Unified Framework for Information Fusion

Dr. P. Valin
Senior Member of Engineering R&D
Lockheed Martin Canada
6111 Royalmount Avenue
Montréal, QC H4P 1K6
pierre.valin@lmco.com

Abstract - Until recently, data fusion problems have been solved heuristically using fuzzy logic, rule-based inference, the Dempster-Shafer (DS) theory of evidence, etc. Data fusion is often just a sequential series of functions, with occasional feedback loops. Beginning in the late 1970s, however, researchers have shown how data fusion can be placed under a purely probabilistic and theoretically rigorous paradigm based on the theory of (closed) random sets. The purpose of this talk is to provide a brief "mathematician's overview" of this work, especially Finite Set Statistics (FISST). Random sets provide a natural setting for data fusion in two respects:

1. as a natural way of formulating Multi-Sensor (MS) Multi-Target (MT) problems
2. as a means of modelling ambiguous evidence.

Keywords: random sets, unification, Bayesian approach, multi-target, multi-sensor.

1. Introduction

Let us consider the diversity of the information fusion dilemna.

Suppose, on the one hand, that one or more sensors collect observations from one or more targets. Then the total observation (taken over all sensors) consists of a finite set of (conventional) measurements that varies randomly (with respect to cardinality as well as individual measurements). Likewise, our total estimate of reality consists of a finite set of (conventional) estimates of individual targets that varies randomly.

On the other hand, consider an English-language statement such as "The helicopter is near the ownship." The concept "near" could be interpreted as a single closed circular region surrounding the ownship. More generally, it could be modelled as a family of such circular regions, each assigned the subjective probability that it is the most likely interpretation of "near". Such a family is a random set that models the concept "near". One easily sees, accordingly, that a "natural" mathematical universe for data fusion is a space whose elements are sets consisting of a finite number of (closed) random sets. However, such a framework is "natural" in an engineering sense only if it can be formulated in a manner that facilitates practical application. This lecture will outline such a formulation.

Random set theory and FISST applications in Multi-Sensor Data Fusion (MSDF) for MS-MT tracking and identification (ID), have been spearheaded since the mid 1990 by researchers at Lockheed Martin NE&SS in Eagan, particularly by Ronald Mahler. This lecture attempts to summarize the approach of his book [1] "Mathematics of Data Fusion". In order not to burden this presentation with many references, all bibliographical references can be obtained from that book.

2. Conventional Approach, Known Problems And Potential Solution

2.1 The Bayesian Approach

Recent years have seen the emergence of a "cookbook Bayesian" viewpoint based on Bayes rule,

$$f(x|z) = \frac{f(z|x)\,f(x)}{f(z)}$$

where *x* denotes the unknown quantities of interest (kinematic and ID), the prior encapsulates our previous knowledge about *x*, *z* is the new data, the likelihood function *f(z|x)* describes the generation of data and the posterior encapsulates our calculated current knowledge about *x*. Finally,

$$f(z) = \int f(z|x)\,f(x)\,dx$$

is a normalization constant. If time has expired before collection of new information *z*, then the prior has to be extrapolated to a new prior that reflects the uncertainties caused by possible interim target motion, which is usually done by evaluating the Markov time-prediction integral

$$f(x) = \int f^{+}(x|y)\,f(y)\,dy$$

where $f^{+}(x|y)$ is the Markov transition density describing the likelihood of the target having state *x* if it previously had state *y*. Translation of the calculated posterior into operator language is obtained through Bayes-optimal state estimators such as the Maximum A Posteriori (MAP) or Expected A Posteriori $\hat{x} = \arg\max_x f(x|z)$

$$\int x f_{posterior}(x|z)\,dx$$

However, it is merely the visible tip of a conceptual iceberg, the existence of which tends to be forgotten precisely because the rest of the iceberg is taken for granted. Both the optimality and simplicity of the Bayesian framework can be taken for granted only within the confines of standard applications addressed by standard textbooks. When one ventures out of these confines, one must exercise proper engineering prudence which includes verifying that standard textbook assumptions still apply. The Bayesian "iceberg" has many facets:

1. Sensor Models: Bayes' rule exploits to the best possible advantage the high-fidelity knowledge about the sensor contained in the likelihood function $f(z|x)$. Many forms of data, e.g., generated by tracking radars, are adequately characterized that $f(z|x)$ can be constructed with sufficient fidelity. Other kinds of data, e.g., Synthetic Aperture Radar (SAR), are proving to be so difficult to simulate that it is unclear whether sufficiently high fidelity will ever be achieved. There is also the kind of data, features extracted from signatures, English-language statements received over datalink, rules drawn from knowledge bases, etc., which is statistically so poorly understood, that probabilistic approaches are not even obviously applicable.
2. Target Models: Much of what has been said about likelihoods $f(z|x)$ applies with equal force to Markov densities $f^{+}(x|y)$. The more accurately that $f^{+}(x|y)$ models target motion, the more effectively Bayes' rule will do its job. In real-world scenarios targets can appear (e.g., MIRVs and decoys emerging from a ballistic missile re-entry vehicle) or disappear (e.g., aircraft that drop beneath radar coverage) in correlated ways. Consequently, MT filters that assume uncorrelated motion and/or constant target number may perform poorly against dynamic MT environments, for the same reason that single-target trackers that assume Great care must be exercised in the selection of a state estimator, however. For example, the EAP estimator plays an important role in theory but often produces erratic and inaccurate solutions when the posterior is multimodal. In the MT case, the dangers of taking state estimation for granted become even more acute. For example, we may fail to notice that the MT versions of the standard MAP and EAP estimators are not even defined, let alone provably optimal.
3. Formal Optimality: The failure of the standard Bayes-optimal state estimators in the MT case has far-reaching consequences for optimality. Because the standard Bayes-optimal state estimators fail in the MT case, we must construct new MT state estimators and prove that they are well behaved.
4. Computability: The prediction integral and Bayes normalization constant must be computed using numerical integration and, since an infinite number of parameters are required to characterize $f_{posterior}(x|z)$, in general, approximation is unavoidable.

