degree to which $f_k(X)$ avoids total confusion. Complete lack of information is described by a multitarget analog of the uniform distribution, $u(X)$. The degree to which an algorithm avoids total confusion is

$$I(f_k;u) = \int f_k(X)\log\left(\frac{f_k(X)}{u(X)}\right)\delta X \tag{62}$$

Relatively efficient algorithms have been devised for both approaches (though Eq. (61) is much more tractable). The approach has been generalized to include general Csiszár information-theoretic functionals [13]. It is also possible to estimate how much information is being provided by some specific sub-function—e.g., detection vs. tracking vs. identification.

The approach is also easily extended to measure the performance of sensor management algorithms [13]. One simply measures the performance of the underlying Level 1 fusion algorithm with and without sensor management. The difference provides a measure of the performance improvement attributable to sensor management.

• *Multitarget miss distance*. This approach can be used with any information fusion algorithm that produces a set of target state-estimates as its output. However, it is applicable only when ground truth is known and it is not as comprehensive a measure of information output as the true information-theoretic metrics just discussed.

Our goal is to measure the "miss distance" between multitarget ground truth $G = \{\mathbf{g}_1,\ldots, \mathbf{g}_\gamma\}$ and the multi-track output $X = \{\mathbf{x}_1,\ldots, \mathbf{x}_n\}$ of an algorithm at any given time-instant. Drummond et. al. have proposed the use of optimal-assignment algorithms to compute such measures [77]. The correct way to implement such an approach is obvious when $\gamma = n$. If otherwise the usual approach is to declare "false" or "missing" tracks and then delete truths and tracks to equalize numbers so that optimal assignment techniques can be applied. When targets are more closely spaced, this procedure becomes arbitrary and subjective. Moreover, it does not produce miss distances in a rigorous mathematical sense—with the consequence that they can exhibit subtle, pathological, counter-intuitive behavior.

We have shown how Drummond's approach can be generalized to a multitarget miss distance that is both mathematically sound and measures the things that one would intuitively hope to see measured [25]. It is based on *Wasserstein distance*, a metric widely used in theoretical statistics that has also attracted great interest for image signal processing applications.

5.2 Robust ATR for SAR

Synthetic Aperture Radar (SAR) data provides an example of data that is unambiguous in itself but may involve considerable ambiguity regarding the manner of its generation. This is because SAR images can be greatly distorted by effects such as wet mud, dents, turret articulations, placement of nonstandard equipment—or nonstandard placement of standard equipment—on the surfaces of targets, and so on. Most of these effects are due to real-world statistical variability, but are essentially impossible to model or statistically characterize.

The LMTS/SSCI team has applied the techniques described in section 4.4 to this problem with good preliminary results. See [26] for more details.

5.3 The PHD approximate multitarget filter

The PHD filter was devised to surmount the daunting combinatorial complexity of multitarget problems described in section 3.1. In certain applications it is unnecessary, or undesirable, to detect and track each and every target in the scenario. This is the case, for example, when one is interested only in large tactical formations such as battalions ("group targets"); when the sheer number of targets is too large to permit effective tracking using techniques such as MHT; or when one must track a dense formation of targets because a few of the targets may have high tactical priority ("cluster tracking").

The conventional approach to such problems would be to attempt to track individual targets and assemble the group from them. Such an approach is most infeasible in the densest parts of formations, where targets of interest are most likely but where confusion is greatest.

In 2000 we proposed a multitarget filtering approach based on the opposite strategy: first track only the over-all group behavior, and then attempt to detect and track individual targets only as the quantity and quality of data permits. This filter, the *probability hypothesis density* (PHD) *filter*, is based on the idea of propagating a *first-order multitarget moment statistic* of the multitarget posterior $f_{k|k}(X|Z^{(k)})$ in place of $f_{k|k}(X|Z^{(k)})$ itself. In this sense it is a statistical analog of the constant-gain Kalman filter (e.g., α-β-γ filters), which propagates the posterior expectation of the single-target posterior $f_{k|k}(X|Z^{(k)})$.

The question of how to define the expected value of a multitarget distribution $f_{k|k}(X|Z^{(k)})$ is an open research question (see section 7.2). This is because integrals of the form $\int X\cdot f_{k|k}(X|Z^{(k)})\delta X$ are undefined, since addition/subtraction of finite sets X are undefined. Consequently, one must define an embedding $X \rightarrow \mathbf{e}_X$ that transforms finite sets X into elements $\mathbf{e}_X$ of some vector space. This embedding should preserve set-theoretic additive structure: $\mathbf{e}_{X\cup Y} = \mathbf{e}_X + \mathbf{e}_Y$ if $X \cap Y = \emptyset$. In this case an *indirect* expected value $\int \mathbf{e}_X \cdot f_{k|k}(X|Z^{(k)})\delta X$ can be constructed.

