Fusion Gain in Multi-Target Tracking

Stefano Coraluppi, Marco Guerriero, and Craig Carthel
Applied Research Department
NATO Undersea Research Centre
Viale S. Bartolomeo 400, 19126 La Spezia, Italy
coraluppi@nurc.nato.int, guerriero@nurc.nato.int, carthel@nurc.nato.int

Abstract - This paper introduces an information quality metric and a definition for fusion gain in multi-target tracking systems. We validate the reasonableness of these quantities and illustrate the relationship between fusion gain, scenario difficulty, and tracker effectiveness. The information quality metric is closely related to the information reduction factor, and has direct application to sensor management.

Keywords: Multi-target tracking, measures of performance, fusion gain.

1 Introduction

Subject to a miss-distance threshold, detector performance can be characterized fully by the target probability of detection, the false alarm rate, and the localization error. This performance characterization, while straightforward, does not generally lead to statistically-consistent performance evaluation. A proposed statistically-consistent approach that relies on a maximum likelihood formulation is discussed in [1].

Performance evaluation for multi-target trackers is more problematic due to the additional time dimension. Indeed, many performance metrics have been proposed in the tracking and fusion literature, and it appears that no set of metrics fully characterizes tracker performance. While most approaches account for a track-continuity requirement, a recently-proposed metric does not [2]. We propose here a simple information-quality metric that has several appealing qualities. In future work, it will be of interest to apply this metric to tracker-oriented sensor management.

Section 2 defines the information quality metric and fusion gain; related derivations are in appendices A-B. Sections 3-4 describe a simulation study aimed at validating that our definition of fusion gain is reasonable, and at relating fusion gain to scenario difficulty and tracker effectiveness. Section 5 summarizes our work and suggests future directions.

2 The information quality metric and fusion gain

Assume that detection-level statistics are given by the *probability of detection* (p_D), the *contact rate per scan* (n), and the *contact measurement error* (Σ). Given a scenario with n_t targets in a measurement space with covariance Σ_Z, we define detector performance as follows. (The approximation assumes $\Sigma_Z >> \Sigma$.)

$$f(p_D, n, \Sigma) \equiv \frac{\mathrm{tr}\left(\frac{p_D n_t}{n}\Sigma^{-1}\right)}{\mathrm{tr}\left(\Sigma_Z^{-1}\right)} \approx \frac{\mathrm{tr}\left(\frac{p_D n_t}{n}\Sigma^{-1} + \frac{n - p_D n_t}{n}\Sigma_Z^{-1}\right)}{\mathrm{tr}\left(\Sigma_Z^{-1}\right)}. \tag{1}$$

An interpretation for the metric is that it quantifies the average information content of an arbitrarily selected detection, under a single-target assumption. As such, it measures the *information quality* (IQ) provided by the detector. Information quality as defined here is dimensionless, thanks to the denominator term that quantifies the information content of a contact in the center of the sensor measurement space.

Correspondingly, assume that track-level statistics are given by the *track hold* (p_D^T), the *average number of tracks* (n^T), and the *track localization error* (Σ_T). Tracker performance under the IQ metric is given by:

$$f\left(p_D^T, n^T, \Sigma_T\right) = \frac{\mathrm{tr}\left(\frac{p_D^T n_t}{n^T}\Sigma_T^{-1}\right)}{\mathrm{tr}\left(\Sigma_Z^{-1}\right)}. \tag{2}$$

Note that n^T is the average number of tracks at any time; as such, it reflects false track duration, which the usual false track rate metric does not.

Further, the track-level statistics may be computed in a global or local fashion. Under the former, there is a single mapping of track to targets. Under the latter, the mapping is scan-based and does not reflect track continuity.

The term $\frac{p_D n_t}{n}$ in equation (1) can be thought of as the *probability of success*, the probability that an arbitrarily selected contact is target originated. The expression is correct under the assumption of a deterministic number of targets and of detector contacts. In the non-deterministic case, we may replace n_t and n with λ_t and $\lambda + p_D\lambda_t$, the expected number of targets and of contacts, respectively. (Note that following standard convention λ denotes the number of false returns.) The resulting expression is correct under a Poisson assumption; the reader is referred to further details in appendices A-B. The relationship between the IQ metric and the previously-introduced information reduction factor is discussed in appendix C.