2.2 Why Multi-Evidence Problems Are Tricky

Given the technical community's increased understanding of the actual complexity of real signature and other kinds of data, it is no longer credible to invoke Bayesian filtering and estimation as a cookbook panacea. Any such approach that does not also include a robustness strategy, i.e., a means of dealing with likelihood functions that cannot be specified with sufficient fidelity or with data that contains inherently difficult-to-characterize uncertainties, is simply evading the real issues. One needs systematic and fully probabilistic methodologies for:

1. modelling uncertainty in poorly characterized likelihoods
2. modelling ambiguous data and likelihoods for such data
3. constructing MS likelihoods for ambiguous data
4. efficiently fusing data from all sources (ambiguous or otherwise)

2.3 Why Multi-Target Problems Are Tricky

Likewise, given the technical community's increased understanding of the actual complexity of MT problems, it is not credible to propose yet another *ad hoc* MT tracking approach. When one tries to apply the standard statistical thinking just described to the MT case with an unknown number of targets, one quickly discovers that Bayes-optimal MT estimation and filtering encounters fundamental conceptual and practical difficulties. One needs a systematic means of constructing:

1. provably true *(as opposed to* heuristic) MS-MT likelihood functions from the underlying sensor models of the sensors
2. provably true *(as opposed to* heuristic) MT Markov densities from the underlying motion models of the targets, which account for target motion correlations and changes in target number
3. stable, efficient, and provably optimal MT state estimators that address the failure of the classical Bayes-optimal estimators and, in particular, simultaneously determine target number, target kinematics, and target identity without resort to operator intervention or optimal report-to-track association

It should be understood that many MS-MT tracking and association problems for positional and ID fusion seen in real-life situations may not require all the complexity espoused here. As usual, "one uses what one can get away with".

2.4 Finite-Set Statistics (FISST)

One of the major goals of FISST is to address the "Bayesian iceberg" issues described previously. FISST deals with imperfectly characterized data and/or measurement models by extending Bayesian approaches in such a way that they are robust with respect to these ambiguities. FISST deals with the difficulties associated with MS and/or MT problems by directly extending engineering-friendly single-sensor, single-target statistical calculus to the MS-MT realm. Finally, FISST provides mathematical tools that may help address the formidable computational difficulties associated with MS-MT filtering (whether optimal or robust).

The basic approach is as follows. A suite of known sensors transmits, to a central data fusion site, the observations they collect regarding targets whose number, positions, velocities, identities, threat states, etc. are unknown. Then:

1. reconceptualize the sensor suite as a single sensor (a "global sensor")
2. reconceptualize the target set as a single target (a "global target") with MT state $X=\{x_1,\ldots,x_n\}$ (a "global state", note the use of a capital letter)
3. reconceptualize the set $Z=\{z_1,\ldots,z_n\}$ of observations, collected by the sensor suite at approximately the same time, as a single measurement (a "global measurement", note the use of a capital letter) of the "global target" observed by the "global sensor"
4. represent statistically ill-characterized ("ambiguous") data as random closed subsets Θ of (MS) observation space. Thus, in general, $Z = \{z_1,\ldots,z_n,\Theta_1,\ldots \Theta_{m'}\}$
5. just as single-sensor, single-target data can be modelled using a measurement model $Z=h(x,W)$, model MT MS data using an MS MT measurement model - a randomly varying finite set $\Sigma=T(X)\cup C(X)$
6. just as single-target motion can be modelled using a motion model $X_{k+1}=\Phi(x_k,V_k)$**,** model the motion of MT systems using an MT motion model - a randomly varying finite set $\Gamma_{k+1}=\Phi(X_k,V_k)\cup B_k(V_k)$.

Given this, we can reformulate MS-MT estimation problems as single-sensor, single-target problems. The basis of this reformulation is the concept of belief-mass β. Belief-mass functions are non-additive generalizations of probability-mass functions. The FISST MS-MT differential and integral calculus is what transforms these mathematical abstractions into a form that can be used in practice. The following general methodology, nicknamed Almost-Parallel Worlds Principle (APWOP), for attacking MS-MT data fusion problems works, modulo certain remarks. Nearly any concept or algorithm phrased in random-vector language can, in principle, be directly translated into a corresponding concept or algorithm in the random-set language. One says "almost-parallel" because, as with any translation process, the correspondence between dictionaries is not precisely one-to-one (for example, vectors can be added and subtracted whereas finite sets cannot). Nevertheless, the parallelism is complete enough that, provided one exercises some care, a hundred years of accumulated knowledge about single-sensor, single-target statistics can be directly brought to bear on MS-MT problems. This is summarized in the following Table 1:

Table 1. Mathematical parallels between single sensor, single target and MS-MT problems

Random Vector, Z	Finite Random Set Σ
sensor	global sensor
target	global target
observation, z	observation-set, Z
parameter, x	parameter-set, X
sensor model, $z = h(x,W)$	MT s. m., $Z = T(X) \cup C(X)$
motion model, $x_{k+1}=\Phi(x_k,V_k)$	MT m. m., $\Gamma_{k+1}=\Phi(X_k,V_k) \cup B_k(V_k)$
differentiation, dp/dz	set differentiation, $\delta\beta/\delta Z$
integration, $\int_S f(z\|x)dz$	set integration, $\int_S f(Z\|X)\delta Z$
probability-mass function, $p(S\|x)$	belief-mass function, β
likelihood function, $f(z\|x)$	MT l. f., $f(Z\|X)$
posterior density	MT posterior
Markov densities	MT Markov densities

The performance of an MT data fusion algorithm can be measured by constructing information-based measures of effectiveness, e.g., the following MT generalization of the Kullback-Leibler discrimination in terms of the densities

One should realize that such a correspondance table occurs in many other scientific fields, for example in physics. There, we have the gradual increase in complexity from large scale linear problems dealt with classical methods (e.g. vector mechanics, or tensor formulations of electromagnetism), to the small-scale low-energy problems (with actions of the order of Plank's constant h) treated by quantum theory, to ultimately, the high-energy small-scale realm of quantum field theory, where even the vacuum is a complicated concept. The same can be said of everyday engineering thermodynamics being underpinned by statistical mechanics, itself a subset of quantum statistical theory. Table 2 parallels the contents of Table 1.