In point process theory the common practice is to choose $\mathbf{e}_X = \delta_X = \Sigma_{\mathbf{x}\in X}\, \delta_{\mathbf{x}}$ where $\delta_{\mathbf{x}}$ is the Dirac delta function concentrated at $\mathbf{x}$. The expectation

$$\begin{aligned} D_{k|k}(\mathbf{x} \mid Z^{(k)}) &= \int \delta_X(\mathbf{x})\cdot f_{k|k}(X \mid Z^{(k)})\delta X \\ &= \int f_{k|k}(\{\mathbf{x}\}\cup W \mid Z^{(k)})\delta W \end{aligned} \tag{63}$$

is called the *first-moment density* or, as we call it, the *probability hypothesis density* (PHD). Its defining property is this: for any region S, $\int_S D_{k|k}(\mathbf{x}|Z^{(k)})d\mathbf{x}$ is the expected number of targets in S. Using section 4.7, one can derive predictor and corrector equations for a PHD-based multitarget filter by using the following fact:

$$D_{k|k}(\mathbf{x} \mid Z^{(k+1)}) = \frac{\delta G_{k|k}}{\delta \mathbf{x}}[1] \tag{64}$$

See [48] for more details.

Particle-system implementations of the PHD filter are being investigated by a number of researchers throughout the world. Instances include Lin, Kirubarajan, and Bar-Shalom [37]; Mahler and Zajic [61]; Panta, Vo, Doucet, and Singh [72]; Punithakumar and Kirubarajan [74]; Shoenfeld [79]; Sidenbladh [80]; Sidenbladh, Svenson, and Schubert [81]; Tobias and Lanterman [87]; and Vo, Singh, and Doucet [89].

The first of these papers [37] is somewhat surprising since it shows that, at least in simple 1-D simulations, a modified form of the PHD filter can outperform a conventional MHT-type multitarget tracker in lower-SNR environments. If this result holds true in more realistic scenarios, the PHD filter may prove useful in conventional multitarget tracking problems, not just the cluster-tracking applications for which it was intended.

5.4 Group-target filtering

Using FISST techniques, we have developed what appears to be the first (and so far only) rigorous, systematic foundation for group-target detection, tracking, and identification [42].

The basis of the approach is the following state representation for a multigroup system. The state of a single group is a pair $(\mathbf{g},X)$ where $\mathbf{g}$ is the state of the group and X is the (nonempty) set of states of the ordinary targets that belong to the group. (For example, $\mathbf{g}$ can have the form $\mathbf{g} = (\mathbf{x},\mathbf{v},N,\tau,\gamma)$ where (i) $\mathbf{x}$ is the geometric centroid and $\mathbf{v}$ is its velocity; (ii) N is the number of targets; (iii) τ is the type (e.g. armored cavalry division); and (iv) γ is a geometric-shape parameter (e.g. chevron, column, etc.).) Likewise, the state of a system consisting of several group targets is a finite set of pairs $\mathfrak{X} = \{(\mathbf{g}_1,X_1),\ldots, (\mathbf{g}_e,X_e)\}$. Given this, the optimal solution for the group target problem is the Bayes recursive filter that propagates multigroup posterior distributions of the form $f_{k|k}(\mathfrak{X}|Z^{(k)})$. This filter will be hopelessly intractable in general. We have proposed a version of the PHD filter that propagates a "group PHD" of the form $D_{k|k}(\mathbf{g},\mathbf{x}|Z^{(k)})$.

5.5 Control-theoretic sensor management

The daunting computational complexity of multisensor-multitarget sensor management was described in section 3.1. To address this compexity, we are developing a theoretically foundational but potentially tractable control-theoretic basis for multisensor-multitarget sensor management using a comprehensive, intuitive, system-level Bayesian paradigm [41, 44, 50, 57]. Some of this work is summarized in another paper presented at this conference [55]. This work currently encompasses the following aspects of multisensor-multitarget sensor management: (1) targets of current or potential tactical interest; (2) multistep look-ahead (control of sensor resources throughout a future time-window); (3) sensors with non-ideal dynamics, including sensors residing on moving platforms such as UAVs; (4) sensors whose states are observed indirectly by internal actuator sensors; and (5) possible communication drop-outs. Our approach also addresses a more subtle issue: *the impossibility of deciding between an infinitude of plausible objective functions*, by concentrating on "probabilistically natural" sensor management goals.

This work is based on a "natural" sensor management objective function, the *posterior expected number of targets* (PENT), that tends to maximize the number of well-localized targets. This objective function is constructed using a new optimization-hedging strategy, "maxi-PIMS," that tractably hedges against the unknowability of future observation-collections. (A previous optimization-hedging strategy, called "maxi-null," proved to be too conservative to yield good sensor management results.) The maxi-PIMS strategy optimizes the likelihood of collecting the *predicted ideal measurement-set* (PIMS). Intuitively speaking, in a PIMS there are no false alarm/clutter observations, every target in the FoV generates an observation, and target-generated observations are noise-free.

The PENT objective function is used in conjunction with one of two approximate multitarget filters: the probability hypothesis density (PHD) filter or the multi-hypothesis correlator (MHC) filter. Preliminary simulations using PENT with an MHC filter has demonstrated good sensor management behavior [14].

6 Imitator-critics of FISST

As we noted in section 1.6, some have responded to FISST by (1) advocating more complicated formulations of point process theory supposedly more general than FISST; or, alternatively, (2) promoting a so-called "plain-vanilla Bayesian approach" as a great advance over FISST because it avoids point process theory altogether. We address these imitator-critics in turn: point process theory and its relationship with FISST in section 6.1; and the "plain-vanilla Bayesian approach" in section 6.2.

