VALIDATION OF REQUIRED SURVEILLANCE PERFORMANCE (RSP) ACCURACY

John W. Andrews, Steven D. Thompson
M.I.T. Lincoln Laboratory, Lexington, MA 02420-9185

ABSTRACT

Required Surveillance Performance (RSP) offers a framework in which new surveillance systems can be approved for use when they meet operationally relevant performance standards without regard to the particular technologies employed by the system.

This paper is concerned with the problem of validating that a proposed system meets RSP measurement accuracy standards. An established form for an accuracy test is the Close Approach Probability (CAP) criterion. When an absolute level-of-safety metric is employed, the CAP test requires determining the probability density that lies far into the tails of the separation error distribution. This paper demonstrates that the inability to know with certainty the form of the error distribution is a formidable obstacle to validating such a criterion. However, when a relative safety approach is employed, validation is less demanding. In this case we show that a simple alternative formulation known as the variance ratio test is equivalent to the CAP test.

Introduction

ICAO Context: Reference System and TLS Approach

The International Civil Aviation Organization (ICAO) has defined two basic approaches to safety validation [1]. They are known as the *reference system* approach and the *target-level-of-safety (TLS)* approach. In the reference system approach, the new system is compared to a similar reference system that is already in use and has been accepted as providing adequate safety. In the target-level-of-safety approach, the safety level of the new system is calculated directly from models and data. The reference system approach has the advantage that it does not require complete modeling of all the factors that impact safety. However it cannot be applied when the new system is substantially different from the old.

ICAO standards call for the safety level of the ATM system as a whole to be better than one fatal event per 10 million operations (or approximately one fatal accident per 1.0E-07 flight hours). Since there are many possible causes of a fatal accident, the rate allocated to any one cause must be a small fraction of this rate. One study [6] has suggested that the failure rate assigned to surveillance position measurement error be 2.0E-12 per surveillance report, which is an extremely demanding requirement.

Required Surveillance Performance (RSP)

The concept for the Next Generation Air Transportation System (NGATS) [2] calls for development of a new standards framework referred to as Required Total System Performance (RTSP). RTSP will involve a set of performance requirements covering communication, navigation, and surveillance. The navigation component, Required Navigation Performance (RNP) is already being applied to some new services. The surveillance component, Required Surveillance Performance (RSP). is less mature. Under RSP, the characteristics of a surveillance system (such as accuracy, integrity, availability, etc.) would be defined in terms of a number of RSP levels. The type of ATM services supported by each level would also be specified. Then a new system - regardless of the technology it employed - would qualify for providing the service if it could demonstrate that it achieved the requisite RSP level. This approach has the advantage of allowing new systems to more readily be brought into service without the need to rewrite specifications for new technologies.

A principal issue in RSP is the ability of new systems to achieve the required safety standards. Sound techniques for safety validation will be important to the success of the RSP approach.

The Close Approach Probability (CAP) Criterion

An accuracy criterion is a fundamental element of an RSP formulation. A conventional form for such a criterion is the Close Approach Probability (CAP) test based on the geometry shown in Figure 1. Here, two aircraft are separated by a standard separation S_0. The diameter of the smallest circle that contains the aircraft is A_W. This is taken as the area that must not overlap to avoid collision.

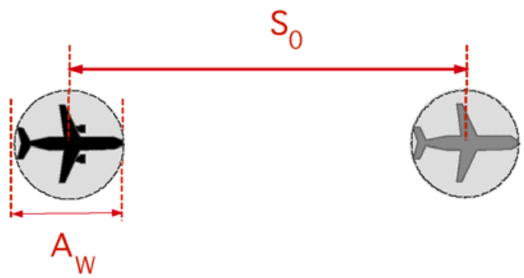

Figure 1 Separation geometry for the CAP test

The CAP criterion requires determination of the probability of close approach, P_{CA}, defined as the probability that the two aircraft will actually be close enough to collide even though their measured separation is S_0. Figure 2 shows that P_{CA} is related to $f_S(S_0)$, the value of the separation error probability density function (pdf) at S_0. At this error, a measured separation of S_0 corresponds to an actual separation of zero.

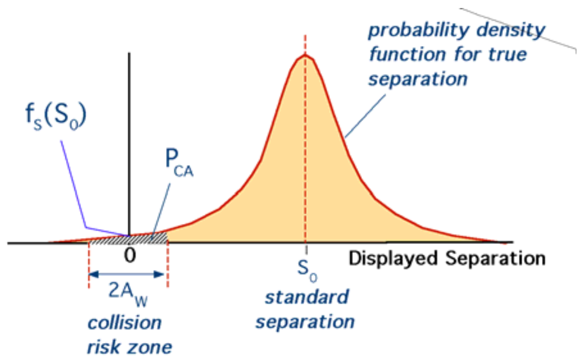

Figure 2 P_{CA} is the probability that the true separation is within the collision risk zone despite a measured separation of S_0.