Table 2. Mathematical correspondances between physical phenomena at various scales

Classical theory	Quantum theory	Quantum field theory
Fixed no. of large particles	Fixed no. of small particles	Variable no. of small particles
3D Momentum variable	3D Momentum operator	4D Momentum operator
State vector	Wavefunction concept	Field concept
Pos. & mom. both defined	Pos. & mom. operators do not commute	Fields commute (bosons) or anti-commute (fermions)

3. Basic Statistics for Tracking & ID

The foundation of applied tracking and identification, namely the recursive Bayesian nonlinear filtering equations, is described in Section 1. The procedure for constructing provably true sensor likelihood functions from sensor models, and provably true Markov transition densities from target motion models, is described in Sections 2 and 3, respectively. The basis for fully automated tracking and identification, Bayes-optimal state estimators, is reviewed in Section 4. Section 5 provides a very brief survey of some of the major computational issues and approaches in real-time nonlinear filtering.

3.1 Bayes Recursive Filtering

Most signal processing engineers are familiar with the Kalman filtering equations. Suppose that we are given a linear sensor measurement model $Z=Hx+W$ and a linear target motion model $X_{k+1}=\Phi_k(X_k)+V_k$ where H, Φ_k are matrices, while W and the set V_k are independent, zero-mean Gaussian noise vectors. Then the Kalman time-update and information-update equations are:

The Kalman filter is a special case of the Bayesian discrete-time recursive nonlinear filter. This more general filter is nothing more than the equations in Paragraph 1.1 applied recursively, namely:

The Bayes nonlinear filter equations assume certain conditional independence properties that need not be discussed here. The practical success of these equations depends upon our ability to effectively construct the likelihood function $f(z|x)$ and the Markov transition density $f_{k+1|k}(x_{k+1}|x_k)$. Though likelihood functions are sometimes constructed via direct statistical analysis of data, more typically they are constructed from sensor measurement models. Markov densities are typically constructed from target motion models.

The usual choices for the "true" likelihood function is $f(z|x)=f_w(z-h(x))$ if the measurement model is $Z = h(x)+W$, and for the true Markov density $f_{k+1|k}(x_{k+1}|x_k)=f_{Vk}(x_{k+1}-\Phi_k(x_k))$ if the motion model is $X_k=\Phi_k(x_k)+V_k$, where both W and V_k are again independent, zero-mean Gaussian noise vectors. How one gets to such expressions is detailed in the next two sections. Finally the most common "good" Bayes state estimators are just the previously defined MAP and EAP evaluated at time k.

3.2 Constructing Likelihoods From Sensor Models

Suppose that a target with randomly varying state X is interrogated by a sensor which generates observations of the form $Z=Z_{X=x}=h(x)+W$, where W is a zero-mean random noise vector with density $f_w(w)$, but which does not generate missed detections or false alarms. The statistical behaviour of Z is characterized by its likelihood function $f(z|x)$ which describes the likelihood of the sensor collecting measurement z given that the target has state x. How do we compute this likelihood function? We begin with the probability mass function (a.k.a. probability measure) of the sensor model:

$$p(S|x) = Pr(Z \in S)$$

This is just the total probability that the random observation Z will be found in any given region S if the target has state x. The total probability mass $p(S|x)$ in a region S is just the sum of all the likelihoods in that region:

$$p(E_z|x) = \int_{E_z} f(y|x)dy \cong f(z|x)\, v(E_z)$$

where $v(E_Z)$ is a small (hyper)volume about the point z. In the limit of infinitely small $v(E_Z)$, division of the above equation by the infinitesimal volume defines the Radon-Nikodyn derivative

Thus we have found the provably true likelihood function, i.e., the density function that faithfully describes the measurement model $Z=h(x,W)$. For the additive version of the model $Z=Z_{X=x}=h(x)+W$, it can easily be shown that this leads to $f(z|x)=f_W(z-h(x))$, namely a "noisy" version about the expected observation.

3.3 Constructing Markov Densities From Motion Models

Suppose that, between the k^{th} and $(k+1)^{st}$ measurement collection times, the motion of the target is best modelled by an equation of the form $X_{k+1}=\Phi_k(X_k)+V_k$. That is, if the target had state X_k at time-step k then it will have state $\Phi_k(X_k)$ at time-step k+l, except that possible error in this belief is accounted for by appending the random variation V_k. How do we construct $f_{k+1|k}(x_{k+1}|x_k)$? It is clear that this situation is completely parallel to the previous section. Specifically, the probability mass function $p_{k+1}(S|x_k)=Pr(X_{k+1}\in S)$ is the total probability that the target will be found in region S at time-step k+1, given that it had state X_k at time-step k, namely

is the "true" Markov density associated with the motion model. For the additive version of the model, it can easily be shown that this leads to

i.e. a "noisy" version about the expected position from the motion model.

3.4 Computability Issues In Single-Target Filtering

The following is a very cursory survey of some of the major computational strategies.