6.1 Point process theory and FISST

Point process theory is the mathematical theory of stochastic multi-object systems [9, 32, 83, 86]. Intuitively speaking, a point process is a *random finite multiset*. A finite multiset (sometimes also called a "bag") is a finite unordered list $L = «\mathbf{y}_1,\ldots, \mathbf{y}_1,\ldots, \mathbf{y}_n,\ldots, \mathbf{y}_n»$ of elements of some space $\boldsymbol{Y}$ in which repetition of elements is allowed: the $\mathbf{y}_1,\ldots,\mathbf{y}_n$ are distinct and there are ν_1 copies of $\mathbf{y}_1$, ν_2 copies of $\mathbf{y}_2$, and so on.

Multisets can be thought of as generalizations of fuzzy sets in which elements have multiple rather than partial membership [2]. That is, they can be represented as membership functions $\mu(\mathbf{y})$ on $\boldsymbol{Y}$ whose values are nonnegative integers: $\mu(\mathbf{y}_i) = \nu_i$ for all $i = 1,\ldots,n$ and $\mu(\mathbf{y}) = 0$ otherwise. Intersection of multisets is defined as $(\mu_1\wedge\mu_2)(\mathbf{y}) = \min\{\mu_1(\mathbf{y}), \mu_2(\mathbf{y})\}$. Union is defined as $(\mu_1\vee\mu_2)(\mathbf{y}) = \max\{\mu_1(\mathbf{y}), \mu_2(\mathbf{y})\}$. Complementation of multisets cannot be meaningfully defined at all.

Multisets are not the only possible formulation of the concept of multiplicity. Other mathematically equivalent representations are as follows (where in each case the $\mathbf{y}_1,\ldots,\mathbf{y}_n$ are distinct):

1) *Dirac sum*:

$$\delta(\mathbf{y}) = \nu_1\delta_{\mathbf{y}_1}(\mathbf{y}) + \ldots + \nu_n\delta_{\mathbf{y}_n}(\mathbf{y}) \tag{65}$$

2) *Counting measure*:

$$\Delta(S) = \int_S \delta(\mathbf{y})d\mathbf{y} = \nu_1\Delta_{\mathbf{y}_1}(S) + ... + \nu_n\Delta_{\mathbf{y}_n}(S) \quad (66)$$

3) *Simple counting measure on pairs*:

$$\Delta(S) = \Delta_{(\nu_1,\mathbf{y}_1)}(S) + ... + \Delta_{(\nu_1,\mathbf{y}_n)}(S) \quad (67)$$

4) *Simple Dirac sum on pairs*:

$$\delta(\nu,\mathbf{y}) = \delta_{(\nu_1,\mathbf{y}_1)}(\nu,\mathbf{y}) + ... + \delta_{(\nu_1,\mathbf{y}_n)}(\nu,\mathbf{y}) \quad (68)$$

5) *Finite set of pairs*:

$$T = \{(\nu_1,\mathbf{y}_1),...,(\nu_n,\mathbf{y}_n)\} \quad (69)$$

and so on, where $\Delta_{\mathbf{y}}(S)$ denotes the Dirac measure: $\Delta_{\mathbf{y}}(S) = 1$ if $\mathbf{y} \in S$ and $\Delta_{\mathbf{y}}(S) = 0$ otherwise.

If $\nu_i = 1$ for all $i = 1,..,n$ then Eqs. (65)-(69) are all equivalent representations of a finite subset of $\boldsymbol{Y}$. In particular, Eq. (69) shows that multiplicity can be formulated in terms of ordinary set theory—and thus that any point process can be formulated as a certain type of random finite set.[4] In particular, let Ψ be a finite random subset of $\boldsymbol{Y}$. Then

$$N_\Psi(S) = |\Psi \cap S|, \qquad \delta_\Psi(\mathbf{y}) = \sum_{\mathbf{w}\in\Psi}\delta_{\mathbf{w}}(\mathbf{y})$$

are equivalent mathematical representations of Ψ (also known as a simple point process, see pp. 100-102 of [86]).

From an engineering point of view, Eqs. (65)-(68) are all essentially just changes of notation. But they increase notational obfuscation while adding no new substance. Moreover, they lose the simple, intuitive tools of ordinary set theory, while adding the mathematical complexity necessary to rigorously deal with them.

For example, represent multitarget states as Dirac sums $\delta_X = \delta_{\mathbf{x}_1} + ... + \delta_{\mathbf{x}_n}$. Then a multitarget probability distribution must be a functional $F[\delta]$—i.e., a function whose argument δ is an ordinary function. If $F[\delta]$ is a probability distribution then we must define a *functional integral* $\int F[\delta]\mu(d\delta)$. The theory of functional integrals is, in general, very difficult. However, in this case functions have the form $\delta = \delta_{\mathbf{x}_1} + ... + \delta_{\mathbf{x}_n}$ and probability distributions the form

$$F[\delta] = F[\delta_{\mathbf{x}_1} + ... + \delta_{\mathbf{x}_n}]\,. \quad (70)$$

So, functional integrals are also simpler. Since δ is a density and differential probability mass $F[\delta]d\delta$ must be unitless, $F[\delta]$ must have the same units of measurement as $\mathbf{x}$. If the functional integral is to produce intuitively desired results then it must have the form

$$\int_S F[\delta]D\delta = \sum_{n=0}^{\infty}\frac{1}{n!}\int_{S^n} F[\delta_{\mathbf{x}_1} + ... + \delta_{\mathbf{x}_n}]d\mathbf{x}_1 \cdots d\mathbf{x}_n \quad (71)$$

In other words, $\int F[\delta]D\delta$ is just the set integral of Eq. (11) written in highly obfuscated form.