Because the probability density function does not vary greatly over the distance $2A_W$, we can use the pdf value at S_0 as the approximate average pdf value in the collision risk zone. Then

$$P_{CA} \;=\; 2A_W \int_{-\infty}^{\infty} g_1(y) g_2(y - S_0)\, dy \;=\; 2A_W f_S(S_0)$$

where g_1 and g_2 are the position measurement error pdf's for each aircraft considered individually. It should be noted that in most actual systems the absolute position errors for the two aircraft are correlated because both position measurements have common errors such as radar registration errors or propagation errors from navigation satellites. In such a case the common error component must be removed from the constituent pdf's before applying the above equation.

VALIDATION USING THE REFERENCE SYSTEM APPROACH

This section will discuss the validation of a CAP requirement under the reference system approach. The reference system approach attempts to establish that a new system performs as good as or better than a reference system with regard to key metrics.

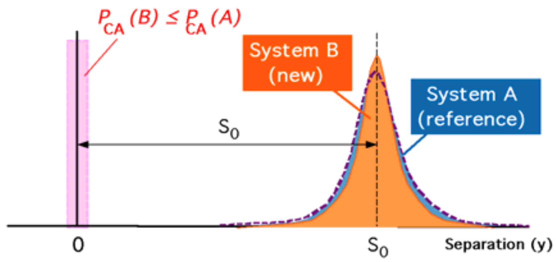

Figure 3 CAP test for reference system approach

As shown in Figure 3, a new System B must pass the following test:

$$Accept B if\ P_{CA}(B) \le P_{CA}(A); otherwise reject.$$

Because P_{CA} is proportional to the value of the separation error pdf at S_0, the required acceptance test is equivalent to

$$Accept\ B\ if\ \ \frac{f_B(S_0)}{f_A(S_0)} \le 1; \quad otherwise\ reject.$$

In evaluating the pdf's used in this test, we face an immediate problem: Because the failure is a rare event for all separation standards of interest, test data is likely to include very few measurements that fall into or near the hazard region. The great bulk of data will come from the central region of the distribution. Extrapolation from this central region to a point far into the tails of the distribution can be justified only if there is a scientific or engineering principle that dictates the rule to be used for this extrapolation. Unfortunately, for most real-world systems there is no such principle.

The Variance Ratio Test

We will now demonstrate that a simplified acceptance test based upon the ratio of error variances can be applied when using the reference system approach. The test hinges upon the definition of similarity between the reference system and the new system. We will assume that if two systems are similar, their error distributions have a similar form. We will offer a proof of the variance ratio test for distributions that are from the exponential family of distributions. The test is valid for other forms, although we shall not discuss them here.

A Generalized Exponential Probability Distribution

When probability modeling is used to demonstrate satisfaction of an RSP standard, it is necessary to justify the employed forms of error probability distributions. Common forms used for the probability density functions are the double exponential and the Gaussian. But these two distributions are just special cases of a family of exponential distributions that could be employed. There is no reason why these two should be best given the many factors that can contribute to surveillance errors. To improve the robustness of the modeling, we will now introduce a general exponential form that includes both the double exponential distribution and the Gaussian distribution as special cases. This allows actual data to be used to find the best fitting form among the family of exponential distributions. This also provides the flexibility needed to avoid unintentional and arbitrary selection of parameters.

The pdf for the *exponential family* of distributions can be written in the following way:

$$f(y) \;=\; \frac{a}{\sigma}\exp\left[-\left(\frac{|y-\mu|}{\sigma b}\right)^k\right] \quad , \quad -\infty \le y \le \infty$$

where

$$a \;=\; \frac{k}{2b\Gamma(1/k)} \quad , \quad b = \sqrt{\Gamma(1/k)/\Gamma(3/k)}\,,$$

and the gamma function is defined as

$$\Gamma(k) = \int_0^\infty t^{k-1}\exp[-t]\,dt, \quad k > 0.$$

The resulting pdfs are plotted for several values of k in Figure 4. Here k=1 produces a double exponential distribution and k=2 produces a Gaussian distribution. The greater the value of k, the more rapidly the pdf decreases in the tails of the distribution.

