E. Blasch, M. Pribilski, B. Daughtery, B. Roscoe, and J. Gunsett, “Fusion Metrics for Dynamic Situation Analysis,” *Proc. SPIE 5429*, April 2004.

Fusion Metrics for Dynamic Situation Analysis

Erik. P. Blasch[1], Mike Pribilski[2], Brian Daughtery[2], Brian Roscoe[2], and Josh Gunsett[2]
1 – Air Force Research Laboratory, Dayton, Ohio, 2 - Jacobs Sverdrup, Dayton, Ohio

ABSTRACT

To design information fusion systems, it is important to develop metrics as part of a test and evaluation strategy. In many cases, fusion systems are designed to (1) meet a specific set of user information needs (IN), (2) continuously validate information pedigree and updates, and (3) maintain this performance under changing conditions. A fusion system’s performance is evaluated in many ways. However, developing a consistent set of metrics is important for standardization. For example, many track and identification metrics have been proposed for fusion analysis. To evaluate a complete fusion system performance, level 4 sensor management and level 5 user refinement metrics need to be developed simultaneously to determine whether or not the fusion system is meeting information needs. To describe fusion performance, the fusion community needs to agree on a minimum set of metrics for user assessment and algorithm comparison. We suggest that such a minimum set should include feasible metrics of **accuracy**, **confidence**, **throughput**, **timeliness**, and **cost**. These metrics can be computed as confidence (probability), accuracy (error), timeliness (delay), throughput (amount) and cost (dollars). In this paper, we explore an aggregate set of metrics for fusion evaluation and demonstrate with information need metrics for dynamic situation analysis.

Key Words: Fusion, Tracking, Metrics, Sensor Management, User Refinement

1. INTRODUCTION

Many academic communities have defined a constrained set of metrics to compare, analyze, or evaluate models. For instance, *Communication Theory* lists five key Quality of Service (QOS) metrics of delay, delay variance, Pr(error), throughput, and cost [24]. *Economics* has Macro (interest rate, money supply, unemployment) and Micro (price and quantity) metrics. The *business* community looks at the balance sheet to determine profits, opportunity cost, perpetuities, and return on investment (the bottom line). *Manufacturing* works with Control Theory (response time, rise time, etc.) and machine capabilities (6 sigma tolerance). Finally, if we look at social and cultural issues, there are also metrics for sports, consumer confidence, stocks, and entertainment which are updated daily. Thus, the fusion community needs to adopt a limited number of metrics to effectively communicate fusion gains. In each field of study, metrics are a set of statistics that can be assessed directly or indirectly through a regression analysis, ANOVA tests, and model validations.

Dynamic situation analysis has three components: (1) dynamical responsiveness to changing conditions, (2) situational awareness, and (3) continual analysis to meet performance requirements. The combination of these three entities are instantiated in a tracking and identification systems such as the Joint Belief-Probability Data Association Filter (JBPDAF) tracker [7], an interactive display to allow the user to make decisions, and metrics for replanning and sensor management. To afford interactions between future fusion system designs and users needs, metrics are required. In this paper, we use the User-Fusion model to highlight a set of fusion metrics to bridge the user-fusion gap. The metrics chosen include timeliness, accuracy, throughput, confidence, and cost. These metrics are similar to the standard QOS metrics in communication theory and human factors literature, as shown in Table 1.

Table 1: Metrics for various Disciplines

COMM	**Human Factors**	**Info Fusion**	**ATR/ID**	**TRACK**
Delay	Reaction Time	**Timeliness**	Acquisition /Run Time	Update Rate
Probability of Error	Confidence	**Confidence**	Prob. (Hit), Prob. (FA)	Prob. of Detection
Delay Variation	Attention	**Accuracy**	Positional Accuracy	Covariance
Throughput	Workload	**Throughput**	No. Images	No. Targets
Cost	Cost	**Cost**	Collection platforms	No. Assets
Stallings 2002	Wickens, 1992	**Blasch, 2003**	Blasch, 1999	Blasch, Hoffman 2000

E. Blasch, M. Pribilski, B. Daughtery, B. Roscoe, and J. Gunsett, "Fusion Metrics for Dynamic Situation Analysis," *Proc. SPIE 5429*, April 2004.

Section 2 overviews fusion levels, evaluation strategies, and tradeoffs. Section 3 describes various metrics associated with the different fusion levels. Section 4 describes a new set of metrics based on information needs (IN) and Section 5 presents results. Finally, Section 6 provides a discussion and conclusions.

2. BACKGROUND

2.1. Fusion Model

An operational system must satisfy the user's functional needs and extend their sensory capabilities. Of interest to the Information Fusion (IF) community are systems which translate data about a region of interest into knowledge, over which the user can reason, make decisions, and take action. [2] A user fuses data and information over time and space and acts through their cognitive model, whether it is in the head or with graphical displays, tools, and techniques. The current paradigm for fusion research, shown in Figure 1, includes:

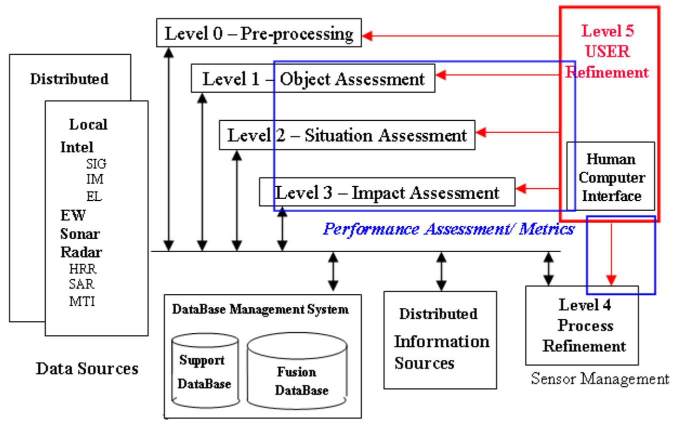


Figure 1. User-Fusion Model

Level 0 – Sub-Object Data Assessment: estimation and prediction of signal/object observable states on the basis of pixel/signal level data association (e.g. information systems collections);