The IQ metric has the following characteristics:

- It is a scalar metric that accounts for both detection and localization objectives;
- It is computed analogously at either to the tracker input or output, allowing for a direct computation of *fusion gain* as a ratio of IQ metrics;
- It accounts for false detection and track information (which the information-metric in [3] does not);
- It reflects the operationally-relevant notion of quality of information, rather than the total information as in [3];
- It is flexible in that it can *require* or *not require* track continuity, depending on whether track-level statistics are evaluated in a global or scan-based fashion.

Note that the IQ metric does not directly reflect track fragmentation.

There are two fundamental requirements for the IQ metric to be of interest in target-tracking applications, including sensor management. The first is that it have a meaningful interpretation with respect to data quality for the tracker input and output. In particular, information quality represents the information in detector (or tracker) data compared with a contact (or track) at the center of the sensor measurement space. A higher input IQ must lead to a higher output IQ for a given scenario and a given multi-target tracker. Second, the ratio of IQ metrics, the fusion gain, must be meaningful. In particular, it must reflect scenario difficulty and tracker effectiveness.

Thus, we are interested in examining the following:

- Is output IQ monotonically increasing as a function of input IQ?
- Does fusion gain decrease for increasingly challenging multi-target scenarios?
- Does fusion gain increase for increasingly sophisticated tracker implementations?

Subject to satisfactory responses to these questions, optimizing the detection-level IQ metric can be proposed as a paradigm for track-oriented sensor management. Further, note that the IQ metric applies both to trackers that include and those that do not include track labeling. We will use a classical multi-hypothesis tracker for our analysis [4]; alternatively, it would be of interest to explore as well a tracker of the latter sort, like the recently-proposed SJPDA [5].

3 Simulation framework

We base our statistical analysis on simulated positional measurement data from a single sensor with a fixed revisit rate. Contact simulation parameters are listed in table 1.

Table 1. Data simulation settings in Monte Carlo studies.

Parameter	Setting
number of Monte Carlo realizations	50
surveillance region	$(180\text{m})^2$
scenario duration	150sec
sensor revisit time	1sec
number of targets n_t	1 or 2
sensor p_D	1, 0.8, or 0.6
sensor data rate n	25 or 20
sensor localization error Σ	$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$, σ =2m, 1.5m, or 1m

We base our output tracker performance statistics on the output of a multi-hypothesis tracker that includes logic-based track management and track-oriented hypothesis management with a linear-programming (LP) relaxation approach to hypothesis pruning. This paradigm has been used successfully in ground, undersea, and maritime surveillance contexts [4]. Key tracker parameters are identified in table 2.

Table 2. Tracker settings in Monte Carlo study.

Parameter	Setting
track filter maneuverability index	$0.1\text{m}^2\text{s}^{-3}$
track filter prior velocity covariance	1m/sec
track correlation gate	99%
track initiation	4-of-4
track kill	4 misses
hypothesis tree depth	0 or 3
track classification distance threshold	2σ

While the tracker is assumed to have knowledge of sensor characteristics, it does not know the number of targets. Thus, evaluation of the input IQ metric in practical settings must use the *expected* number of targets, as mentioned in the previous section and as discussed in greater length in the appendices.

The input IQ metric is computed immediately by equation (1) using the appropriate values in table 1. Output performance requires a more involved procedure. For each realization of tracker output, we determine a global mapping of tracks to truth using the distance threshold (see table 2), and subsequently compute track hold, average number of tracks, and track localization. The results are averaged over all realizations in the Monte Carlo study. That is, we first determine estimates of p_D^T, n^T, and Σ^T, and subsequently evaluate output IQ by equation (2).