In summary: those who claim that more complicated point process formulations such as Eqs. (65)-(68) produce a more general theory than FISST, must answer the following question: Why aren't they promoting mere changes of notation that add no new substance?

[4] Matters are not as straightforward as this discussion implies. One must define topologies on the spaces of objects that one wants to randomize. To say that two point process formulations are equivalent, one must show that they are topologically equivalent. Eq. (66) and Eq. (67) are equivalent because of pp. 5, 16 of [32]. Eq. (68) and (69) are equivalent because of [76] and [5].

6.2 The "plain-vanilla Bayesian approach"

As noted earlier, a few anonymous partisans have claimed that a so-called "plain-vanilla Bayesian approach" suffices as down-to-earth, general "first principles" for Bayes multitarget filtering. *Ergo*, FISST is unnecessary or worse. In two recent papers [38, 43] we have demonstrated the shallowness of their criticisms and the speciousness of the "plain-vanilla Bayesian approach" itself. We summarize these papers in this section.

- The "plain-vanilla Bayesian approach" is so heedlessly formulated that it is not even Bayesian—and moreover, disparages random set concepts even while unwittingly assuming them. For example, one such partisan has: (1) boasted that his "plain-vanilla" approach addresses multitarget Bayes filtering with all necessary rigor and generality while being "straightforward"; (2) asserted that, therefore, FISST is pointless "obfuscation"; (3) subsequently introduced a complicated ur-theory for Bayes multitarget filtering based on the Dirac sum representation of a point process, Eq. (65); (4) failed to notice that this should be unnecessary since—as per his boast—the "plain-vanilla" approach already covers all of the bases; and (5) failed to comprehend that his ur-theory just unwittingly re-invents basic random set concepts in highly "obfuscated" notation—see Eqs. (70), (71). How can this be said to meet even minimal standards of logic?
- Also, note the double standards being applied. Anything more complicated than "plain-vanilla" is "obfuscation"—except when a "plain-vanilla" partisan is the obfuscator. Likewise: If FISST is illustrated using familiar observation models (e.g., post-detection reports), this proves that FISST is not "general." But explicit, general, rigorous methods for observation models (and many other things) prove only that FISST is not "simple." Whereas the "plain-vanilla" approach is simple *because* it has no such methods—but yet, magically, is elastically all-subsuming. Similarly: A partisan avers that his approach addresses "the problem of search, track and identification, with the confounding issue that target count is unknown and must be estimated too"—whereas FISST addresses something else: unified expert-systems theory. But FISST addresses both. When did misrepresentation and puffery come to suffice as "first principles"?
- In asserting no significant difference between single-target and multitarget Bayes statistics, such partisans also fail to account for the actual, major differences—most seriously, by *erroneously presuming that the naïve multitarget generalizations of the single-target Bayes-optimal state estimators exist* (section 4.5). What is the credibility of "plain-vanilla Bayesian" when one of its central decision procedures is "not invariant under even a change of units"—especially given that these are the words used by one partisan in criticizing the same type of error in work that preceded his own?
- The "plain-vanilla" stance essentially repudiates the formal statistical modeling standard that FISST directly extends to multisensor-multitarget problems. But "plain-vanilla" implementation has produced a succession of *ad hoc*, brute force algorithms afflicted by inherent—but less than candidly acknowledged—computational "logjams"

(see section 3.1 and [43]). When did "logjams" become exemplars of down-to-earth engineering practicality?

7 Possible directions for the future

We divide this discussion into three parts: applications of FISST (section 7.1), unresolved theoretical questions (section 7.2), and "blue sky" issues (section 7.3).

7.1 Applications

- Further development of particle filter implementations of the PHD filter and general Bayes filter.
- Further refinement of control-theoretic sensor management.
- Investigation of point process computational techniques being currently used in, e.g., medical imaging.
- Extension of FISST to Electronic Counter-Countermeasures (ECCM), as proposed on p. 19 of [45].
- Extension of FISST to include communication network effects, as initiated in [50].
- Extension of FISST to account for sensor spatial and temporal biases.
- Extension of FISST to distributed fusion.
- Extension to FISST to threat assessment, perhaps along the lines of [7] and/or [57].

7.2 Unresolved theoretical questions

- What is the relationship between the multi-hypothesis correlator-tracker and the general multitarget Bayes filter?
- Can the PHD filter be generalized so that a Poisson approximation is unnecessary for the corrector step?
- Interacting multitarget motion models (e.g., interacting-particle methods) for group-target filtering, to model coordinated multitarget motion.
- Can the PHD filter be *tractably* generalized to produce a *second-order* multitarget filter along the lines of [47]?
- The multitarget filter is computationally daunting in part because one must deal with a mixed discrete (target number) and continuous (target states) system. In [51] we proposed a method for making target number a continuous variable. Can this approach be refined with the aim of producing more computationally efficient approaches to general multitarget filtering?
- Can E. Kamen's symmetric measurement equation (SME) filter be rigorously and systematically formulated along the lines suggested on p. 179 of [20]?