Level 1 – Object Assessment: estimation and prediction of entity states on the basis of observation-to-track association, continuous state estimation and discrete state estimation (e.g. data processing); [16]

Level 2 – Situation Assessment: (element of situational awareness) estimation and prediction of relations among entities, to include force structure and force relations, communications, etc. (e.g. information processing);

Level 3 – Impact Assessment: estimation and prediction of effects on situations of planned or estimated actions by the participants; to include interactions between action plans of multiple players (e.g. assessing threat actions to planned actions and mission requirements, performance evaluation);

Level 4 – Process Refinement (an element of Resource Management): adaptive data acquisition and processing to support mission objectives (e.g. sensor management and information systems dissemination, C^2). [16]

Level 5 – User Refinement (an element of Knowledge Management): adaptive determination of who queries information and who has access to information (e.g. information operations) and adaptive data retrieved and displayed to support cognitive decision making and actions (e.g. altering the sensor display).

Level 1 and Level 4 functions require quality data for effective processing – similar to a user's requirement for IF for confident decision making. For example, it is difficult for a human to infer the target identity when the visual field is obstructed by clouds, where the user must wait for the clouds to disappear. Likewise, automation systems must receive a certain integrity level of input data to fuse it successfully into usable outputs; one example is to make a target ID despite cloud occlusion of sensor data. Further, the current understanding of a given situation must be thorough to ensure that the right mixture of the correct sensors is tasked for future sensing needs. Selection of tasks is based on a user's situational awareness (SA).

Issues associated with user SA are **(1) workload, (2) attention**, and **(3) trust** [2]. In the case of workload, we minimize the amount of information the human must process which also helps highlight the need for different types of "cause" data for the user. The highlighted information focuses the user's attention to a limited set of metrics to increase trust. Trust is important to convey extended sensory capabilities about the world. False data would hinder the direct connection between the user and the world around them. The issues of reaction time, trust, attention, and workload, are related to **timeliness, confidence, accuracy,** and **throughput**; respectively. Fusion system outputs include things (confidence/accuracy) and processes (time/throughput). It is the user that must interact with a fusion system to determine the importance of information collection which requires metrics. To reason over a changing situation, the user needs the correct data at the correct time.

Level 2 and 3 functions require contextual information for efficient processing – similar to a user’s requirement for dimensional reduction for decision making. For example, it is easy for a human to infer the target identity when the target is surrounded by vehicles of the same affiliation. Likewise, the automation system must be given a set of object, situation, and impact input data to be able to successfully reduce the search space. Finally, the current contextual information must be thorough to ensure that enough sensors are tasked for future sensing needs. User refinement is essentially the control of the fusion process, which is evaluated at each time step.

2.2. Fusion Performance Evaluation

Waltz and Llinas (1990) present a methodology for fusion test and evaluation. Key to fusion analysis is continuous evaluation, which is overviewed in “*Ch 11: Fusion Evaluation in Multi-Sensor Data Fusion*”. In the chapter, Command & Control fusion systems are designed to meet users’ needs. Defining what these needs are is the key to fusion performance; however, user needs change as a function of a dynamic context. Commanders interested in targeting desire high accuracy and resolution with minimum time delay. To meet these IF requirements, adequate models of algorithms and users are needed. Issues associated with modeling include formal, symbolic, semantic, and mental mathematical models for given scenarios with various operational complexities. For each model, fidelity varies as a function of sensor geometry, measurement accuracy and parametric representation. While all fusion systems can not be tested in the field, simulations can be developed with scenario generators, models, and data bases. Various testbed simulations can be used to test the robustness of fusion systems over such conditions are sensor variances, terrain content, and user types.

To determine fusion quality, an evaluation can be conducted over (1) fixed constants, (2) change in factors, and (3) Monte Carlo measurement and parameter variance. Based on the evaluation, desired quality objectives should be ranked, quantified, and qualified. Using objective rankings, candidate alternatives for fusion system designs can be determined based on the figures of merit [17]. ***Figures of Merit*** include measure of effectiveness (MOE), worth (MOW), Force Effectiveness (MOF) and performance (MOP). We utilize ideas presented by Waltz to arrive at a metric taxonomy of (1) effectiveness **MOE** (time), performance **MOP** [P(Detection) and P(False alarm)], (3) Dimensionality (throughput), and (4) Force Effectiveness **MOFE** (cost). Performance tradeoffs are derived from various metrics which afford evaluation, as compared to truth, for a fusion system to meet the utility, quality, and figure of merit constraints from the user.

2.3. User-Fusion Performance Tradeoffs

The user needs reliable information that is accurate, timely, and confident. If a machine presents results that conflict with the user’s expectations, the user would experience cognitive dissonance. Figure 2 shows technological and task operational conditions that affect the user’s ability to make informed decisions [9]. The IF system must produce quality results for effective and efficient decision making. An IF design should be robust in response to object, data, and environment variations. Using high fidelity models would increase the quality inputs to the IF system which would enhance user-fusion performance capabilities.