1. Approximate Filters: approximate filters achieve computational tractability by replacing likelihoods and posteriors by their linear, quadratic, higher-order, or other kinds of approximations. The best known approximate filters are the Extended Kalman Filter (EKF), the Iterated EKF (IEKF), and quadratic and higher-order filters. A major problem with approximate filters is the fact that, over time, approximation error tends to accumulate and can cause poor performance or even divergence.
2. Exact Filters: an exact filter is one that achieves computational tractability by assuming that posterior distributions belong to some family of densities, e.g., Gaussians or generalized exponential. Exact filters will perform badly when the real-world posteriors deviate significantly from this form. Besides the Kalman and Kalman-Bucy filters, the best known examples of exact filters are those of Benes and Daum.
3. Infinite-Dimensional Exact Filter: Kouritzin's convolution filter is apparently the only existing infinite-dimensional exact filtering technique. Assume that the old posterior and its time-update must be related by some (unknown) convolution kernel K. Once K has been determined off-line, one can compute $f_{k+1|k}$ in real time using Fast Fourier Transforms.
4. Spectral Separation Filter: the Spectral Separation Scheme (S^3) filter of Rozovskii, Lototsky, et. al. is also an off-line technique. It is based on the fact that the unnormalized posterior can be approximated as a truncation of a rapidly convergent stochastic infinite series. The S^3 filter can accommodate quite general continuous-time motion models, but observation noise is assumed to be additive and Gaussian. It is therefore typically used in conjunction with non-parametric estimators to determine and pre-process the actual noise backgrounds.

4. MT Sensor and Motion Models

MT measurement models have been constructed for the following successively more realistic situations:

1. MT measurement models with no missed detections and no false alarms/clutter
2. MT measurement models with missed detections
3. MT measurement models with missed detections and false alarms or clutter
4. MT measurement models for clutter caused by targets equipped with Electronic Counter-Measures (ECM)
5. MT measurement models for the multiple-sensor case

6. single-target and MT measurement models for data that is imperfectly characterized or otherwise" ambiguous".

MT motion models have also been constructed for the following successively more realistic situations:

1. MT motion models assuming that target number does not change.
2. MT motion models assuming that target number can decrease. These models are analogous to MT measurement models with missed detections.
3. MT motion models assuming that target number can decrease or increase. These models are analogous to MT measurement models with missed detections/clutter.

5. Unification through the FISST MS-MT Calculus and Statistics

This chapter introduces the mathematical core of FISST: the FISST MT integral and differential calculus. The basic purpose is to establish the parallelisms between single-target problems and MS-MT shown in Table 3:

Table 3. Parallelisms in calculus and statistics

probability-mass function, p(S|x)	belief-mass function, β
differentiation, dp/dz	set differentiation, δβ/δZ
integration, $\int_S f(z\|x)dz$	set integration, $\int_S f(Z\|X)\delta Z$

This will allow true MS-MT likelihood functions to be constructed from the measurement models of the individual sensors and true MT Markov transition densities to be constructed from the motion models of the individual targets. Thus FISST provides a formalism that unifies a multitude of separate single target algorithms into a single coherent picture of MS-MT.

5.1 The Belief-Mass Function Of A Sensor Model

Just as the statistical behaviour of a random observation vector Z is characterized by its probability mass function *p(Slx) = Pr(Z⊂S)*, so the statistical behaviour of the random observation-set Σ is characterized by its belief-mass function (a.k.a. belief measure):

$$\beta(S|X) = Pr(\Sigma \subseteq S)$$

where Γ is the random MT state. The belief mass is just the total probability that all observations in a sensor (or multi-sensor) scan will be found in any given region S, if targets have MT state X. For example, if $X=\{x\}$ and $\Sigma=\{Z\}$ where Z is a random vector then

$$\beta(S|X) = Pr(Z \in S) = p(S|x)$$

i.e., the belief mass of a random vector is just its probability mass.

5.2 The Belief-Mass Function of a Motion Model

In single-target problems, the statistics of a motion model $X_{k+1}=\Phi(x_k,V_k)$ are described by the probability-mass function $p_{k+1}(S|x_k)=Pr(X_{k+1}\in S)$, which is the probability that the target-state will be found in the region S if it previously had state $\mathbf{X_k}$. In like manner, suppose that $\Gamma_{k+1}=\Phi(X_k,V_k)$ is an MT motion model, then the statistics of the finitely varying random state-set Γ_{k+1} is described by its belief-mass function

$$\beta(S|X) = Pr(\Gamma_{k+1} \subseteq S)$$

which is the total probability of finding all targets in region S at time-step k+1 if, in time-step k, they had MT state $X_k=\{x_{k,1},\ldots,x_{k\ n(k)}\}$.

5.3 FISST, the Set Integral and Set Derivative

The conventional probability summation equation generalizes to a set integral

where the meaning of any set integral is the following:

Note that the set notation (curly brackets) and the regular ordering notation (round brackets) are just related by $F(\{y_1,\ldots y_j\})=j!F(y_1,\ldots y_j)$, so that the factorial in the above equation does not carry any deep significance. In the above, F could be an MS-MT likelihood $F(Z)=f(Z|X)$, an MT Markov density $F(X)=f_{k+1|k}(X|X_k)$ or an MT prior or posterior $F(X)=f_{k|k}(X|Z_k)$. Some care must be exercised with set notation and integration when the variable *y* is completely discrete, but this is outside the scope of this presentation.

The inverse function is the set derivative which, for Z having distinct entries, reads

With this definition the following standard rules are well defined:

1. Sum rules (linear combination distributivity)
2. Product rules (differentiate each term leaving others in product constant, then sum)

3. Constant rule (differentiation gives zero)
4. Chain rule (for composite functions)
5. Power rule (with appropriate factorials).