Note that, for a surveillance region of size $Z = X \cdot Y$ [m^2] and assuming positional measurements, in equations (1-2) we have the following:

$$\mathrm{tr}\Sigma_Z^{-1} = \frac{12}{X^2} + \frac{12}{Y^2} . \qquad (3)$$

Also, note that in equation (2), assuming identical errors in both dimensions and letting σ_T be the average track positional error, we have:

$$\mathrm{tr}\Sigma_T^{-1} = \left(\frac{\sqrt{2}}{\sigma_T}\right)^2 + \left(\frac{\sqrt{2}}{\sigma_T}\right)^2 = \frac{4}{\sigma_T^2} . \qquad (4)$$

4 Simulation results

Our simulation study is intended to address the three issues posed in section 2 regarding the characteristics of input IQ, output IQ, and fusion gain. We now address these in turn.

All the simulation realizations shown in the figures use the same color code convention: target trajectories are in magenta, target contacts are magenta dots, false contacts are black dots, true tracks are blue, and false tracks are red.

4.1 Monotonicity in output IQ

We examine a single-target scenario, with initial position of (-75m, 5m) and fixed velocity (1m/s, -0.067m/s). We consider three combinations of sensor performance characteristics (p_D, n, and σ), leading to three input IQ values. For each, we run a separate Monte Carlo study; the resulting estimates of p_D^T, n^T, and Σ^T are used in assessing output IQ. Results are given in table 3, and a realization is illustrated in figure 1.

Table 3. Single-target performance results as a function of sensor characteristics.

Scenario	Input IQ	Output IQ	Fusion gain
p_D=0.6, n=25, σ=2m	16.20	340.77	21.04
p_D=0.8, n=20, σ=2m	27.00	1,057.66	39.17
p_D=0.8, n=20, σ=1.5m	48.00	2,171.07	45.23

The results in table 3 provide evidence that input and output IQ are meaningful representations of detector and tracker performance, respectively. Indeed, higher input IQ leads to higher output IQ. Note however that, as seen here, fusion gain is generally not fixed for a given ground truth scenario and tracker implementation. This may partly depend on the fact that we have not tuned our tracker to input data detection characteristics; had we done so, fusion gain might be more consistent across a range of detector characteristics, thus suggesting that it should be solely a function of scenario difficulty and tracker effectiveness.

4.2 Fusion gain dependence on scenario difficulty

We now fix the sensor detection and localization characteristics (p_D=0.8, n=20, and σ=1.5m), and examine tracker performance for a number of scenarios of varying difficulties:

- Low difficulty – single target with initial position of (-75m, 15m) and fixed velocity of (1m/s, -0.2m/s);
- Intermediate difficulty – two crossing targets with initial positions of (-75m, 15m) and (-75m, -15m) and fixed velocities of (1m/s, -0.2m/s) and (1m/s, 0.2m/s), respectively;
- High difficulty – two slowly-crossing targets with initial positions of (-75m, 5m) and (-75m, -5m) and fixed velocities of (1m/s, -0.067m/s) and (1m/s, 0.067m/s), respectively.

Results are given in table 4, and a few realizations are illustrated in figures 2-5. We see that there is progressively lower fusion gain with increasing scenario difficulty. This suggests that fusion gain as defined here is a meaningful representation of pre-tracker to post-tracker performance improvement.

Table 4. Tracking performance results as a function of scenario degree-of-difficulty.

Scenario	Input IQ	Output IQ	Fusion gain
single target	48.00	2,171.07	45.23
crossing targets	96.00	1,146.36	11.94
slowly-crossing targets	96.00	883.56	9.20

4.3 Fusion gain dependence on tracker effectiveness

Finally, we fix (idealized) sensor detection and localization characteristics (p_D=1, n=2, and σ=1m) as well as the scenario (slowly-crossing targets, as above). We examine performance for two choices of hypothesis tree depth (*n-scan*); results are in table 5. (Note that the results in previous sections were based on n-scan=0.)

Table 5. Tracking performance results as a function of tracker complexity.