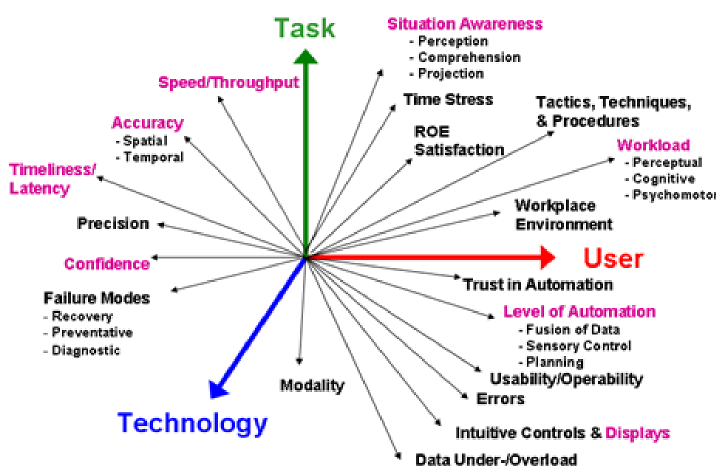


Figure 2. Machine-User interaction Performance

Metrics for (1) **user capabilities** of **trust, workload, attention**, (2) **cognitive modeling** for **situation awareness**, and (3) user **interaction** with the fusion system have been described in previous papers. [2, 3, 15, 19]. To explore actionable user functional control opportunities, we explore human capabilities to support fusion-system computations.

2.4. User Action

Using Figure 3, the OODA loop, the fusion system conducting an automatic target recognition (ATR) observes the sensor data and orients the data for the user. The user must decide if the ATR system is correct and acts on the results. To continue the loop, the user tasks the ATR sensor to collect data for fusion evaluation. The human can cue the fusion system by:

- **Observe –** Perception (Fuse Data)
- **Orient –** Situation Context (Memory of locations)
- **Decide –** Assess (Compare)
- **Act –** Plan /Scripts (order of sensor tasking)

In Figure 3, (1) is the intersection of the ATR system and the user (where ATR is one of many fusion operations). The ATR ID confidence must increase to enhance user trust. (2) represents the user tasking for more ATR data, and (3) shows that the OODA is iterated for efficiency (i.e. cycle time). From Figure 3, we see that the human processes the observation and orientation information from fusion evaluation. The user looks at the fusion system outputs and determines the final decision of the target type and the action to be performed. The user-fusion system performance is typically evaluated with a receiver operator curve (ROC).

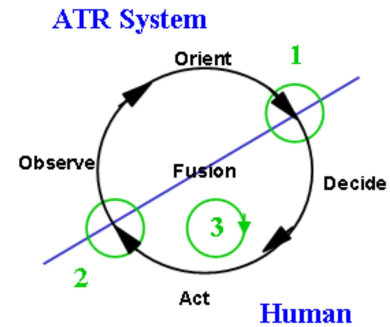


Figure 3. An ATR-Human OODA control Loop.

2.5. System Level ROC Curve Analysis

To develop system level metrics, it is a daunting task. One way to convey the decision criteria metrics across a fusion system is to utilize a *receiver operator curve* (ROC). ROC curves are used to display system performance for varying thresholds. ROC curves have been used in signal detection theory for human performance [4], sensor / radar ATR classification [5], and decision analysis [4]. Some ROC modifications have been proposed such an integrated human-ATR system [5], track and ID ROC curve (TIROC) [5], 3D ROC curve [1], time-dependent 3D ROC Curve [2], uncertainty operating curve UOC [6] and a **system ROC (SOC) curve** [12]. The development of different ROC curves are based on varying the threshold of detection so as to allow the user or the fusion algorithm to choose a conservative or relaxed threshold for a probability of detection for an acceptable false alarm rate.

Since the goal of a fusion system is to convey information to the user, we can use the ROC analysis combined with an acceptable determination of P_D versus P_{FA}. Thus, the goal is to determine various fusion levels metrics that ultimately increase a fusion gain. The metrics (**1) accuracy, (2) confidence, (3) throughput,** and **(3) timeliness**, are utilized in a ROC curve presentation with P_D and P_{FA} results. We seek a fusion gain [12] to evaluate whether or not the fusion system is combining data into useful information. Unlike [12], were we derived a fusion gain achieved as the area difference between two ROC curves; here we are interested in an information theoretic gain based on probabilities, time, and information needs satisfaction. To develop the SOC, we propose a minimum set of metrics to include in the analysis.

3. METRICS

Typical fusion parameters include **(1) probability, (2) error**, and **(3) information**. Probability is a normalized ratio of performance over a complete set of possibilities. Error is associated with uncertainty. Uncertainty is the result of the randomness of situational constraints that result from Fusion system performance in real-world testing. Such an example of uncertainty is the latency associated with incoming data. Information is a quality metric associated with the value of the data combination to meet functional needs. To develop any metric for system level performance, we use **probabilities (P), errors (σ),** or **time (t),** as contributing to the system metrics. Information metrics, such as mutual information or entropy describe the fusion gain [4-6]. Error and probability relate to confidence and accuracy. Finally, to meet real-world constraints, we need cost and time constraints to obtain the information required.

Using an evaluation taxonomy, we see that probability, error, timeliness, throughput, and cost are fusion requirements as well as user constraints. Fusion results can be related to a Kalman filter either in an ATR or tracking domain, where the uncertainty with the measurement includes an error (which is typically modeled as Gaussian white noise). The Kalman filter has three steps: (1) filter past state information, (2) estimate current state, and (3) predict future state. The KF minimizes the error and we can evaluate the performance of an ATR or tracking system [7] with the estimated state versus the truth data. The evaluation can also be assessed as to the P_{CA} versus P_{FA}. Thus, to evaluate a fusion system, we need to incorporate the desires of the user which become fusion INs. To accomplish the user-fusion interaction, we present a set of metrics that can be evaluated as to whether or not information needs are satisfied.