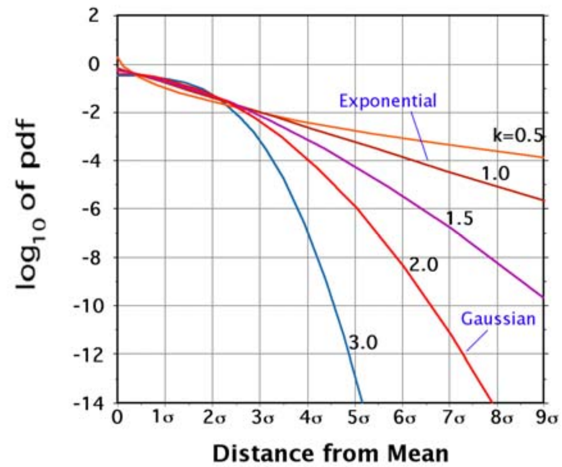

Figure 4 A family of exponential probability distributions

The Variance Ratio Test

We will now apply the CAP test using a reference system approach. We will assume that the pdfs for the two compared systems are both drawn from the general exponential family described earlier. This assumption is less restrictive than simply assuming a textbook form for the distribution. We will further assume that the existence of similarity allows us to use the same value of k for the two systems being compared.

As shown earlier, the acceptance test hinges upon the ratio of the separation error pdf's at S_0. The acceptance criterion can be written

$$\frac{f_B(S_0)}{f_A(S_0)} \;=\; \frac{\sigma_A}{\sigma_B}\exp\left[\left(\frac{|S_0-\mu|}{b\sigma_A}\right)^k - \left(\frac{|S_0-\mu|}{b\sigma_B}\right)^k\right] \;\le\; 1$$

Or, in logarithmic terms,

$$\ln\left[\frac{\sigma_A}{\sigma_B}\right] \;+\; \left(\frac{|S_0-\mu|}{b}\right)^k\left(\frac{1}{\sigma_A^k}-\frac{1}{\sigma_B^k}\right) \;\le\; 0$$

Letting $\rho = \sigma_B/\sigma_A$, this can be written

$$\frac{|S_0-\mu|}{\sigma_A} \;\ge\; b\rho\left[\frac{|\ln\rho|}{|\rho^k-1|}\right]^{1/k} \quad if\ \rho \le 1$$

$$\frac{|S_0-\mu|}{\sigma_A} \;\le\; b\rho\left[\frac{|\ln\rho|}{|\rho^k-1|}\right]^{1/k} \quad if\ \rho > 1$$

Now the quantity on the lefthand side of the above inequalities is the number of standard deviations from the mean at which the reference system pdf is evaluated for computing P_{CA}. The test thresholds are plotted in Figure 5. Because of the low risks permitted in aviation, the evaluation must occur in the tails of the distribution and the ordinate value is unlikely to be less than about 4 or 5. As shown in the figure, the inequality for $\rho \leq 1$ is always satisfied in the tails of the distribution. For $\rho > 1$, the inequality is satisfied only when the separation standard is so small that P_{CA} is evaluated near the central peak of the distribution - a situation that would never arise if the reference system were operating at an acceptable level of safety to begin with.

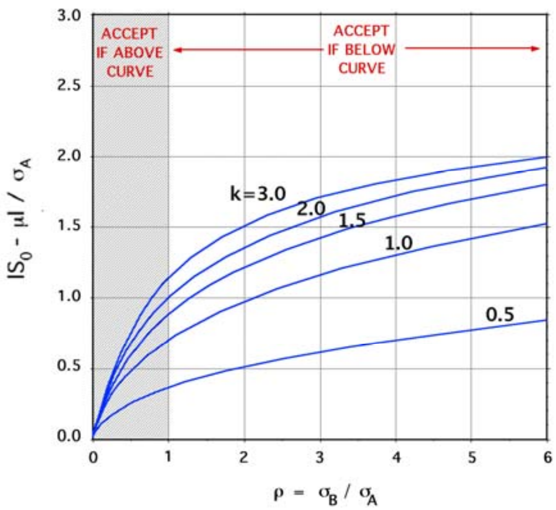

Figure 5 Acceptance test threshold

This analysis shows that when similarity between two systems allows us to assume similar forms for the error pdfs, the acceptance test depends entirely on whether σ_B is is less than or greater than σ_A. We can write this in terms of the variances as follows:

$$Accept\ B\ if\ \ \frac{\sigma_B^2}{\sigma_A^2} \leq 1; \quad otherwise, reject.$$

This *variance ratio test* is intuitively reasonable since the variance is a statistic that places increasing weight upon larger deviations from the mean, and for reasonable forms, we do not expect one distribution to have a larger variance than another while simultaneously having a smaller pdf in the tails of the distribution.

The variance ratio test is applicable to all the generalized exponential forms, including commonly used exponential and Gaussian distributions. Further mathematical investigation would undoubtedly define a wider range of pdf's that satisfy the assumptions of the test.