With these tools, it is a straightforward but tedious task to show that the remaining correspondences of the introduction follow naturally, as shown in Table 4:

Table 4. Parallelisms for likelihoods and densities

likelihood function, $f(z\|x)$	MT likelihood $f(Z\|X)$
posterior density $f_{k\|k}(x\|Z^k)$	MT posterior $f_{k\|k}(X\|Z^k)$
Markov densities $f_{k+1\|k}(x_{k+1}\|x_k)$	MT Markov densities $f_{k+1\|k}(X_{k+1}\|X_k)$

6. Bayes Fusion, Tracking & ID

6.1 Optimal Bayes Fusion, Tracking & ID

Bayesian MT filtering is inherently nonlinear. The easiest way to see this is to look at the simplest true MT problem: two targets observed by a single Gaussian sensor with noise model $N_Q(z\text{-}Hx)$, perfect detections, and no false alarms. In this case

[illegible]$NzHx$

When the targets are greatly separated from each other, one of the two terms dominates and this distribution is essentially Gaussian. If they are not greatly separated it becomes a Gaussian mixture. This particular Gaussian mixture is highly non-Gaussian when the two targets have intermediate separation, i.e., not too far from each other but also not too near**.** In other words, multi-target nonlinear filtering is unavoidable if our goal is optimal-Bayes tracking of multiple, closely-spaced targets. Using FISST,

the nonlinear filtering equations can be generalized to MS-MT problems.

The single-sensor, single-target Bayesian nonlinear filtering equations of Section 2 are already computationally demanding, and computational difficulties can get only worse when we attempt to implement the MT nonlinear filtering equations. Instead of the single numerical multivariate integrals of the single-target case, we are now faced with numerical set integrals. For example, in a scenario with as many as M targets, numerical evaluation of a set integral will require as many as $M^2/2$ single-target approximate multivariate integrals. Likewise, the already difficult problem of compressing a single-target posterior into a representation with a small number N of parameters expands into the much more difficult problem of compressing a multi-target posterior into a representation that requires on the order of N^M parameters. Clearly, if multi-target optimal Bayesian filtering is ever to become practical, drastic approximations will be required. The usual approximations involve the Gaussian approximation since it is well known that a convolution of Gaussians will yield another Gaussian.

6.2 Robust-Bayes Fusion, Tracking & ID

Up until this point one has assumed that observations *z* are precisely defined and that their corresponding likelihood functions *f(zlx)* are known with sufficient fidelity to support optimal-Bayes MS-MT fusion, detection, tracking, and identification. However, as noted previously, it is unclear whether real-time, sufficiently high-fidelity likelihoods will ever be achieved for certain kinds of data, e.g. SAR images or HRR range profiles. In such situations optimal-Bayes algorithms can be non-robust, i.e., perform more poorly than anticipated because of the discrepancies between the modelled and actual likelihoods. Similarly, other kinds of data, such as features extracted from signatures, English-language statements received over link, etc., are so "ambiguous" (poorly understood from a statistical point of view) that probabilistic approaches are not even obviously applicable. The purpose of this chapter is to address the following question: How does one extend Bayesian (or other forms of probabilistic) inference to situations in which likelihood functions and/or data are imperfectly understood? This section presumes working familiarity with several reasoning frameworks outlined in my previous lecture, such as fuzzy logic and the Dempster-Shafer theory of evidence.

6.3 Random Set Models Of "Ambiguous" Data

The FISST approach to data that is difficult to statistically characterize is based on the following key point: ambiguous data can be probabilistically represented as random *closed subsets* of (MS) measurement space.
Consider a simple example. Suppose that a near-sighted observer (like myself) examines a distant object that is known to belong to the set of objects
V = {redball, greenball, redcube, greencube, bluecube}

Suppose that the observations available to this observer are

U = {RED, GREEN, BLUE, ROUND, SQUARE, UNCERTAIN}.

These observations U can be associated with certain subsets of V as shown in Table 5:

Table 5. Observations and their associated subsets

Red	{redball, redcube}
Green	{greenball, greencube}
Blue	{bluecube}

Round	{redball, greenball}
Square	{redcube, greencube, bluecube}
Uncertain	V

In other words, any particular observation implies only a constraint on what the identity of the target could be. Suppose, however, that the observation is equivocal in the sense that the observer can at best specify only a range of hypotheses about what he or she actually saw. That is, the observer may specify a list $T_1 \ldots T_N$, of subsets of V, believed to be valid with respective degrees of confidence m_I where each T_i is drawn from the six subsets of V listed above. This sort of ambiguous evidence can be represented as a random subset Θ with probability distribution

$$Pr(\Theta = T_i) = m_i \quad (i = 1 \ldots, N)$$

Because Θ is a random subset of the state space V, ambiguous observations constrain the state of the target directly. In general, however, many state variables may be hidden from the direct view of the observer. Accordingly, more general ambiguous observations will be random subsets of observation space U. In this case it is the data that is directly constrained by the ambiguous evidence rather than the state. Nevertheless, the state will still be constrained indirectly since a constraint on data implies a constraint on the state.

6.4 Unification Of Techniques For "Ambiguous" Data

It is one thing to recognize that random sets provide a common probabilistic foundation for various kinds of statistically ill-characterized data. It is quite another to construct practical random set representations of such data. Certain aspects of expert-systems theory, such as fuzzy logic, the Dempster-Shafer theory of evidence, and rule-based inference, provide well-known methodologies for mathematically encoding many kinds of ambiguous evidence. As noted in the last section, data that exhibits a great deal of imprecision can be represented as lists of evidential hypotheses, more commonly known as Dempster-Shafer bodies of evidence or "mass assignments." Likewise, "vague" concepts such as the "near" in the English-language statement "The helicopter is near the ownship" can be represented as fuzzy sets. The purpose of this section is to show how two kinds of ambiguous data, namely imprecise and vague data, can be represented probabilistically by random sets (it can be shown that contingent, and constrained-probabilistic data can also be put in random set form). This will provide us with the following recipe for processing ambiguous data:

1. represent data using expert-system models
2. transform these models to probabilistic random set mathematics
3. compute probabilistic formulas.