3.1. Metrics for Fusion

Metrics determine fusion performance and include (1) an **Objective** (desired) and (2) **Threshold** (minimum acceptable requirement). If we look at the fusion system from top down (satisfy user needs) or bottom-up (minimize uncertainty), the goal is to define the metrics for the development of fusion systems and operator performance.

Level 5: UR – Reaction Time, Trust, Throughput	(Satisfaction)
Level 4: PR – Sensor revisit rate, sensor coverage	(Timeliness)
Level 3: IA – Survivability, Vulnerability, Risk	(Cost)
Level 2: SA – Area of coverage, No. of observations	(Throughput)
Level 1: OA – Probability of Tracking and ATR / ID	(Confidence)
Level 0: Registration – positional error	(Accuracy)

Previously, we addressed the issues of **trust, workload**, and **attention**. [2] The interaction between a sensor manager and a user include displayed data and actionable information. Without human inputs, process refinement is based only on the data received and the fusion output and not the user needs.

3.2. Registration Metrics (Level 0)

Registration metrics usually entail a circular error probability (CE90) which is based on a distance metric such as Image Pixels on target. The distance metrics include a bias and a variance (mean) from the desired target and can be referenced to an accuracy metric. Non-IMINT data (i.e. SIGINT and MTI) can also be mapped with a bias and variance. For the evaluation, we use the area of interest (AOI) as the constant, the fiducial accuracies and sensor model variances, as well as the terrain accuracy levels for the factor analysis.

- **Geo-Location Error** — The geo-location error is defined as: $\varepsilon_{\text{Geo-error}} = \sqrt{\frac{1}{N_C} \sum \varepsilon_{\text{Lat}}^2 + \varepsilon_{\text{Long}}^2}$ (meters)

where: N_C is the number of control points,

$\varepsilon_{\text{Lat}} = (\text{Latitude}_{\text{Reported}} - \text{Latitude}_{\text{Truth}})\,\alpha_{\text{Latitude}}$ $\quad\varepsilon_{\text{Long}} = (\text{Longitude}_{\text{Reported}} - \text{Longitude}_{\text{Truth}})\,\alpha_{\text{Longitude}}$

Conversion: α_{Latitude} = Latitude Error (WGS-84) to Meters, $\alpha_{\text{Longitude}}$= Longitude Error (WGS-84) to Meters

- **Accuracy modeling** - KS Statistic, Chi-Square Test, or Wald Test:

The Kolmogorov-Smirnov (Goodness of Fit) test statistic is defined as

$$D = \max_{1 \le i \le n} \left| F(Y_i) - (i / N) \right|$$

where ***F*** is the theoretical cumulative distribution being tested, Y_i are the ordered set of points from 1 to N, and D is the statistic compared to a table (based on sample size N) to determine if the observed registration is within the truth registration distribution.

3.3. Detection, ATR, and Identification Metrics: (Level 1)

Target information can be modeled as per the NIIRS rating: detection, recognition, classification, and identification.

- **Probability of Detection** (P_D) - The ratio of the number of recorded detections (N_D) to the number of detection opportunities (N_{DO}). ($P_D = N_D / N_{DO}$).

Note: P_D is applicable to stationary and moving targets, where emitters can be inferred as detections. A moving target is said to be detected if a set of reports corresponding to the target are associated and a vehicle track is declared. A stationary target group is said to be detected if more than X% of the targets within the group are detected and associated with one another, where X is a parameter. A moving target group is said to be detected if a set of reports corresponding to more than Y% of the targets comprising the group are associated and a group track is declared, where Y is a parameter.

- **Probability of recognition** (P_R): ratio of correct type declarations (N_R) to opportunities (N_{RO}). ($P_R = P_R / N_{RO}$)
- **Probability of classification** using **Confusion Matrices** - A common reporting format for ATR systems is classification probabilities, including cross target probabilities, associated with a given population. For example,

a Confusion matrix records the probabilities P(i | j) = probability of declaring a vehicle as type i given that it is really type j, where i, j = different features.

- **Probability of Correct Identification** (P_{ID}) — Ratio of the number of times a target, emitter, or group is correctly identified (N_{ID}) to the number of occurrences (N_{IDO}). ($P_{ID} = N_{ID} / N_{IDO}$).

For example, we can use shape metrics for ID evaluation, e.g., RMS errors on length and width target attributes:

$$RMS_{Length} = \sqrt{\frac{1}{N} \Sigma \left[Length_{True} - Length_{Estmated} \right]^2} \; ; \; RMS_{width} = \sqrt{\frac{1}{N} \Sigma \left[Width_{True} - Width_{Estmated} \right]^2}$$

Others metrics include a Log-likelihood ATR, Maximum A Posteriori (MAP), and maximum likelihood (ML) [23]. The ML is based on the measurement information while the MAP is based on the expectation from the filtering analysis. As described by the Kalman Filter, we see that ML is used in the association. Estimation and prediction filtering use the MAP which is achieved from a Bayes analysis. In determining the true target analysis, we also desire to determine the error of the analysis using a false alarm metric:

- **False Alarm Rate** (FAR) - Number of false detections per square kilometer.