VALIDATION USING TARGET LEVEL OF SAFETY APPROACH

As noted earlier, when a suitable reference system cannot be found, it is necessary to qualify a system by showing that it meets an absolute risk threshold known as the Target Level of Safety (TLS). ICAO guidance [2] notes that the TLS analysis should consider all elements of the separation assurance process, including communication, navigation, surveillance, and procedures. A common practice is to allocate the total risk among all possible faults and to analyze the faults separately. One such fault is separation failure caused by surveillance inaccuracy. The Closest Approach Probability (CAP) test is one approach to defining surveillance accuracy requirements. While not intended to be a realistic model of the entire separation process, it nevertheless characterizes one way in which surveillance error could contribute to loss of separation.

We will now focus upon the question of validation of a TLS requirement using the CAP criterion. In [6], the allowable risk threshold allocated to failure of the CAP criterion is 2.0E-12. This implies that there be only one measurement in each 500 billion measurements that would make aircraft in the collision risk zone appear to be properly spaced. The implications of such a stringent requirement will be discussed later. For the moment, we will merely look at the question of how we can validate that a new surveillance system (System B) satisfies the CAP criterion.

CAP Criterion for TLS Approach

As shown earlier, the application of the CAP test to determine a TLS requires an absolute value for P_{CA}. This requires evaluating the pdf of the separation error distribution at a point in the tails of the distribution. Direct validation through compilation of an error histogram is usually infeasible due to the prohibitive amount of data required to establish confidence in the P_{CA} value found in this way. Instead, the usual approach is to assume a form for the probability distribution for the separation errors, estimate the parameters of the distribution, and use the resulting pdf to calculate P_{CA}. However, an issue in this approach is validating the assumed form of the pdf.

In modeling radar errors, the most commonly used forms for the error pdf are the Gaussian and the double exponential distributions [4][6]. But it is important to acknowledge that nothing requires that the real world system to choose its error distributions from textbooks. The distributions may be combinations of textbook distributions or may be entirely new distributions created by the particular characteristics of the hardware and software in use.

For example, Figure 6 provides a histogram of separation errors produced by a data-driven surveillance error model based upon flight test data and modeling of display errors [5]. A casual inspection reveals peculiar features such as a narrow central peak, a number of smaller local peaks, and an extended tail. These features are produced by the error properties of electronic processes, asynchronous measurements, propagation, etc. No textbook distribution can replicate this probability distribution exactly. Even if one decides that RSP standards can focus solely on approximate fit (or an upper-bound fit) to the tails of the distribution, one is left with the question of the accuracy with which model parameters can be determined.

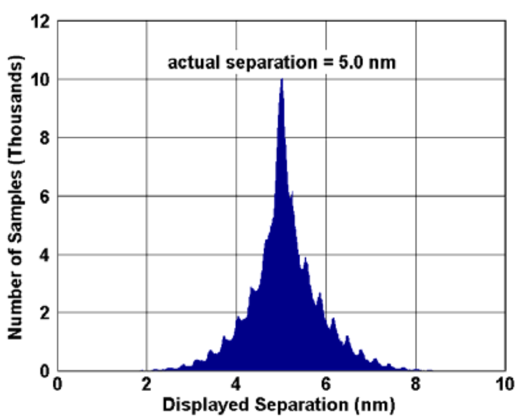

Figure 6 Displayed separation error for long-range sliding window radar tracking (σ=0.81 nm)

To better fit observed data for one particular long-range radar, Nagaoka and Amai [4] postulated a composite distribution consisting of both a Gaussian and exponential term. We will discuss composite distributions in the next section.

A further consideration is that P_{CA} must be evaluated in the tails of the error distribution. Often P_{CA} is evaluated at 6 or 7 standard deviations from the mean. Although the Gaussian form is sometimes applied with little thought, it is well-known that real world distributions for complex technical systems are seldom, if ever, Gaussian in the tails. The principal reason for this is that the Gaussian form is assured by the Central Limit Theorem only when the tails are produced by the chance addition of a large number of comparable error sources. Unfortunately, for real-world technical systems, the tails are usually produced by only a handful of rare, distinct failure modes. The specific statistical characteristics of those failure modes dominate the tails of the distribution. There is no *a priori* principle or theory that suggests a single form for the error probability distribution.

Composite Probability Distributions

Real world error distributions are often due to a variety of error sources, and the form of the distribution may change with time due to factors such as equipment calibration, atmospheric propagation, etc.. An appropriate error probability model is such cases is a *composite distribution* that consists of a sum of N terms as follows:

$$f(x) = \sum_{i=1}^{N} w_i f_i(x) \quad where \quad \sum_{i=1}^{N} w_i = 1$$

and each $f_i(x)$ is assumed to be a properly formulated pdf in its own right. One way of interpreting this form is that the surveillance system samples can be collected under N distinct sets of conditions, each condition occurring with probability w_i and having its own unique pdf, $f_i(x)$. This composite form is readily applied to a fault-based safety analysis in which each fault produces a different error pdf.