7.3 Blue sky issues

- What is the expected value of a random finite set?
- Fully integrate possibility (Sugeno) measures by extending random set theory to random filter theory, along the lines of [34].
- Full rigorous foundation for possibility theory as a subdiscipline of real analysis, along the lines of [23].

8 Errata

Given the unusual breadth of FISST, it is inevitable that errors or oversights will occur. Over the last decade we have tried to promptly alert readers about these. In this section we list all errors that we know, with references to publications in which they are described in more detail.

- Garbled bibliographic references in Chapter 2 of [20]: see p. viii of [45].
- Erroneous statement about multitarget estimators in [59]: see p. 6 of [45].
- Erroneous statement about upper/lower probabilities in [20]: see pp. 59-60 of [45].
- Erroneous "user-defined information," second equation from bottom, p. 311 of [20]: see p. 302 of Hoffman, Mahler, Zajic, "User-defined information and scientific performance evaluation," SPIE Vol. 4380, 2001.
- Error in the PHD corrector equation in early PHD filter papers: see footnote 17, p. 1168, of [48].
- Erroneous "pseudo-sensor" approximation for multisensor PHD corrector equation in early PHD filter papers: see p. 1169 of [48];
- Major typo in Eqs. (10), (106) of [48]: should be

$$D_{k+1|k+1}(\mathbf{x}) \cong F_{k+1}^{[1]}(Z_{k+1}^{[1]} \mid \mathbf{x}) \cdots F_{k+1}^{[s]}(Z_{k+1}^{[s]} \mid \mathbf{x}) D_{k+1|k}(\mathbf{x})$$

Also, note that this is a much more drastic approximation than simply applying the PHD corrector successively using the observations produced by different sensors.

- Failure of "maxi-null" optimization-hedging strategy for sensor management in 2003 papers: see section 4 of [55], section 3 of [57], or sections 1.1, 3.1 of [50].
- The careless operations count in sections 1.1, 1.2 of [43] has been corrected in section 3.1 above.

9 Conclusions

In this invited keynote paper we have provided a comprehensive, high-level assessment of FISST: of its antecedents; its beginnings and motivations; its techniques, tools, conceptual evolution, and current state of the art; its applications; its imitator-critics; and its possible future directions. For those already or potentially interested in learning more about FISST, we hope that this paper has provided a useful road map for navigating through FISST-related publicationss and techniques. The best entry points into FISST are the tutorial [56] and the technical monograph [45] or its condensed version [52].

Acknowledgements

The basic research reported in this paper was supported by Lockheed Martin in 1993-1994; and since then by the U.S. Army Research Office (ARO) under contracts DAAH04-94-C-0011 and DAAG55-98-C-0039; and the U.S. Air Force Office of Scientific Research (AFOSR) under contract F49620-01-C-0031. The content does not necessarily reflect the position or policy of the U.S. Government. No official endorsement should be inferred.

The author thanks the following individuals for their help, encouragement and/or support during the past decade: Dr. I.R. Goodman of SPAWAR Systems Center; Dr. William Sander of ARO; Stanton Musick of AFRL; Craig Poling, Charles Mills, and Drs. John Hoffman and Tim Zajic of Lockheed Martin; and Drs. Ravi Ravichandran, Adel El-Fallah, and Raman Mehra of Scientific Systems Co., Inc.

We also gratefully acknowledge the thought-provoking correspondence of Professor Ba-Ngu Vo of the University of Melbourne (Australia), e.g. regarding the issues raised in section 4.8; and leading to Eq. (43) of [50].

References

[1] C. Agate and K. Sullivan. Particle filtering algorithm for tracking multiple road-constrained targets. In I. Kadar, ed., *Sig. Proc. Sensor Fusion, and Targ. Recog. XII*, 256-266, SPIE (Vol. 5096), Bellingham WA, 2003.

[2] J. Albert. Algebraic properties of bag data types. In *Proc. 17th Int'l Conf. on Very Large Data Bases*, 211-219, Barcelona, September, 1991.

[3] D. Ballantyne, H. Chan, and M. Kouritzin. A branching particle-based nonlinear filter for multi-target tracking. In *Proc. 4th Int'l Conf. on Inf. Fusion*, Montreal, Aug. 7-10 2001, Int. Soc. Inf. Fusion, Sunnyvale, CA, 2001.

[4] Y. Bar-Shalom and X.-R. Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Ann Arbor, MI, 1993.

[5] M. Baudin. Multidimensional Point Processes and Random Closed Sets. *J. Applied Prob.*, (21): 173-178, 1984.

[6] R. Bethel and G. Paras. A PDF multitarget-tracker. *IEEE Trans AES*, 30(2): 386-403, 1994.

[7] D. Brown and H. Liu. A point process transition density model for threat assessment. In *Proc. 2001 MSS Nat'l Symp. on Sensor and Data Fusion*, Vol. I, San Diego, June 25-28, 2001.

[8] S. Challa, B.-N. Vo, and X. Wang. Bayesian approaches to track existence—IPDA and random sets. In *Proc. 5th Int'l Conf. on Information Fusion*, Vol. II, 1228-1235, Annapolis MD, July 8-11, 2002, Int. Soc. Inf. Fusion, Sunnyvale CA, 2002.

[9] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York, 1988.