Scenario	Input IQ	Output IQ	Fusion gain
n-scan=0	2,700.00	3,001.44	1.11
n-scan=3	2,700.00	3,290.77	1.22

Interestingly, given the high input IQ, fusion gain is close to unity. We see that fusion gain increases with increasing tracker complexity; this provides further confidence in the notion of fusion gain as defined in this work.

It is instructive to examine the estimates of p_D^T, n^T, and σ_T that give rise to the results in table 5. In particular, we have the following:

- *n-scan*=0: p_D^T=0.52, n^T=1.99, σ_T=0.684m ;
- *n-scan*=3 : p_D^T=0.58, n^T=1.99, σ_T=0.690m

These numbers illustrated why fusion gain is difficult to achieve in high-difficulty scenarios, particularly where the input IQ is already high. Indeed, though p_D=1, the output track hold is only slightly above 0.5, due to numerous track-swap occurrences that lead to false track classification and adversely affect the track hold metric. An illustration of this phenomenon is given in figures 6-7.

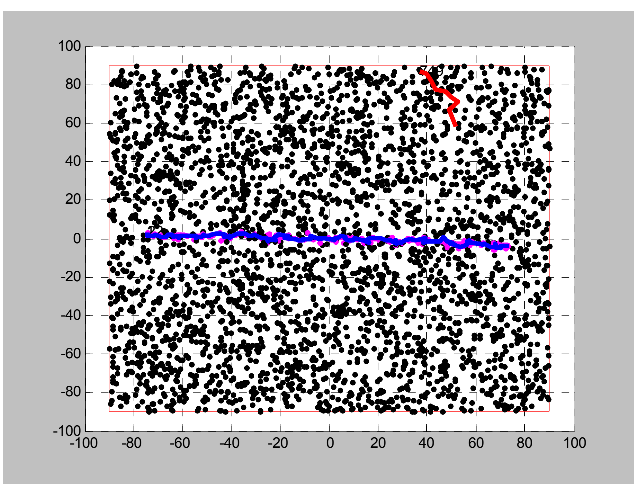

Figure 1. One realization of the single-target scenario (case 3 – best-performing sensor).

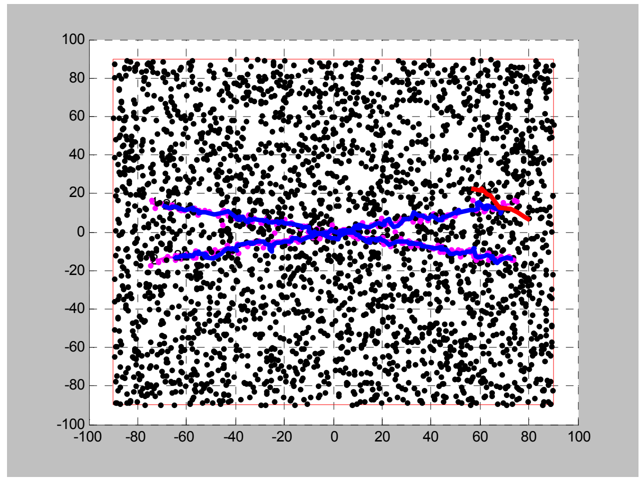

Figure 2. One realization of the crossing-targets scenario.

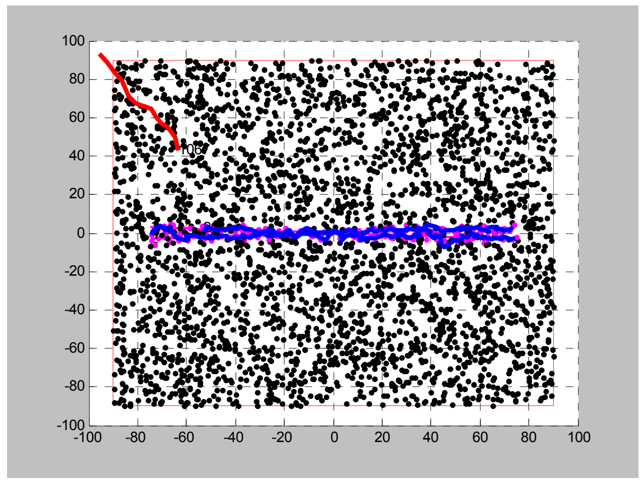

Figure 3. One realization of the slowly-crossing targets scenario.