3.4. Track Metrics: (Level 1)

- **Probability of Track Detection** (P_{DT}) - Ratio of detected tracks (N_{DT}) to true track number (N_{TT}). ($P_{DT} = N_{DT} / N_{TT}$).
- **Track False Alarm Fraction** (F_T) - Ratio of false tracks (N_{FT}) to total tracks (N_{TT}). ($F_{FT} = N_{FT} / N_{TT}$).
- **Track Continuity** - Average number of tracks formed per trajectory of a single vehicle/group. Ideally equal to 1.
- **Track Purity** (T_P) — Ratio of track segments in an integrated track that belong to same vehicle, or group of vehicles, (N_{TS}) to total number of segments in a track (N_{TST}). ($T_P = N_{TS} / N_{TST}$)
- **Track Position Accuracy** — Root Mean Square Error between ground truth and tracker target positional estimates:

$$RSME_{TPA}(N) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \left[(x_i - \alpha_i)^2 + (y_i - \beta_i)^2 \right]} \quad \text{(meters)}$$

where, x_i, y_i are tracker position estimates, α_i, β_i are ground truth positions, and N is a specified number of samples defining the observation period.

- **Track Heading Accuracy** - RMS Error between ground truth and tracker vehicle/group heading estimates:

$$RSME_{THA}(N) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (\theta_i - \phi_i)^2} \quad \text{(degrees)}$$

where θ_i is tracker heading estimates with respect to North, ϕ_i is ground truth heading estimates with respect to North, and N is a specified number of samples

- **Track Velocity Estimate Accuracy** – RMS Error between ground truth and tracker vehicle/group velocity estimates:

$$RSME_{TVA}(N) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (v_i - \nu_i)^2}$$

where v_i are tracker velocity estimates, ν_i are ground truth velocities, and N is a specified number of samples.

- **Flow Rate Accuracy** - Root Mean Square Error between estimated flow rate and ground truth flow rate:

$$RSME_{FE}(N) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (\xi_i - \gamma_i)^2} \quad \text{(vehicles/sec)}$$

where; ξ_i is estimated flow rate (vehicles/sec), γ_i is actual flow rate, and N is a specified number of samples.

3.5. Situation Assessment Metrics (Level 2):

Situation awareness or assessment is typically evaluated based on user requests for a given situation. At higher levels of fusion, the lower level metrics are aggregated. For instance, individual object metrics of accuracy can be aggregated for group metrics such as group spacing, group identity, and relational aspects of group members (how likely are they to be members of the same group) [10, 11].

Situational metrics are correlated with user needs for situational awareness. Metrics include: attention, workload, trust, and dependability [2]. Attention and workload correlate to the communications *throughput* of the information. While lots of data could be time-consuming, it is assumed that the fusion system would deliver a parsimonious, reliable set of results to the user. Trust is related to *confidence* in presented results. Finally, dependability is related to *cost* since the situational content can either take time away from the user (opportunity cost) or minimize the effort needed to explore alternatives.

As an example of situational metrics, we suggest data **association** (matching of level one attributes of tracking and identification) metrics:

- **Probability of Correct Association** (P_{CA}) - Ratio of the number of correct correspondences across sensors, looks, etc. to the number of opportunities. ($P_{CA} = N_{CA} / N_{CAO}$)
- **Probability of Correct Track Association** (P_{TCA}) - Ratio of the number of correct track correspondences across sensor types (N_{TC}) to number of opportunities (N_{TCO}). ($P_{TCA} = N_{TC} / N_{TCO}$).

3.6. Impact Assessment Metrics (Level 3):

Impact assessment relates to benefits, costs, and risks. Since a fusion system is employed to reduce uncertainty or maximize information, it is important to choose metrics that address tradeoffs as a function of risk [18]. For example, risk metrics include: (1) Survivability – level of anxiety to risk and (2) Vulnerability – aggressive action to uncertainty.

A level 3 metric could include an exponential time decay (based on the *a priori* information) on the confidence of information received over time. The longer the delay means the higher the uncertainty and the greater the risk. The two additional metrics we propose are:

- **Percentage of plans achieved** (P_{PA}) - Ratio of plans completed to the number requested. ($P_{PA} = N_{PC} / N_{PCR}$)
- **Benefit to Cost Ratio Confidence** (BC) - Confidence of high value object accuracy/ID to opportunity cost to other objects ($BC_{HV} = HV_C / N_C$)

3.7. Process Refinement Metrics (Level 4):

Process refinement metrics include many factors in an optimization function. The difficulty is relating these factors in a single information value function. The value function is based on the information utility, span/region of responsibility, and sensor controllability. Some factors to consider are:

- **Timeliness** - Economic asset use (efficiency), Revisit rate of sensors, Update rate to display
- **Accuracy** - Sensor coverage area: unknown (search), known (track maintenance)
- **Throughput** - Density of performance, quality of information (Effectiveness)
- **Cost** - Opportunity cost of lost information, monetary value of sensor use
- **Confidence** - Satisfaction to commander's needs

Typically, profess refinement (1) minimizes operational cost or (2) maximizes information needs satisfaction. Using an optimization function, a derivative analysis can determine the operating point that maximizes or minimizes performance. Since impact assessment determines the associated costs, the sensor manager can then optimize information-need objectives given timeliness, confidence, and accuracy constraints. The constraints associated with the sensor manager include: (1) span of control over people and assets, (2) area of coverage control, and (3) timeliness for information usefulness. A sensor manger can utilize various value functions such as priorities, cueing, control (controllability, observability, stability), spatio-temporal constraints, utility analysis, and opportunity cost. The fusion system objectives are also affected by social, cultural, and regional factors. Finally, to determine information collections, throughput is essential to convey results to the user.

3.8. User Refinement Metrics (Level 5)

Typical user performance modeling includes **(1) actions, (2) time,** and **(3) knowledge**. [2] Such examples of actions include a tasks analysis to determine what (and which order) actions are processed. User models include **response times** such as physical responsiveness. Knowledge is measured as to whether the person accomplishes the task given. Typical tasks are broken down into simple things such as (1) Skills, Rules, and Knowledge tasks [21]; (2) Cognitive, Associative, and Automatic actions [20]; and (3) a mixture of the tasks and actions [4] for fusion system modeling Fusion systems and humans are ***trained to conform*** to a set of rules for normal conditions; however, the SRK model was used to explore how a human adapts from trained scenarios to novel situations [22]. Thus, as the fusion system explores new extended operating conditions, it is important to relate fusion system information to the user's cognitive requirements. Such an example is that of the human cognitively selecting a region of coverage and allowing the fusion system to determine all the targets in the area.