[10] A. Doucet, N. de Freitas, and N. Gordon, eds. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[11] A. Doucet, B. Vo, C. Andrieu, and M. Davy. Particle filtering for multi-target tracking and sensor management. In *Proc. 5th Int'l Conf. on Inf. Fusion*, Annapolis MD, July 8-12, 2002, Int. Soc. Inf. Fusion, Sunnyvale CA, 2002.

[12] D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42: 87-101, 1991.

[13] A. El-Fallah, R. Mahler, T. Zajic, E. Sorensen, M. Alford, and R. Mehra. Scientific performance evaluation for sensor management. In I. Kadar, ed., *Sig. Proc., Sensor Fusion, and Targ. Recog. IX*, 183-194, SPIE (Vol. 4052), Bellingham WA, 2000.

[14] A. El-Fallah, M. Perloff, B. Ravichandran, R.Mehra, R. Mahler, T. Zajic, and C. Stelzig. Multitarget sensor management with target preference. In I. Kadar, ed., *Sig. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[15] N. Everett, S. Tham, and D. Salmond. A novel MCMC tracker for stressing scenarios. In O. Drummond, ed., *Signal and Data Processing of Small Targets 2004*, SPIE (Vol. 5428), Bellingham WA, 2004.

[16] L. Fisher. A survey of the mathematical theory of multidimensional point processes. In P. Lewis, ed., *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*. John Wiley & Sons, New York, 1972.

[17] D. Fixsen and R. Mahler. The modified Dempster-Shafer approach to classification. *IEEE Trans. SMC-Part A*, 27(1): 96-104, 1997.

[18] I. Goodman. Fuzzy sets as equivalence classes of random sets. In R. Yager, ed., *Fuzzy Sets and Possibility Theory: Recent Developments*, 327-343, Permagon Press, New York, 1982.

[19] I. Goodman and H. Nguyen. *Uncertainty Models for Knowledge Based Systems*. North-Holland, Amsterdam, 1985.

[20] I. Goodman, R. Mahler, and H. Nguyen. *Mathematics of Data Fusion*. Kluwer Academic Publishers, Dordrecht, 1997.

[21] J. Goutsias, R. Mahler, and H. Nguyen, eds. *Random Sets: Theory and Applications*. Springer-Verlag, New York, 1997.

[22] M. Grabisch, H. Nguyen, and E. Walker. *Fundamentals of Uncertainty Calculi With Applications to Fuzzy Inference*. Kluwer Academic Publishers, Dordrecht, 1995.

[23] J. Graf. A Radon-Nikodým theorem for capacities. *J. für Reine und Angewandte Mathematik*, (320): 192-214, 1980

[24] K. Hestir, H. Nguyen, and G. Rogers. A random set formalism for evidential reasoning. In I. Goodman, M. Gupta, H. Nguyen and G. Rogers, eds., *Conditional Logic in Expert Systems*, 309-344, North-Holland, Amsterdam, 1991.

[25] J. Hoffman and R. Mahler. Multitarget miss distance via optimal assignment. *IEEE Trans. Sys., Man, and Cybernetics—Part A*, 34(3): 327-336, 2004.

[26] J. Hoffman, R. Mahler, R. Ravichandran, R. Mehra, and S. Musick. Robust SAR ATR via Set-Valued Classifiers: New results. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XII*, 139-150, SPIE (Vol. 5096)), Bellingham WA, 2003.

[27] A. Höhle. A mathematical theory of uncertainty: fuzzy experiments and their realizations. In R.R. Yager, ed., *Recent Developments in Fuzzy Set and Possibility Theory*, 344-355, Permagon Press, New York, 1981.

[28] P. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

[29] C. Hue, J.-P. Le Cadre, and P. Peréz. Sequential Monte Carlo methods for multiple target tracking and data fusion. *IEEE Trans. Sig. Proc.*, 50(2): 309-325, 2002.

[30] M. Isard and J. MacCormick. BraMBLe: A Bayesian multiple-blob tracker. In *Proc. 8th IEEE*

Int'l Conf. on Comp. Vision, 2: 34-41, Vancouver, July 9-12, 2001.

[31] K. Kastella. A microdensity approach to multitarget tracking. In *Proc. 4th Int'l Conf. on Inf. Fusion*, Paris, 2000, Int. Soc. Inf. Fusion, Sunnyvale CA, 2000.

[32] A.F. Karr. *Point Processes and Their Statistical Inference*, Second Edition. Marcel Dekker, New York, 1991.

[33] Y. Kharin. *Robustness in Statistical Pattern Recognition*. Kluwer Academic Publishers, Dordrecht, 1996.

[34] J. Kohlas. Support and plausibility functions induced by filter-valued mapping. *Int'l J. of General Systems*, (21): 343-363, 1993.

[35] R. Kruse, E. Schwencke, and J. Heinsohn. *Uncertainty and Vagueness in Knowledge-Based Systems*. Springer-Verlag, New York, 1991.

[36] A.D. Lanterman, M.I. Miller, D.L. Snyder, and W.J. Miceli. Jump-diffusion processes for the automated understanding of FLIR scenes. In F.A. Sadjadi, ed., *Automatic Target Recognition IV*, 416-427, SPIE (Vol. 2234), Bellingham WA, 1994.