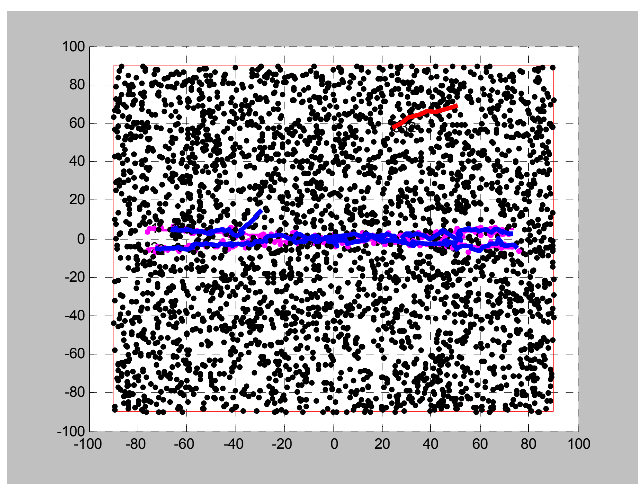

Figure 4. Another realization of the slowly-crossing targets scenario.

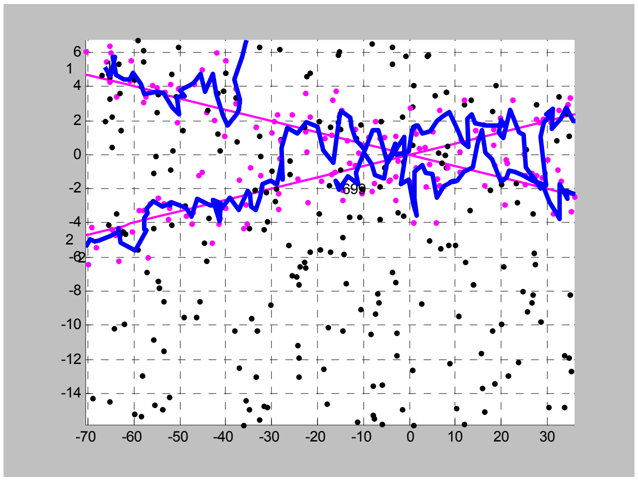

Figure 5. Close-up of target crossing in figure 4: track fragmentation due to target proximity.

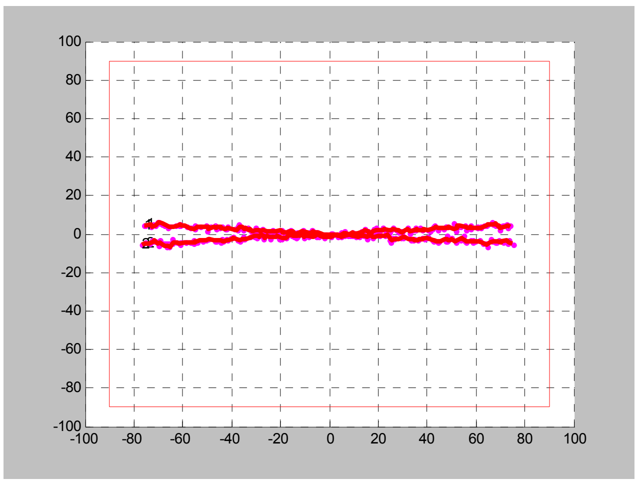

Figure 6. A track swap example, with ideal sensor data.