Just like fusion **timeliness**, the user has been modeled in a response time analysis. A response-time analysis can be of the form of physical, symbolic, or cognitive [26]. To determine the timeliness needed by users, human performance process models include [13] typically include physical, perceptual, and mental models:

- **Physical: Fitt's Law**. The time T_{pos}, to move the hand to a target of size S which lies a distance D away is given by:

$$T_{pos} = I_m \operatorname{Log}_2 (D / S + 0.5) \qquad \text{where } I_m = 100 [70 \sim 120] \text{ msec/bit.}$$

- **Perceptual: Power Law**. The time T_n to perform a task on the nth trial follows a power law:

$$T_n = T_1 n^{-\alpha} \qquad \text{where } \alpha = .4 [0.2 \sim 0.6].$$

- **Mental: Uncertainty Principle**. Decision time T increases with uncertainty about the judgment or decision to be made: $T = I_C H$, where *H* is the information-theoretic decision entropy and $I_C = 150$ [0 – 157] msec/bit.

  For n equally probable alternatives (called **Hick's Law**), $H = \operatorname{Log}_2 (n + 1)$

  For n alternatives with different probabilities, p_i, of occurrence, $H = \sum_i p_i \operatorname{Log}_2 \left(\frac{1}{p_i + 1} \right)$

To correlate fusion track/ID **confidence** to the user's trust, we utilize these properties

- **Trust:** Predictable, Dependable, Faith
  - **Faith -** Closure against doubt in an uncertain,
  - **Dependability –** results are consistent
  - **Predicable** – control actions carried out
- **Confidence – LEVEL 1→ 5** Object assessment correct
- **Reliability – LEVEL 2 → 5** accurate, consistent, verified
- **Security – LEVEL 3 → 5** Impacts assessed – mitigation of risks & uncertainty
- **Familiarity – LEVEL 4 → 5** practice and training, control
- **Integrity –** only show what is known that has been verified and validated

While there are many issues associated with user satisfaction for different users, the metrics would have to relate the quality of user ratings that can be ordered, normalized, and qualified. Polling different users would allow for an aggregation of results for a general measure of satisfaction.

Satisfaction = Predictability + Dependability + Faith + Competence + Reliability + Security + Integrity

In the next section, we try to objectively quantify the "satisfaction" by assessing the basic **information needs**, that a user or fusion system would desire [14]. If user or fusion requests are achieved, such as a (1) confidence bound, (2) location accuracy, (3) track life, or (4) communications timeliness; we then assume the fusion system has met user constraints. The use of information needs would discount user preferences to changing opinions and requests to differing situations.

4. INFORMATION NEEDS

The key to the sustained usefulness of a fusion system is the interaction between the fusion system and the user for decision making. In order to coordinate user requests and fusion outputs, we utilize the concept of "Information Needs". For example, the user selects an object to keep track of over some region of interest. Specifically, the user requests a certain **accuracy** [positional covariance (m^2)], **timeliness** [latency (s)], and **confidence** [belief - Pr(ID)]. The instantiation of these constraints will result in user satisfaction with fusion system outputs. Information needs could be about a target (position, velocity, ID, force structure), sensor (coverage area, timeliness), or the environment (terrain, weather). Information needs help a user plan the next course of actions and thus, if the objectives are met, the user will be satisfied. Likewise, a central fusion processor could also make requests to reduce the uncertainty (accuracy/confidence) and time constraints. Thus, the fusion system and human should have similar information needs.

These simultaneous user/fusion system information needs are shown in the figure below. Figure 4a shows the different sensor taskings as related to achieving desired IN results. Figure 4b shows the sensor coverage over the desired location of the track. If multiple coverage areas of the sensor are repeated, as in the case of multiple looks to increase ID confidence, then the track is labeled with higher probability belief.

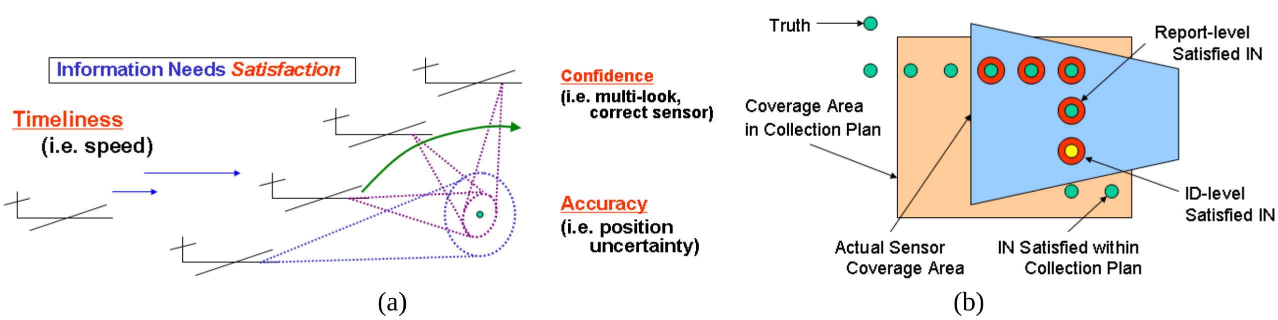


Figure 4. (a) Information needs, and (b) determination of satisfaction based on spatio-temporal/confidence assessment.