[37] L. Lin, T. Kirubarajan, and Y. Bar-Shalom. Data association combined with the probability hypothesis density filter for multitarget tracking. In O. Drummond, ed., *Sign. and Data Proc. of Small Targets 2004*, SPIE (Vol. 5428), Bellingham WA, 2004.

[38] R. Mahler. Bayesian vs. "plain-vanilla Bayesian" multitarget statistics. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[39] R. Mahler. Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, I: Boolean Logic. *IEEE Trans. SMC—Part A*, 26: 27-41, 1996.

[40] R. Mahler. Combining Ambiguous Evidence With Respect to Ambiguous *a priori* Knowledge, II: Fuzzy Logic. *Fuzzy Sets and Systems*, 75: 319-354, 1995.

[41] R. Mahler. Global posterior densities for sensor management. In M.K. Kasten and L.A. Stockum, eds., *Acquisition, Tracking, and Pointing XII*, 252-263, SPIE (Vol. 3365), Bellingham WA, 1998.

[42] R. Mahler. An extended first-order Bayes filter for force aggregation. In O. Drummond, ed., *Sign. and Data Proc. of Small Targets 2002*, 196-207, SPIE (Vol. 4729), Bellingham WA, 2002.

[43] R. Mahler. Formal vs. heuristic modeling for multitarget Bayesian filtering. In O. Drummond, ed., *Sign. and Data Proc. of Small Targets 2004*, SPIE (Vol. 5428), Bellingham WA, 2004.

[44] R. Mahler. Global optimal sensor allocation. In *Proc. Ninth Nat'l Symp. on Sensor Fusion*, Vol. I, 347-366, Naval Postgraduate School, Monterey CA, Mar. 12-14, 1996.

[45] R. Mahler. An Introduction to Multisource-Multitarget Statistics and Its Applications. Technical Monograph. Lockheed Martin, Eagan MN, 2000.

[46] R. Mahler. Measurement models for ambiguous evidence using conditional random sets. In I. Kadar, ed., *Sig. Proc., Sensor Fusion, and Targ. Recog. VI.*, 40-51, SPIE (Vol. 3068,), Bellingham WA, 1997.

[47] R. Mahler. Multitarget moments and their application to multitarget tracking. In *Proc. Workshop on Estimation, Tracking, and Fusion: A Tribute to Y. Bar-Shalom*, 134-166, Naval Postgraduate School, Monterey CA, May 17, 2001.

[48] R. Mahler. Multitarget filtering via first-order multitarget moments. *IEEE Trans. AES*, 39(4): 1152-1178, 2003.

[49] R. Mahler. Multitarget motion models. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. VIII*, 47-57, SPIE (Vol. 3720), Bellingham WA, 1999.

[50] R. Mahler. Multitarget sensor management of dispersed mobile sensors. In D. Grundel, R. Murphey, and P. Paralos, eds., *Theory and Algorithms for Cooperative Systems*, World Scientific, Singapore, 2005.

[51] R. Mahler. Point target clusters and continuous-state multitarget statistics. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XI*, 163-174, SPIE (Vol. 4729), Bellingham WA, 2002.

[52] R. Mahler. Random set theory for target tracking and identification. In D.L. Hall and J. Llinas, eds., *Handbook of Multisensor Data Fusion*, CRC Press, Boca Raton FL, 2002.

[53] R. Mahler. Representing rules as random sets, I: Statistical correlations between rules. *Information Sciences*, 88: 47-68, 1996.

[54] R. Mahler. Representing rules as random sets, II: Iterated rules. *Int'l Jour. Intell. Sys.*, 11: 583-610, 1996.

[55] R. Mahler. Sensor management with non-ideal sensor dynamics. In *Proc. 7th Int'l Conf. on Inf. Fusion*, Stockholm, June 28-July 1, 2004.

[56] R. Mahler. "Statistics 101" for multisensor, multitarget data fusion. *IEEE AES Mag., Part 2: Tutorials*, 19(1): 53-64, 2004.

[57] R. Mahler. Target preference in multitarget sensor management: A unified approach. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[58] R. Mahler. Unified Bayes multitarget fusion of ambiguous data sources. In *Proc.IEEE Int'l Conf. on Integration of Knowledge Intensive Multi-Agent Systems*, 343-348, Cambridge MA, Sept. 30-Oct. 4, 2003.

[59] R. Mahler. A unified foundation for data fusion. In F.A. Sadjadi, ed., *Selected Papers on Sensor and Data Fusion*, 325-354, SPIE (Vol. MS-124), Bellingham WA. Reprinted from *Proc. 7th Joint Service Data Fusion Symp.*, 1: 153-173, Johns Hopkins Applied Research Laboratory, Laurel, Maryland, Oct. 25-28 1994.

[60] R. Mahler, P. Leavitt, J. Warner, and R. Myre. Nonlinear filtering with really bad data. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. VIII*, 59-70, SPIE (Vol. 3720), Bellingham WA, 1999.

[61] R. Mahler and T. Zajic. Bulk multitarget tracking using a first-order multitarget tracking filter. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XI*, 175-186, SPIE (Vol. 4729), Bellingham WA, 2002.

[62] G. Mathéron. *Random Sets and Integral Geometry*. J. Wiley, New York, 1975.