6.5 Vague Data: Fuzzy Logic

A fuzzy membership function on some (finite or infinite) universe U is a function that assigns a number *f(u)* between zero and one to each member u of U. The value of *f(u)* is interpreted as the "degree of membership" of u in the fuzzy set represented by f, with f(u)=1 indicating unequivocal membership and f(u)=0 unequivocal non-membership. Fuzzy evidence, as well as the operations of fuzzy logic, can be represented in random set form as follows. Let Σ be a random subset of some (finite or infinite) universe. The one-point covering function μ_Σ of Σ is the fuzzy membership function defined as $\mu_\Sigma(u) = p(u \in \Sigma)$. Conversely, let f: U → [0,1] be a fuzzy membership function of the universe and let A be a uniformly distributed random number on the unit interval [0,1]. Then the random subset $\Sigma_A(f)$, called the canonical random set representation of the fuzzy subset *f*, is

$$\Sigma_A(f) = \{u \in U \mid A \le f(u)\}$$

That is, the random set $\Sigma_A(f)$ can be interpreted as a faithful probabilistic representation of the fuzzy membership function *f*. The conventional fuzzy logic complementation operator can be regarded as the result of forcing algebraic closure on the canonical random set representation of the complement. Thus the celebrated non-Boolean behaviour of fuzzy logic can be interpreted as the result of ignoring the statistical correlations between fuzzy subsets. Also, the fact that there are an infinite number of possible fuzzy logics can be interpreted to mean that many fuzzy logics correspond to different assumptions about the statistical correlations between fuzzy sets.

6.6 Imprecise Data: Dempster-Shafer Bodies of Evidence

A Dempster-Shafer body of evidence B on some space U consists of nonempty subsets $B: B_1, \ldots, B_b$ of U and nonnegative weights $m_1, \ldots, m_b$ that sum to one. Each B is interpreted as a hypothesis "$u \in B_i$" constraining the possible value of the unknown quantity u; and the quantity m_i is the amount of mass that accrues to the hypothesis B_i alone and not to any strictly smaller subset of B_i. Every body of evidence B has an associated belief function and plausibility function defined by

$$Bel(S) = \sum_{B_i \subseteq S} m_i$$

$$Pl(S) = \sum_{B_i \cap S \neq \emptyset} m_i$$

If one defines the random subset Σ of U by $p(\Sigma = B_I) = m_I$, for i=1,…b, Σ is said to be the random set representation of B and one writes $B = B^\Sigma$. The mass function of the random set Σ is the same thing as the belief function of B^Σ.

The usual Dempster rule of combination that provides a means of combining two independent bodies of evidence B, C into a single body of evidence B*C can be interpreted as an algebraic operation that mimics set-theoretic intersection of the nonempty, independent random subsets of U. Let us recall that, for each nonempty hypothesis D_k obtained by this set intersection, the combination rule (a.k.a. the orthogonal sum) assigns the mass $(B*C)_k=d_k/(l-K)$ where d_k is the sum of all $b_i c_j$ such that $B_i \cap C_j = D_k$ and where K is the sum of all $b_i c_j$ such that $B_i \cap C_j = 0$.

7. Some Recent Applications of FISST

The success of any approach in engineering mathematics is measured by the applications that it enables. The purpose of this section is to summarize some recent applications of FISST to practical engineering R&D problems. This work is being conducted cooperatively by Lockheed Martin, Eagan (MN), Scientific Systems Company, Inc. (SSCI), and Summit Research Corporation (SRC). The applications to be described in the following subparagraphs are:

1. MS datalink fusion and Non-Cooperative Target Identification (NCTI)
2. naval passive-acoustic antisubmarine fusion and target identification
3. air-to-ground Automatic Target Recognition (ATR) using SAR
4. scientific performance evaluation of MS-MT data fusion algorithms
5. unified detection and tracking using true MT likelihood functions and approximate nonlinear filters
6. joint tracking, pose estimation, and target identification using High Range Resolution Radar (HRRR).

7.1 Offboard Multisource Attribute Fusion and NCTI

The goal of this work is to develop a concept-feasibility prototype of an optimally self-reconfiguring adaptive and robust NCTI algorithm based on fusion of multi-source datalink features. On many surveillance aircraft, workstation operators examine various forms of data such as images, extract "features" from this data (e.g., number of hubs or wheels, presence or absence of a gun, etc.), and then transmit these features onto a datalink in a standard alphanumeric format. Features of this kind are corrupted by certain irreducible uncertainties, such as "fat-fingering" (i.e., the operator accidentally hit the wrong key), degradations in operator effectiveness due to fatigue and/or data overload, differences in training, or differences in individual ability. In principle, any individual operator could be "calibrated" to statistically characterize these effects. Because this is not possible in reality, one can never know any given operator's likelihood function $f_{operator}(feature|state)$ with any great certainty.

Test data for this problem came from a simulated Persian Gulf scenario called "LONEWOLF-98". It consists of features parsed from datalink message lines derived from Synthetic Aperture Radar (SAR), Long Range Optical (LRO), Moving Target Indicator (MTI) radar, Electronic Intelligence (ELINT), and Signals Intelligence (SIGINT). The possible target types are tanks, self-propelled guns, missile launchers, and personnel carriers.