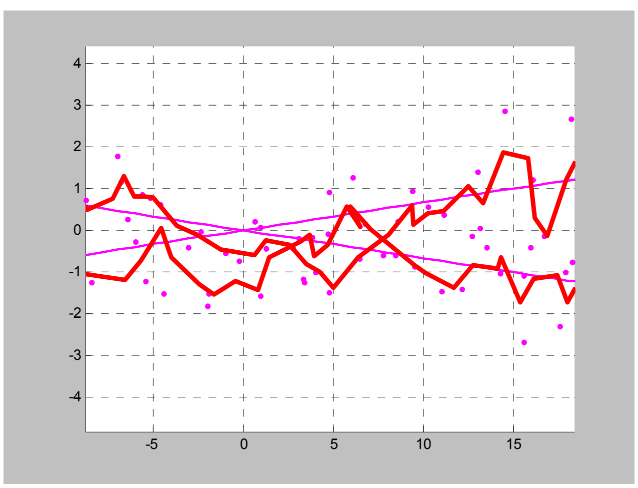

Figure 7. Close-up on the track swap in figure 6.

5 Conclusion

This paper introduces an information quality metric and a definition for fusion gain in multi-target tracking systems. We validate the reasonableness of these quantities and illustrate the relationship between fusion gain, scenario difficulty, and tracker effectiveness. Additionally, we study the relationship between the information quality metric and the previously-introduced information reduction factor.

The IQ metric has many appealing features. Principally, it couples detection and localization performance and can be applied analogously at the tracker input or output. We have seen in section 4.1 that, for a fixed scenario and tracker instantiation, higher input IQ leads to higher output IQ. Thus, we would suggest that the metric can be fruitfully applied in sensor management application: sensor selection based on input IQ optimization should correspondingly lead to good performance in terms of output IQ.

As a side note, it would be of interest to examine the reference, *dumb* detector or tracker used to characterize the denominator in equations (1-2). It would be interesting to consider an improved, possibly non-deterministic, *a priori* solution that accounts for a non-unity expected number of targets λ_t.

References

[1] C. Carthel, S. Coraluppi, P. Willett, M. Maratea, and A. Maguer, Maximum Likelihood Approach to HF Radar Performance Characterization, in *Proceedings of the 12th International Conference on Information Fusion*, Seattle WA, USA, July 2009.

[2] D. Schuhmacher, B.T. Vo, and B.N. Vo, On Performance Evaluation of Multi-Object Filters, in *Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, July 2008.

[3] O. Erdinc, P. Willett, and S. Coraluppi, Sonobuoy Placement for Optimal Multistatic Detection and Localization, *ISIF Journal of Advances in Information Fusion*, vol. 2(1), June 2007.

[4] S. Coraluppi and C. Carthel, Multi-Stage Data Fusion in Military Surveillance Systems, in *NATO RTO SET-133 Lecture Series on Multistatic Surveillance and Reconnaissance: Sensor, Signals and Data Fusion*, April 2009.

[5] L. Svensson, D. Svensson, and P. Willett, Set JPDA Algorithm for Tracking Unordered Sets of Targets, in *Proceedings of the 12th International Conference on Information Fusion*, Seattle WA, USA, July 2009.

[6] J. F. C. Kingman, *Poisson Processes*, Oxford University Press, 1993.

[7] Y. Bar-Shalom, X. Li., *Multitarget-Multisensor Tracking: Principles and Techniques,* Storrs, YBS Publishing, 1995.

[8] X. Zhang, P. Willett, Y. Bar-Shalom, Dynamic Cramer-Rao Bound for Target Tracking in Clutter, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4 October 2005.

A The Case of Poisson Targets and False Returns

Assume that the number of targets n_t obeys the Poisson distribution with mean λ_t, that is:

$$p(n_t = i) = \frac{\lambda_t^i}{i!} e^{-\lambda_t} . \tag{5}$$

Assume further that the number of false returns is Poisson with mean λ. As before, assume each target is observed with probability p_D.

The probability that a contact is target originated is given by:

$$p(\text{success}) = \frac{p_D \lambda_t}{\lambda + p_D \lambda_t} . \tag{6}$$

In deriving the above result, we invoke the *merging* and *thinning* properties of the Poisson process [6].

B The Case of One Target and Poisson False Returns

We evaluate here the prior probability that a contact is target originated, under the assumption of a single target and of a Poisson distributed number of false returns with mean λ.