When N=1, we have only one term and the composite form is equivalent to the simple single-form model. It should be noted that as N goes to infinity, the Central Limit Theorem does *not* force the composite pdf to converge to a Gaussian distribution. The Central Limit Theorem does not apply because we do not add the errors from the different distributions simultaneously in order to produce a single random variable. Instead, we switch between distinct pdf's in a random way.

The mean and variance for the composite distribution can be determined as follows:

$$E[X] = \sum_{i=1}^{N} w_i E_i[X]$$

$$E[X^2] = \sum_{i=1}^{N} w_i E_i[X^2]$$

$$\sigma^2 = E[X^2] - E^2[X] = \sum_{i=1}^{N} w_i \left[E_i\,[X^2] \right] - \left(\sum_{i=1}^{N} w_i E_i\,[X] \right)^2$$

where the notation $E_i[Y]$ means that the expectation is taken using the i^{th} component pdf.

Since

$$E_i[X^2] = \sigma_i^2 + \mu_i^2$$

The variance of the composite distribution is

$$\sigma^2 = \sum_{i=1}^{N} w_i \sigma_i^2 + \sum_{i=1}^{N} w_i \mu_i^2 - \left(\sum_{i=1}^{N} w_i \mu_i \right)^2$$

Note that the last two terms correspond to the variance of the means when we view the weights w_i as corresponding to discrete probabilities for the occurrence of the random variables σ_i and μ_i. Thus, the *variance of the composite distribution is the average of the variances of the component pdfs plus the variance among the means themselves.* This is an intuitively reasonable result that suggests how the components contribute to the variance of the composite pdf.

If all variances are zero, then the variance of X is simply the variance of the means. If all means have the same value (e.g. all means are zero), then there is no variance among the means and the composite variance is just the weighted average of the component variances.

Note that if one of the composite pdf's has a very small probability w_i, it will contribute very little to the composite variance even if its own variance is manyfold greater than that of other terms. Thus, a quantity of data sufficient for accurate determination the variance of the composite distribution may provide few clues as to the presence of high-variance terms. But, as we shall see later, such terms are important because they can quickly become dominant in the tails of the composite distribution.

Separation Standard for Composite Distributions

We will now look at the implications of the composite distribution for the specific problem of determining the separation standard required to achieve a specified value of P_{CA}. Table 2 defines four different composite pdf's. The first is a pure Gaussian and the second a pure double exponential distribution. The third is a Gaussian/Double Exponential composite and the fourth a Gaussian/Uniform composite. Uniform distributions can occur when errors are due to asynchronous position measurements and they present an interesting challenge to validation. All distributions have zero mean. The Gaussian and exponential pdfs all have σ=0.228 nm (a value appropriate for monopulse radar surveillance). Figure 7 provides a plot of the achieved P_{CA} as a function of the separation standard for the four distributions.

Table 2 Composite distributions examined

Case	p_i	Pdf Forms
1	1.0000	Gaussian
2	1.0000	Double Exponential
3	0.9999 0.0001	Gaussian Double Exponential
4	0.99999 0.00001	Gaussian Uniform, U[-4,4]

Inspection of these curves leads to the following observations:

• The separation that achieves the required value of P_{CA} (2.0E-12) differs significantly for the four curves. From smallest to largest, it is 1.6 nm for the pure Gaussian, 2.5 nm for the Gaussian/exponential distribution, and 4.0 nm for the pure exponential and Gaussian/uniform.

• The Gaussian/exponential composite distribution is almost identical to the Gaussian out to 1.2 nm. At that point, the exponential term begins to dominate and the pdf asymptotically assumes the shape of the exponential distribution displaced downward by its weighting factor of p_2= 0.0001.

• At the separation standard suggested by the pure Gaussian curve (1.6 nm), the composite Gaussian/exponential distribution has a risk that is 430 times greater.

• The Gaussian/uniform distribution cannot achieve the targeted P_{CA} until the upper limit of the uniform distribution is reached at 4.0 nm. Beyond that that point, the effect of the uniform component disappears and the P_{CA} drops to be essentially the same as that of the pure Gaussian.

The uniform distribution exerts considerable influence out to its cut-off point, which can be shown to be $x = \mu - \sigma\sqrt{3}$. Acceptance in terms of absolute safety is unlikely if a uniformly distributed error component is non-zero at the separation standard.

From the above it can be concluded that even a slight contamination of a pure Gaussian form by an exponential or uniform distribution can have very significant effects upon the required separation standard. This is because the minority components can become dominant in the tails of the composite distribution.