A FISST robust-Bayes classifier was constructed based on a knowledge base of model features for the nine target types. A FISST Bayes-rule classifier was tested against two types of data: (1) low-quality synthetic "pseudo-data"; and (2) "as-real-as-possible-but-unclassified" data from LONEWOLF-98. The pseudo-data was highly corrupted by erroneous clutter features. The FISST robust-Bayes classifier correctly identified the correct target with high confidence (greater than 98%) in all nine instances. The FISST classifier was also blind-tested against sixteen unknown targets in LONEWOLF-98. Data for LONEWOLF-98 consisted of nearly a hundred datalink messages containing a total of 145 message lines containing feature information. The FISST classifier identified the correct target in all instances with very high confidence (98% or greater in ten cases and 93% or greater in the remaining seven).

7.2 Robust Naval ASW / ASuW Target Identification

The system being developed under the SPAWARSYSCOM Passive Acoustic Classification System (PACS) program is a prototype MT detection, tracking, and identification system for Anti-Submarine Warfare (ASW) and Anti-Surface Warfare (ASuW). It uses the following data sources: time-frequency passive-acoustic; and datalink attributes from passive-acoustic, magnetic-field, electric-field, and ELINT. The system includes the following major subsystems:

1. a robust-Bayes classifier algorithm
2. a supporting knowledge base of robust target signature models, which mimics the "hierarchical" decision-making process used by master-level INTELL (in particular, acoustics intelligence) analysts
3. a multi-hypothesis correlator-tracker
4. a semi-automated passive-acoustic operator-assist interface
5. a naval contact manager.

The automated classifier is based on the "Fuzzy Conditioned Dempster-Shafer (FCDS)" approach that LM at Eagan has used in practical classifiers. The general idea is as follows. A typical passive-acoustic feature z consists of a time-varying random centre frequency surrounded by a band with time-varying random width. There are thousands of distinct acoustic features, most of which are extremely difficult to characterize statistically using a standard likelihood function $f(z|v)$, where v denotes target type. Instead, we use the concept of model-matching using generalized likelihood functions. When used in conjunction with the SRC acoustic model-signature knowledge base, classifiers based on both approaches have proved effective in identifying subsurface targets using passive-acoustic data in either datalink-feature form or time-frequency form. They have been blind-tested on real acoustic data with an approximate 85% correct-identification rate.

7.3 Robust ATR For SAR

One familiar approach to ATR of ground targets using sensors such as SAR is to compute a Maximum Likelihood Estimate (MLE) on the image vector. That is, let z_i be the intensity in the i^{th} pixel of the image, let $f_i(z|\theta,v)$ be the likelihood function for the i^{th} pixel given pose θ and target class v and, for simplicity, assume pixel-to-pixel independence. Then the MLE estimate of the target type is that v which maximizes the value of the joint likelihood function $f(z|\theta,v)=f_1(z_1|\theta,v)\ldots f_M(z_M|\theta,v)$ of the entire image-vector $z=(z_1,\ldots, z_M)$. However, real-world SAR images are corrupted by irreducible uncertainties that affect target signatures-e.g. dents, wet mud, irregular placement of surface equipment such as grenade launchers, etc. Such uncharacterizable statistical irregularities make it impossible to specify the likelihood functions $f_l(z_i|\theta,v)$ with enough fidelity to ensure proper performance of MLE (or related Bayesian) techniques. FISST provides a means of modelling these kinds of uncertainties, potentially resulting in algorithms more robust with respect to model-mismatch errors. Preliminary tests using simulated SAR data show that the MLE technique based on this approach is less likely to mis-identify a target than the conventional MLE, i.e., it is more robust, but it is also less confident than the conventional MLE about correct identifications.

7.4 Scientific Performance Evaluation for Data Fusion

The ability to meaningfully assess the competence of algorithms is a crucial part of developing and comparing practical MS-MT data fusion systems, whether this be MS integration (Level 1), threat and situation refinement (Levels 2 and 3), or sensor/resource management (Level 4). However, probably no other vital aspect of MS data fusion has been more heuristic, more poorly understood, and less deeply examined than metrology. Whereas considerable effort has been put into developing fusion algorithms that adhere to sound standards of statistical reasoning, fusion Measures of Effectiveness (MoEs) tend to be cobbled together as afterthoughts of the development of particular fusion algorithms. This situation is true even in Level 1 fusion where, despite the existence of a plethora of metrics and a vast infrastructure of statistical and engineering insight, ad hoc approaches to metrology lead to serious difficulties. In Level 1 fusion, the traditional approach is to use MoEs that measure some particular ("local") aspect of algorithm competence, e.g., localization miss distance, track purity, etc. Such MoEs often produce more confusion than clarity because optimization of an algorithm with respect to one particular local MoE (e.g., target localization accuracy) will not infrequently result in algorithm degradation as measured by another local MoE (e.g., target identity accuracy).

In Levels 2, 3, and 4 fusion the situation is even worse. There are few MoEs of any kind, *ad hoc* or otherwise. This section describes ongoing work whose purpose is to develop a systematic and scientifically defensible and practical approach to data fusion metrology at all Levels. The basic concept is that of directly generalizing information theory to the MS-MT realm.

Mathematical information is constructed and computed using information MoEs. Suppose, for example, that a Kalman tracker algorithm is tracking a single target whose position at any time we know. At any instant, how much information is the tracker producing about the target, compared to ground truth? One useful approach is to compute the Kullback-Leibler cross-entropy or discrimination

$$K(f_{Grnd}; f) = \int f_{Grnd}(x) \log \frac{f_{Grnd}(x)}{f(x)} dx$$

where f denotes the probability distribution associated with the Kalman filter output, and f_{Grnd} is a probability distribution that corresponds to perfect knowledge of ground truth. This metric is always nonnegative and vanishes if and only if $f= f_{Grnd}$. It measures the "directed information distance" between the tracker's estimate and ground truth in the sense that the larger the value of $K(f_{Grnd}; f)$, the less useful information the tracker is producing.