Conditioned on the total number of contacts n_c, the probability of a contact *i* being target originated is denoted by $\pi_i(n_c)$ and is given by [7]:

$$\pi_i(n_c) = \frac{p_D}{p_D n_c + (1-p_D)\lambda}, \qquad i = 1, \ldots n_c , \tag{7}$$

$$\pi_i(n_c) = \frac{(1-p_D)\lambda}{p_D n_c + (1-p_D)\lambda}, \qquad i = 0 , \tag{8}$$

where *i=0* denotes the event that none of the contacts is target originated.

The unconditional probability π (constant for all $i \geq 1$) of a contact being target originated, is obtained by averaging $\pi_i(n_c)$ over the probability mass function of n_c, which is given by:

$$P(n_c = m) = p_D \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} + (1-p_D) \frac{\lambda^m}{m!} e^{-\lambda} . \tag{9}$$

Thus, we have:

$$\begin{aligned} q_3 &= \sum_{m=1}^{\infty} \frac{p_D}{p_D m + (1-p_D)\lambda} \left(p_D \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} + (1-p_D) \frac{\lambda^m}{(m)!} e^{-\lambda} \right) \\ &= p_D \left(\frac{1-e^{-\lambda}}{\lambda} \right) \approx \frac{p_D}{\lambda} \quad as \ \lambda \to \infty \end{aligned} \tag{10}$$

The expression in equation (10) differs from the *probability of success* as obtained in the deterministic case (section 2) and the Poisson case (appendix A). Nonetheless, in the limit of a large number of false returns, the expressions are equivalent.

C Information Quality and Information Reduction Factor

In the case where the number of targets is known and equal to one, fusion gain is defined as:

$$f\left(p_D, \lambda, \Sigma\right) \equiv \frac{\mathrm{tr}\left(p_D \left(\frac{1-e^{-\lambda}}{\lambda} \right) \Sigma^{-1} \right)}{\mathrm{tr}\left(\Sigma_Z^{-1}\right)} . \tag{11}$$

The numerator in equation (11) is approximately equal to the trace of the *Fisher Information Matrix* (FIM) J_Z which is given by [8]:

$$J_Z = q_2 \Sigma^{-1} . \tag{12}$$

where q_2 is the information reduction factor (IRF) which depends on the probability of detection, the density of false contacts, and the measurement noise covariance.

Unfortunately, the expression for q_2 cannot be expressed in closed form: its evaluation requires heavy numerical integration. We prefer instead q_3, which has a simple interpretation as well: it represents the prior probability that a measurement is target originated given that there are false contacts with density λ and a true detection with probability p_D.

As a sanity check on the similarity in these factors, we see that $q_3 = p_D$ for $\lambda = 0$ and $q_3 = 0$ for $\lambda \to \infty$. Figure 8 motivates the use of q_3 as a new information reduction factor by showing promising accuracy of the approximation.

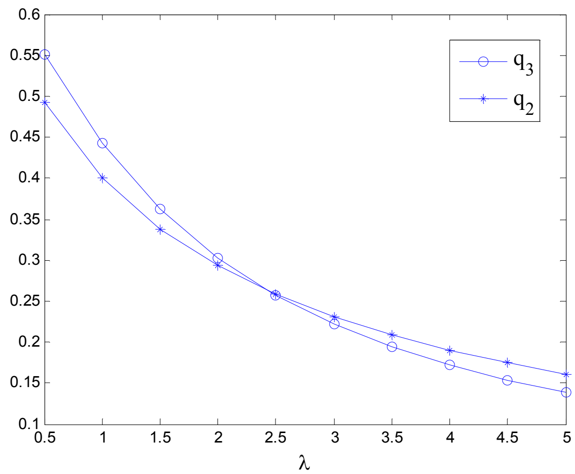

Figure 8. The dependence of the IRF q_2 and q_3 on average number of false contacts for various detection probabilities. The plot assumes a 2-dimensional measurement vector with g=4 sigma gate and $p_D = 0.7$.