The satisfaction of timeliness needs is obvious – did the sensed data become available at the desired time? An upper bound on delay was used for the metric

$$\textbf{Timeliness IN} = \begin{cases} 1 & \text{if Information received} < \text{acceptable delay} \\ 0 & \text{otherwise} \end{cases}$$

For the case of accuracy, it was assumed that the sensor manager tasked the correct sensor (high resolution) or was able to fuse the results from many (low resolution) observations to reduce the error covariance. Since fusion and the user could request IN on targets, it was determined that a satisfied IN was when the correct sensor was tasked to prosecute a target. If object density is high, a higher resolution sensor was needed to get a measurement update. Likewise, if the track accuracy was low, another observation could be used to reduce uncertainty in position. The tasking of the correct sensor over the coverage area is shown in Figure 4(b). Thus, the metric was

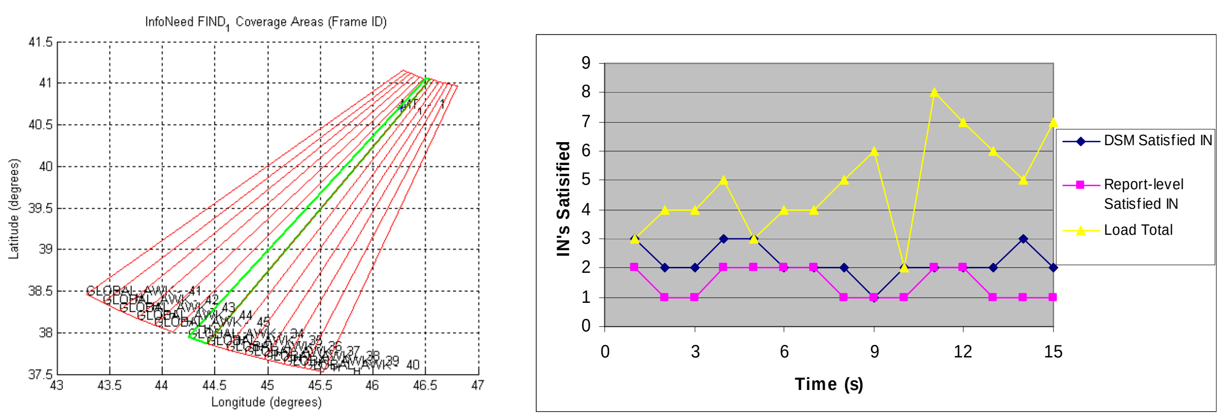
$$\textbf{Accuracy IN} = \begin{cases} 1 & \text{if sensor coverage at desired location} \\ 0 & \text{otherwise} \end{cases}$$

Finally, each sensor has a reporting confidence associated with the object ID. The ID confidence in the fusion system is determined to be the same value reported to the user. Thus, if one sensor can give high confidence in ID (such as vision system), then the IN is satisfied. If only a low-resolution sensor is available, then multiple looks is needed. For this analysis, the metric is:

$$\textbf{Confidence IN} = \begin{cases} 1 & \text{Pr(ID)} > \text{desired Pr(ID)} \\ 0 & \text{otherwise} \end{cases}$$

5. RESULTS

We simulate a tack/ID system and a user conducting a target ID for 3 targets. At each time instant, the user/fusion system requests a detection radar, classification radar, or an ID imaging system. The fusion system is trying to optimize the track life which can be achieved by having or robust ID confidence for an extended track. The limitation is that all the sensors can not be used for all targets at each time instant. The user chooses the constraints desired and the sensor manager weights the users requests with the fusion requests based on current track/ID hypotheses and then tasks sensors. If the Boolean requests were achieved, such as in Figure 7a (for the detection radar), then the IN score was incremented. At the next time step, different IN were desired and the result tabulated. The time history is shown in Figure 7b. Throughput and cost are still being formulated as communications affects throughput and cost is a normalization factor of benefits.



(a) (b)

Figure 7. Performance Information needs metrics evaluation for the user-fusion system

6. DISCUSSION & CONCLUSIONS

Fusion metrics include (1) evaluation metrics for fusion system development, (2) testing metrics to explain and compare fusion system algorithmic results, and (3) performance metrics to communicate added value to customers. In this paper, we discussed metrics that span the entire fusion model including:

Level 0 – registration – location accuracy, error
Level 1 – Track, ID (P_D, P_{FA}) confidence
Level 2 – Sensor coverage area
Level 3 – Risk and uncertainty
Level 4 – Throughput, revisit rate, re-tasking
Level 5 – Timeliness, satisfaction
System – Track/ID life, Benefit/Cost Ratio

Other important issues for giving feedback to fusion development through metric evaluation include: (1) **Cost** – number assets available and associated monetary value, (2) **Display metrics** – to relay fusion performance to the user, and (3) **Simulation** - verification of fusion models, validation of realism over operating conditions, and performance robustness. For fusion upgrades, there should be a business metric associated with the return on investment for IF machines, as shown in Figure 8. The use of and ROIF would complement similar metrics detailing a fusion gain that can be used for a benefit/cost ratio.

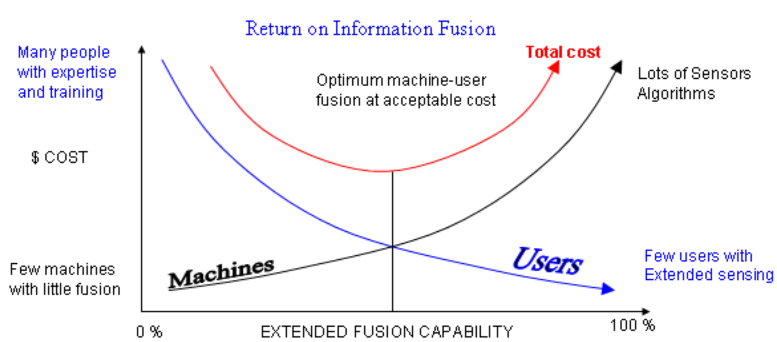


Figure 8. Return on Information Fusion (Investment).