While overall information (whether with respect to complete knowledge or complete ignorance) is desirable, end users may be more interested in performance with respect to some specific function, e.g., target identification. If desired, the amount of information that is attributable to some specific fusion function (e.g., competence in target identification) can be parsed out of this total information score as a percentage. This quantity can be computed by altering the output of the data fusion algorithm to disable the function of interest, and then computing the difference in the information produced by the original and disabled algorithms:

$$NI_{enabled} - I_{disabled}$$

This reasoning is easily extended to Level 4 fusion. The reason is that sensor management is a support function, i.e., its purpose is not to estimate parameters of interest but rather re-allocate sensor dwells in order to collect higher-quality input data for the core fusion algorithm(s). In other words, sensor management subsystem is performing well if the core fusion system generates more information than would be the case if the sensor management subsystem were absent. The following formula encapsulates this idea into an implementable mathematical form:

$$NI_{DF+SM} - I_{DF}$$

where I_{DF+SM} is the information produced by the data fusion algorithm with the assistance of sensor management and where I_{DF} is the information produced by the algorithm without such assistance. Through suitable choice of the underlying information MoE, it can be computed relative to (1) perfect knowledge of ground truth, (2) imperfect knowledge of ground truth, or (3) complete ignorance of ground truth. Since information provides a common language for all performance measurements, it can be expressed as a percent increase *above the* baseline provided by the core data fusion algorithm (without sensor management). As before, the effectiveness of the sensor management system in aiding specific fusion functions (e.g., target identification) can be specified as numerical percentages of ΔI_{SM}. Current results appear to verify the potential utility of the approach against simple scenarios.

7.5 Track-Before-Detect Filtering In Clutter

LM at Eagan has designed and implemented an approximate nonlinear filter with the following properties:

1. its computational complexity is O(logN) or O(N), depending on the computational strategy
2. it is theoretically guaranteed to not diverge in the non-moving target case (a result which carries over to moving-target scenarios in which the data rate is high enough).

This filter has been tested successfully in several simple one-dimensional model problems. A generalization of this filter, intended to jointly detect and track more than one target in clutter, is currently under development.

7.6 Robust Joint Tracking and Identification for HRRR

A HRRR signature is a plot of an RF intensity vector I_i versus range bin i, representing the time-sequence of RF returns from a target as the HRRR plane wave passes through the target extent. This signature is highly aspect-dependent (pose θ) and is rich with features that can aid identification of target type T. The goal of this project is to prove the feasibility of an HRRR joint tracking and identification algorithm that:

1. is robust with respect to uncertainties in the likelihood $f(I_i|T,\theta)$
2. is robust with respect to novel target types
3. optimally estimates pose θ, target type T, and position/velocity
4. will potentially offer real-time operation when hosted on those processors expected to be resident on future generations of fighter aircraft.

A real-time kinematic nonlinear filter was integrated with an existing HRRR target identification algorithm from Airforce Rome Lab, called STaF, which was designed to overcome the lack of robustness of conventional Bayesian approaches to ATR for HRRR. Let i denote the i^{th} range bin of an observed HRRR profile and let $I_i=a(i)$ be its corresponding intensity amplitude. STaF uses the following procedure:

1. extract features, i.e., intensity peaks in the range-profile that are attributable to scatterers on the target surface rather than to random noise
2. from each extracted feature, construct modified posterior distributions that hedge against the possibility of novel or unknown target types; and
3. use Dempster's rule to combine the modified posteriors into n fused posteriors, each corresponding to a different target-identity hypothesis.

A concept-feasibility prototype algorithm based on this approach has been implemented in a restricted two-dimensional problem. Aircraft are assumed to move at constant altitude and to manoeuvre using coordinates turns (i.e., turns in which the length of the aircraft body is tangential to the flight path). Under such assumptions, the complete target state reduces to (x, y, v_x, v_y, w, T), where w denotes planar angular velocity, and between-measurements motion can be modelled using a nonlinear Markov state transition based on a small-angle approximation for each type T:

[illegible]

The approximate nonlinear filter of the previous section was modified to make use of this motion model and the FISST likelihood function, and integrated with an SSCI-developed implementation of the STaF algorithm. The integrated

[illegible]

algorithm was tested on real, simultaneously collected data I, z for three target types Tl, T2, and T3 with the following results. STaF was trained to recognize Tl and T2 but not T3. The integrated classifier-tracker algorithm was tested on data for Tl, T2, and T3 that had not been used in the training. When fed data from Tl or T2, the integrated algorithm decisively chose the correct target type. When fed data from T3, on the other hand, the algorithm was unable to make a determination between type Tl or type T2.

8. Summary and Conclusions

FISST has been created in part to address the issues in probabilistic inference that the "cookbook Bayesian" viewpoint encourages us to ignore or even be unaware of. These issues include:

1. dealing with poorly characterized sensor likelihoods
2. dealing with "ambiguous" data
3. constructing likelihoods for "ambiguous" data
4. constructing true MT likelihoods and Markov transition densities
5. dealing with the "curse of dimensionality" in MT problems
6. providing a single, fully probabilistic, systematic, and unified foundation for MS-MT detection, tracking, ID, data fusion, sensor management, performance estimation, and threat estimation and prediction, while
7. accomplishing all of this within the framework of a direct, relatively simple generalization of standard statistics and undergraduate calculus

During the last two years FISST has emerged from the realm of basic research to a range of practical engineering research applications. The purpose of this plenary talk has been to summarize the FISST approach and its use in such applications. The main challenges ahead are to increase the calculability of MT filtering, in general, beyond the Gaussian approximation.

9. References

[1] Goodman, I.R., Mahler, R.P.S., and Nguyen, H.T., (1997) *Mathematics of Data Fusion*, Kluwer Academic Publishers, 1997, ISBN 0-7923-4674-2, and references therein.