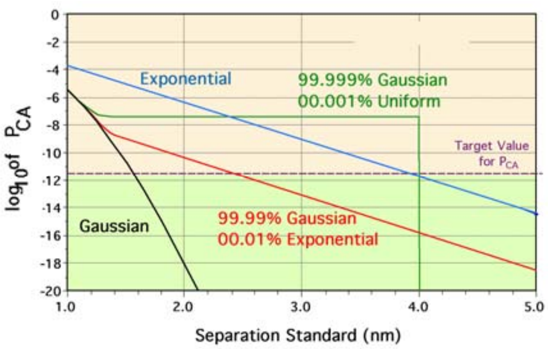

Figure 7 P_{CA} results for composite distributions with same variances.

Gaussian Components with Different Variances

We will now examine a case in which the constituent terms in the composite distribution are all Gaussian, but possess different variances. Such a composite form could arise when there are different special conditions in which the surveillance system performance is degraded. It is worth asking how rare such cases must be in order to be ignored. Figure 8 shows the P_{CA} values produced for single-term Gaussian pdf's (N=1) of different variances. It can be seen that increasing the standard deviation by a factor of three or four has dramatic effects on the value of P_{CA}.

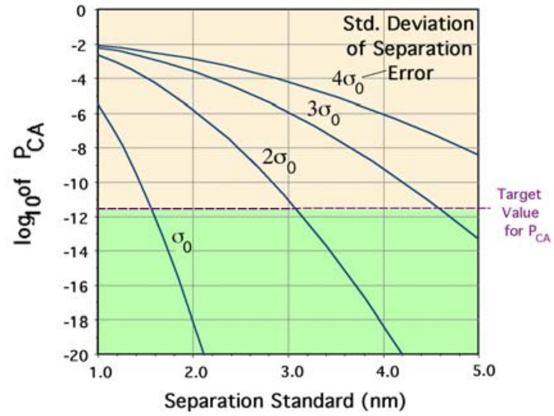

Figure 8 P_{CA} results for pure Gaussian distributions with different variances.

Figure 9 shows the effect of adding a second constituent pdf with a standard deviation that is three times that of the original. It can be seen that the second constituent has a dramatic effect on the required separation standard. When present at only a probability of one-thousandth of a percent (0.00001), it shifts the required separation standard from 1.6 nm to 3.2 nm. The presence such a component would be difficult to detect unless very large data sets are available for analysis.

The examples presented here lead to the following conclusion: In order to apply the probability modeling approach to the determination of a separation standard, one must verify that the distribution does not include composite components with larger tails (such as exponential or uniform forms) or components with standard deviations that are significantly greater than the most common term. Such components can occur at probabilities as low as 1 part in a million and still invalidate the calculation of the separation standard. Validation at this level may be impractical due to the volume of data required for confidence in the result.

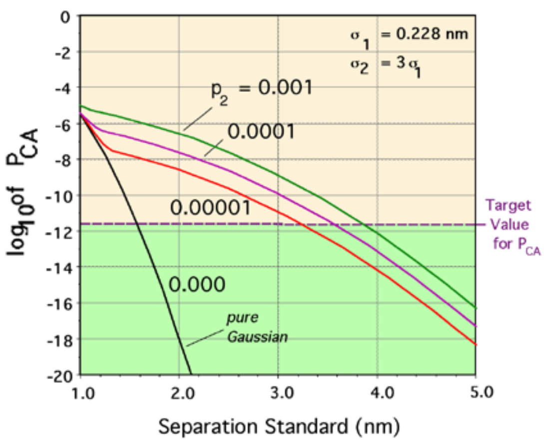

Figure 9 Effect of adding a second Gaussian term to the composite distribution with probability p_2 *and variance three times greater than* σ_1.

Importance of Form of the Distribution

The need to justify the form of the probability density function poses a formidable problem for the target-level-of-safety approach. Surveillance systems exhibit complex behavior that depends upon the signal propagation environment, the state of electronic subsystems, signal processing, and display system characteristics. For such systems, there is little theoretical justification for choosing any particular mix of classic distributions. The credibility of any choice must be based on data.

The vast majority of data collected during testing will lie in the part of the curve where it is impossible to distinguish the pure Gaussian pdf from a composite distribution with some exponential contamination. For the example in Figure 7, the point of divergence between the pure Gaussian and the mixed distribution is about 1.4 nm which is 6.1 standard deviations from the mean. Even at a low level of confidence, billions of data points would be required to distinguish between the distributions.