In this paper, we highlighted various fusion metrics while postulating a **Information Needs Satisfaction**. We stressed the importance of developing fusion metrics that match the user needs for fusion system design. The Quality of Service (QOS) metrics for fusion systems are: **timelines**, **accuracy**, **confidence**, **cost**, and **throughput**. While there are many metrics in the literature, the fusion community needs to adopt a minimum set so as to focus designs, afford comparisons, allow for realization of implementation, and ultimately meet user information needs.

Timeliness	Accuracy	Confidence	Throughput	Cost
Time to detect	Track uncertainty	Correct ID	Coverage Area	Asset utilization
Track initiation time	Prediction accuracy	Declaration rate	% targets tracked	Revisit rates
Time tracked	Platform error, bias	False alarms	% **users needs met**	User salary
Plan approval time	Coverage Area	Trust	Number of images	Machine repair costs

7. REFERENCES

[1] S. Alsing, E. P. Blasch, and R. Bauer, "3D ROC Surface concepts for Evaluation of Target Recognition algorithms faced with the Unknown target detection problem," *SPIE - Vol. 3718,* 13-17 April 1999, pg. 449 – 458.

[2] E. Blasch & S. Plano, "JDL Level 5 Fusion Model 'User Refinement' Issues and Applications in Group Tracking", *SPIE*, 2002.

[3] E. P. Blasch, "Fusion of HRR and SAR information for Automatic Target Recognition and Classification," *Fusion 99,* Las Vegas, NV. July 6-9, 1999 , pp. 1221-1227.

[4] E. P. Blasch "Assembling an Information-fused Human-Computer Cognitive Decision Making Tool," *IEEE Aerospace and Electronic Systems Magazine*, June 2000, pp. 11-17.

[5] E. P. Blasch, "SAR Image Exploitation Using an Information Filter Metric," *Intl Conf on Imaging Science, Systems, and Technology 2000,* Las Vegas, NV, June 2000.

[6] E. P. Blasch and Jim Schmitz, "Uncertainty Issues in Data Fusion", *National Symposium on Sensor and Data Fusion,* 2002.

[7] E. Blasch, "Data Association through Fusion of Target track and Identification Sets," *Fusion 2000,* Paris, France, 2000.

[8] E. Blasch, "Performance Metrics for Fusion Evaluation," Proc. *MSS NSSDF*, 2003.

[9] E. Blasch, "Information Fusion for Decision Making", In NATO ASI: *Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management*, 2004.

[10] E. P. Blasch and T. Connare, "Improving track accuracy through Group Information Feedback", *Fusion 2001*, 2001.

[11] E. P. Blasch and T. Connare, "Improving Track maintenance Through Group Tracking," *Proc of the Workshop on Estimation, Tracking, and Fusion; A Tribute to Yaakov Bar Shalom*, Monterey, CA, May 2001, 360 –371.

[12] E. P. Blasch, J. Hoffman, and J. Petty, "Defining a Fusion Gain – System Operation Characteristic Curve," *SPIE Int. Sym. On Aerospace/Defense Simulation & Control, Conf 4380 – Sensor Fusion X*, Orlando, FL, April 2001.

[13] S. K. Card, T. P. Moran & A. Newell *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum, Hillsdale, NJ, 1983.

[14] B. E. Fabian and E. P. Blasch, "Information Architecture for Actionable Information Production," *NSSDF*, 2002.

[15] D. L. Hall, *Mathematical Techniques in Multisensor Fusion*, Artech House, Norwood, MA ,1992.

[16] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*, SciTech Publishing, Inc, 2001.

[17] I. Kadar, K. C. Chang, K. O'Connor, and M. Liggins, "Figures-of-Merit to bridge fusion, long-term prediction and dynamic sensor management," *SPIE*, 2004.

[18] D. Kettani, & J. Roy, "A Qualitative Spatial Model For Information Fusion and Situation Analysis," SPIE 00.

[19] S. Plano and E. P. Blasch, "User Performance Improvement via Multimodal Interface Fusion Augmentation", *Fusion03,* 2003.

[20] P.M. Fitts and M. I. Posner, *Human performance*, Brooks, Cole, 1967

[21] J. Rasmussen, "Skills, rules and knowledge; signals, signs and symbols; and other distinctions in human performance models," *IEEE Trans. Syst., Man, Cybern., vol.* SMC-13, pp. 257-266, Jan. 1983.

[22] J. Rasmussen, *Information Processing and Human-Machine Interaction: Approach to Cognitive Eng.*, North-Holland, NY, 1986.

[23] T. D. Ross, J. Bradley, L. Hudson, M. O'Conner, "SAR ATR - So What's The Problem? - An MSTAR Perspective", Proc. SPIE, Vol. 3721, Algorithms for Synthetic Aperture Radar Imagery VI, April 1999.

[24] W. Stalling, *High-Speed Networks and Internets, Performance and Quality of Service*, 2^{nd} Ed., Prentice Hall, NJ, 2001.

[25] E. L. Waltz and J. Llinas, *Multi-Sensor Data Fusion*, Artech House, Norwood, MA, 1990.

[26] E. L. Waltz, "Data Fusion in Offensive and Defensive Information Operations," *NSSDF*, 2000.

[27] C. D. Wickens, *Engineering psychology and human performance.* Harper Collins, 1992.