[63] M. Moreland and S. Challa. A multi-target tracking algorithm based on random sets. In *Proc. 6th Int'l. Conf. on Inf. Fusion*, Cairns, Australia, July 8-11 2003, Int. Soc. Inf. Fusion, Sunnyvale, CA, 2003.

[64] S. Mori, C.-Y. Chong, E. Tse, and R.P. Wishner. Tracking and classifying multiple targets without A priori identification. *IEEE Trans. Auto. Contr*., 31(5): 401-409, 1986.

[65] J.E. Moyal. The general theory of stochastic population processes. *Acta Mathematica,* 108: 1-31, 1962.

[66] S. Musick, K. Kastella, and R. Mahler. A practical implementation of joint multitarget probabilities. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. VII*, 26-37, SPIE (Vol. 3374), Bellingham WA, 1998.

[67] J. Naylor and A. Smith. Application of a method for the efficient computation of posterior distributions. *Applied Statistics*, 31(3): 214-225, 1982.

[68] H.T. Nguyen. On random sets and belief functions. *J. Math. Anal. and Appl*., 65: 531-542, 1978.

[69] A.I. Orlov. Relationships between fuzzy and random sets: fuzzy tolerances. *Issledovania po Veroyatnostnostatishesk. Modelirovaniu Realnikh System*, Moscow, 1977.

[70] A.I. Orlov. Fuzzy and random sets. *Prikladnoi Mnogomerni Statisticheskii Analys*, 262-280, Adademiya Nauk CCCP, Moscow, 1978.

[71] M. Orton and W. Fitzgerald. A Bayesian approach to tracking multiple targets using sensor arrays and particle filters. *IEEE Trans. Sign. Proc.*, 50(2): 216-223, 2002.

[72] K. Panta, B. Vo, A. Doucet, and S. Singh. Probability hypothesis density filter versus multiple hypothesis testing. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[73] N. Portenko, H. Salehi, and A. Skorokhod. On optimal filtering of multitarget tracking systems based on point processes observations. *Random Operators and Stochastic Equations*, 1: 1-34, 1997.

[74] K. Punithakumar and T. Kirubarajan. A multiple model probability hypothesis density filter for tracking maneuvering targets. In O. Drummond, ed., *Sign. and Data Proc. of Small Targets 2004*, SPIE (Vol. 5428), Bellingham WA, 2004.

[75] P. Quinio and T. Matsuyama. Random closed sets: A unified approach to the representation of imprecision and uncertainty. In R. Kruse and P. Siegel, eds., *Symbolic and Quantitative Approaches to Uncertainty*, Springer-Verlag, New York, 1991.

[76] B. Ripley. Locally finite random sets: foundations for point process theory. *Annals of Prob*., 4(6): 983-994, 1976.

[77] R. Rothrock and O. Drummond. Performance metrics for multiple-sensor, multiple-target tracking. In O. Drummond, ed., *Sig. and Data Proc. of Small Targets 2000,* 521-531, SPIE (Vol. 4048), Bellingham WA, 2000.

[78] L. Ryder. *Quantum Field Theory*, 2nd Edition. Cambridge U. Press, Cambridge UK, 1996.

[79] P. Shoenfeld. Particle filter algorithm for the multitarget probability hypothesis density. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[80] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. In *Proc. 6th Int'l Conf. on Information Fusion*, Cairns Australia, July 8-11, 2003, Int. Soc. Inf. Fusion, Sunnyvale CA, 2003.

[81] H. Sidenbladh, P. Svenson, and J. Schubert. Comparing multitarget trackers on difference force level units. In I. Kadar, ed., *Sig. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[82] H. Sidenbladh and S.-L. Wirkander. Tracking Random Sets of Vehicles in Terrain. In *Proc. 2003 IEEE Workshop on Multi-Object Tracking*, Madison WI, June 21, 2003.

[83] D.L. Snyder and M.I. Miller. *Random Point Processes in Time and Space*, Second Edition. Springer, New York, 1991.

[84] E. Sorensen, T. Brundage, and R. Mahler. None-of-the-above (NOTA) capability for INTELL-based NCTI. In I. Kadar, ed., *Sig. Proc., Sensor Fusion, and Targ. Recog. X*, 281-287, SPIE (Vol. 4380), Bellingham WA, 2001.

[85] L. Stone, C. Barlow, and T. Corwin. *Bayesian Multiple Target Tracking*. Artech House, Ann Arbor MI, 1999.

[86] D. Stoyan, W.S. Kendall, and J. Meche. *Stochastic Geometry and Its Applications*, Second Edition. John Wiley & Sons, 1995.

[87] M. Tobias and A. Lanterman. Multitarget tracking using multiple bistatic range measurements with probability hypothesis densities. In I. Kadar, ed., *Sign. Proc., Sensor Fusion, and Targ. Recog. XIII*, SPIE (Vol. 5429), Bellingham WA, 2004.

[88] H. van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, 1968.

[89] B. Vo, S. Singh and A. Doucet. Sequential implementation of the PHD filter for multi-target tracking. In *Proc. 6th Int'l. Conf. on Inf. Fusion*, Cairns, Australia, July 8-11 2003, Int. Soc. Inf. Fusion, Sunnyvale, CA, 2003.

[90] R.B. Washburn. A Random Point Process Approach to Multi-Object Tracking. In *Proc. Amer. Control Conf.*, 3: 1846-1852, Minneapolis, June 10-12, 1987.