In addition to the sheer quantity of data required, one must consider the possibility of time variation in performance. Suppose, for example, that an exponentially distributed pdf component is present on only 1 out of 5,000 days (This could be due to some time-varying effect such as sunspot activity that creates adverse signal propagation conditions, unusual multipath conditions, or two critical subsystems, which interact adversely only when both are in a particular state.). Under this case, collecting an enormous amount of data over a period of a few months could fail to detect the underlying vulnerability and result in a false declaration that the distribution was purely Gaussian.

FURTHER PRACTICAL ISSUES

The previous sections have considered the probability modeling issues that arise in applying the CAP criterion as an accuracy requirement for surveillance systems. In this section we provide a broader look at the CAP test and its role in the overall safety assurance process.

Data Rejection Logic

The output of any sensor employed by a surveillance system is likely to use data rejection logic that rejects measurements that are highly inconsistent with aircraft performance capabilities or recent track history. Figure 10 illustrates a typical process in which a correlation box is placed around the expected position as extrapolated from track history. The size of the correlation box is initially chosen so that only a small fraction r of measurements fall outside the box. If no measurement is found within the correlation box, the track is coasted and the size of the box is increased for the next update.

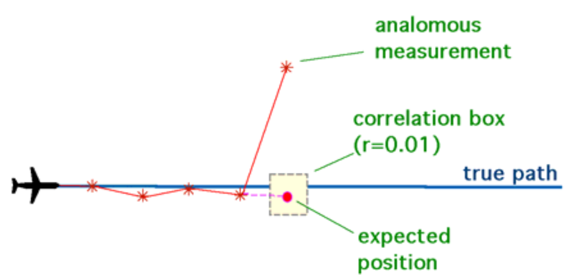

Figure 10 Data rejection suppresses the tails of the sensor error distribution

Such rejection of bad data tends to suppress the tails of the measurement error distribution, and this has a profound impact upon CAP tests performed in the tails. Large errors that occur rarely will result in coasts, not position measurement errors. Any probability estimation model that extrapolates far into the tails in order to apply a CAP criterion is basing its decision on measurements that the actual surveillance system will be designed to reject. A CAP test must take data rejection fully into account fully in order to properly assess RSP.

Component Failure Rates

In any surveillance system, hardware and software components will inevitably fail, and performance in this regard is part of the RSP specification. Complex electronic systems used in aviation today generally have MTBFs between 3000 hours and 35000 hours. Modern radar units are generally required to have MTBFs of around 20,000 hours.

A practical relationship exists between measurement error rates and equipment failure rates. Often a failing surveillance component will generate a short series of anomalous outputs before the failure is detected and the unit is shut down. Such errors are infrequent, but they can be major contributors to the tails of the measurement error distribution. Their characteristics may not be specified in unit performance specifications focused on nominal operations and may not be hinted at in data collected from units that are operating normally.

A reasonable balance should exist between the probability of large output error (CAP violation) for nominal operation and the component failure rate. A requirement for output error that is orders of magnitude more stringent than the component MBTF could greatly increase equipment cost without reducing actual risk to any significant degree.

For example, if the targeted probability of an unacceptable measurement error is 2.0E-12 per report and the unit generates 900 reports per hour, then one critical error will occur in approximately 5.5E08/hour. If the component MTBF of that unit is 20,000 hours, then the specifications allow 27,778 component failures for each unacceptable measurement permitted. The two specifications are difficult to reconcile because tens of thousands of complete surveillance failures are much more of a concern than a single isolated large measurement error under non-failure conditions. Practical specifications for permitted large error rates should not greatly exceed the inherent reliability of the system components unless there is some special engineering analysis that says that this makes sense.

Point Evaluation Versus Cumulative Probability Test

Because CAP test evaluates P_{CA} only within the collision risk zone, it largely ignores that part of the measurement error distribution that lies further than S_0+A_W from the mean. In other words, it does not penalize the system if aircraft can collide when they have separation greater than S_0. This ignored risk is clearly of interest because the ATM process often spaces aircraft at more than the separation standard. Furthermore, before aircraft reach S_0 they must pass through the regions of greater separation. This suggests that the cumulative probability of measurement errors beyond S_0 might be a more complete measure of risk than P_{CA}, which is based upon a narrow range of separation values.

Need for a Complete Suite of RSP Criteria

It should be clear that the CAP test alone does not provide a complete RSP characterization of the error characteristics of a surveillance system. For example, the CAP test does not reflect the ability of the system to estimate aircraft velocities. Normally, velocities are estimated from the time history of surveillance reports by tracking algorithms that smooth time-varying errors. Hence the time correlation of errors is an important feature of a surveillance system that must be addressed by additional performance criteria. The resulting suite of RSP criteria must be calibrated to ensure consistency among themselves. Careful modeling of each performance criterion is important to this calibration.

SUMMARY AND CONCLUSIONS

This paper has shown that RSP acceptance tests using probability models are sensitive to the assumptions used in formulating the models. Such assumptions must be validated with data to provide confidence in the RSP assessment.

• The Close Approach Probability (CAP) criterion can be applied to either the reference system approach or the Target Level of Safety (TLS) approach. Under the reference system approach, validation is simpler and less data is required to validate. In such cases, a simple test we call the *variance ratio test* can often be substituted for more involved statistical criteria. The variance ratio test says that if the error variance of a new system is less than the error variance of the reference system, then the new system can be accepted.

• Validation of a TLS through probability modeling requires an absolute determination of the pdf value in the tails of the error probability distribution. This can run into an intractable validation problem because the outcome is quite sensitive to the f*orm* of the probability density function employed. When we consider more general distributions, such as general exponential distributions or a composite pdf, the value of the pdf in the tails can be altered by orders of magnitude by different choices for the form of the pdf. The credibility of the CAP test thus hinges on validating the form of the pdf, e.g. proving that “fat tail” components are absent. The data quantity required for this validation may make the TLS approach impractical.

The CAP criterion, as commonly formulated, has a number of other issues that affect its practical application. It must be augmented to reflect other important RSP requirements, such as those involving velocity estimation. It must be modified to consider the data rejection capabilities of surveillance systems. And its risk target must be reasonable when compared to allowed equipment failure rates.

It is clear that all safety models used as part of an RSP formulation should clearly state their assumptions and indicate the extent to which such assumptions have been validated. Because the reference system approach if easier to validate, it should be used whenever it can be justified. Even here though, further work is needed to more carefully define the conditions under which the reference system approach may be applied. Efforts to extend the reference system approach to systems that have greater dissimilarities would be of great practical value.

The RSP criteria are intended to be independent of particular technologies. However, the error probability models and data collection methods used to verify RSP conformance must carefully consider the specific ways in which the technology can produce measurement errors. To ensure that this has been done properly, regulatory authorities employing an RSP approach may find it necessary to retain a sound understanding of the technologies underlying the systems in question.

REFERENCES

[1] "Concept of Operations for the Next Generation Air Transportation System", Joint Planning and Development Office, Draft Version 0.2, July 24, 2006

[2] "Manual on Airspace Planning Methodology for the Determination of Separation Minima", DO9689m, International Civil Aviation Organization, 1998 Amended 30 August 2000.

[3] "Minimum Aviation System Performance Standards for Aircraft Surveillance Applications (ASA)", RTCA DO-289, December 9, 2003.

[4] Nagaoka S., O. Amai, , 1991, "Estimation Accuracy of Close Approach Probability for Establishing a Radar Separation Minimum", Journal of the Institute of Navigation, Vol. 44, No, 1

[5] Thompson, Steve, J.W. Andrews, G.S. Harris, and K.A. Sinclair, 1 November 2006, "Required Surveillance Performance to Support 3-Mile and 5-Mile Separation in the National Airspace System", ATC-323, M.I.T. Lincoln Laboratory

[6] Jones, Stanley, R., March 2005, "ADS-B Surveillance Requirements to Support ATC Separation Standards", The MITRE Center for Advanced Aviation System Development, March 2005

[7] Jones, Stanley, R., 2003,"ADS-B Surveillance Quality Indicators: Their Relationship to System Operational Capability and Aircraft Separation Standards", ATC Quarterly, Vol. II pp 225-250

Author Biographies

John W. Andrews

John Andrews is a senior staff member in the Surveillance Systems Group at M.I.T. Lincoln Laboratory. He was one of the senior system analysts during the development of the Traffic Alert and Collision Avoidance System and has made contributions to tracking techniques, algorithm analysis, human subject flight testing, and the modeling of pilot visual acquisition performance. He has served as a consultant to the National Transportation Traffic Board in the investigation of mid-air collisions. Mr. Andrews holds a B.S. in Physics from the Georgia Institute of Technology and a Master's degree in aeronautical engineering from M.I.T.

Steven D. Thompson

Steven D. Thompson is a member of the technical staff of the Surveillance Systems Group at M.I.T. Lincoln Laboratory. He has worked in air traffic control for the past seventeen years and is currently focused on surveillance performance requirements. Prior to joining Lincoln Laboratory, he was at Bell Laboratories in Holmdel, New Jersey and Argonne National Laboratory in Chicago, Illinois. He received a B.S. degree in aerospace engineering and M.S. and Ph. D. degrees in nuclear engineering, all from Georgia Institute of Technology. He also received an M.B.A. form the University of Chicago. He is an active pilot with an airline transport pilot's certificate and holds a type rating in the Citation jet. He also holds a flight instructor's certificate with single and multi-engine airplane, instrument, and glider ratings.