A Unified Methodology for the Evaluation of Hazard Alerting Systems

by

JAMES K. KUCHAR

S.B., Massachusetts Institute of Technology, 1990
S.M., Massachusetts Institute of Technology, 1991

SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN
AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January, 1995

Signature of Author _____
Department of Aeronautics and Astronautics
January, 1995

Certified by _____
Associate Professor R. John Hansman, Jr.
Department of Aeronautics and Astronautics
Thesis Supervisor

Certified by _____
Professor Robert Simpson
Department of Aeronautics and Astronautics

Certified by _____
Dr. Milton Adams
Charles Stark Draper Laboratory, Inc.

Certified by _____
Dr. Gregory Zacharias
Charles River Analytics, Inc.

Accepted by _____
Professor Harold Y. Wachman
Chairman, Department Graduate Committee

A Unified Methodology for the Evaluation of Hazard Alerting Systems

by
James K. Kuchar

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy

Abstract

Hazard alerting systems monitor potential threats and issue warnings to human operators when undesirable incidents are projected to occur. Typically, alerting systems are developed through an *ad hoc*, evolutionary process, although many design issues are common across applications. Often these issues are not clear, and evaluating system performance can be difficult.

This thesis presents a unifying methodology to model and evaluate alerting systems. A generic description of measurement sources, alerting thresholds and displays, the human operator, and the situation dynamics is provided. The relationships between hazards, restrictions on maneuverability, unobservable states, and alerting thresholds are shown. The methodology accounts for modeling and measurement uncertainties and incorporates a probabilistic description of the operator's response to alerts. Because the methodology is developed in a generalized manner, it can be used in vehicle, transportation system, and process control applications.

A probabilistic analysis method is developed to provide quantitative measures of the probability of false alarm and missed detection. The alerting decision is recast as a signal detection problem, allowing the use of classical Signal Detection Theory methods to analyze the effects of threshold placement on system performance.

Two aeronautical examples are used to illustrate the flexibility of the methodology. First, the Ground Proximity Warning System (GPWS) is evaluated using a probabilistic model of the human response to an alert. Over flat terrain, the GPWS alerting threshold is shown to provide a level of safety that is relatively constant regardless of descent rate. Over a probabilistic terrain field, a decrease in the ability of GPWS to determine the severity of terrain hazards is demonstrated as the variability of terrain increases. The methodology is therefore able to show the constraints to system performance due to limited terrain information.

In a second example, the methodology is applied to the Traffic Alert and Collision Avoidance System (TCAS). Performance tradeoffs between two versions of TCAS are quantified in an example known to produce false alarms in actual practice. The updated TCAS thresholds are shown to effectively maximize system performance (low probabilities of false alarm and missed detection) given the quality of information available to the system.

Thesis Supervisor: Prof. R. John Hansman, Jr.
Associate Professor of Aeronautics and Astronautics

Acknowledgments

This work was supported by the Federal Aviation Administration under Grant #92-6-0001, and by the Boeing Commercial Airplane Company.

I must thank Professor R. John Hansman for his invaluable assistance in this work. He provided an important source of ideas, discussion, motivation, and criticism along the windy path that started in electronic charts, moved to terrain displays, and then on to the bigger picture of alerting systems in general.

I also wish to thank my Doctoral Committee: Professor Robert Simpson, Dr. Milt Adams, and Dr. Greg Zacharias, for their valuable input, friendship, and help through the whole doctoral process.

Much of the inspiration and motivation for this work has come from sources in the aerospace industry who are devoted to improving flight safety. I'd especially like to thank Don Bateman (Allied Signal) and George Boucek (Boeing) for their willingness to share information and ideas. I also am very appreciative of the efforts of Bob Giuda and United Airlines for their input and for jumpseat privileges that provided the means to monitor pilots in action on actual flights.

The other ASL-ites (John-Paul Clarke, Adam Dershowitz, Eric Johnson, Amy Pritchett, Sanjay Vakil, Tom Vaneck, Craig Wanke, and Lee Yang) deserve thanks also for their help in bouncing ideas and for participating in diversionous activities to keep grad school fun.

Finally, I thank my parents and my family for their wonderful support and love. I couldn't have done it without you. Thanks, Alex, for everything -- you've been incredibly understanding during the thesis roller coaster and were always there for me with a smile and remarkable patience.

Table of Contents

Definitions

Alert Space	The subset of state-space in which an alert is issued from the alerting system.
Alert stage	The state of an alerting system with respect to the information that is presented to the human operator.
Alerting threshold	The state value boundary between measurements that do not result in alerts and measurements that do result in alerts.
Avoidance trajectory	The probabilistic trajectory (A) that is followed given that an alert is issued at a certain moment.
Caution	An alert intended to make the operator aware of a hazard that may necessitate action at a later time.
Complexity constraint	A constraint on the future state trajectory due to limitations on the possible number of changes in control behavior.
Correct Detection	An event in which the alerting system correctly issues an alert.
Correct Rejection	An event in which the alerting system correctly does not issue an alert.
Current State Extrapolation	Extrapolating the state vector assuming that the highest state derivatives are constant.
Deferred Action	An event in which an alert is not issued, but an alert that is issued at a later time will still allow an incident to be avoided.
Dimension constraint	A constraint on the future state trajectory due to limitations on the number of states that can be affected during a maneuver.
Exclusion zone	The region in space where a hazard reference point must be located for an encounter to occur along a trajectory.
Exposure-dependent hazard	A hazard for which the probability of an incident is a function of the time that the state vector is in Hazard Space.
Exposure-independent hazard	A hazard for which the probability of an incident is independent of the time the state vector is in Hazard Space.
False Alarm	An event in which the alerting system incorrectly issues an alert.

False Alarm Rate	The rate at which false alarms occur, taking into consideration a set of potential hazard situations.
Hard hazard	A hazard for which an incident always occurs when an encounter occurs: $P(I \mid E) = 1$.
Hazard encounter	An event (E) in which the system state vector lies within Hazard Space.
Hazard Space	The subset of state-space in which an incident can occur.
Hazard Situation	A single set of initial states and future maneuvering behavior.
Incident	An event (I) that the alerting system is designed to prevent.
Intended Trajectory Extrapolation	Extrapolating the state vector based on an intended state trajectory.
Magnitude constraint	A constraint on the future state trajectory due to acceleration, rate, and time limitations.
Maneuver Limit Space	The subset of state-space in which an unavoidable entry into Hazard Space will occur in the future due to maneuverability limits or other constraints.
Measurement-induced error	An alerting decision error due to errors in the state measurement.
Missed Detection	An event in which the alerting system incorrectly does not issue an alert or issues an alert too late to prevent an incident.
Missed Detection Rate	The rate at which missed detections occur, taking into consideration a set of potential hazard situations.
Nominal trajectory	The probabilistic trajectory (N) that is followed given that no alert is issued.
Observable state-space	The subset of state-space that is observable by the alerting system.
Safety ratio	The ratio of the probability of an incident with the alerting system in operation to the probability of an incident without the alerting system.
Single-stage system	One in which the system transitions from a non-alert state into a single Warning state when the measured hazard level crosses the alerting threshold.
Soft hazard	A hazard for which an incident may not occur even though an encounter does occur: $P(I \mid E) < 1$.

State trajectory	The set of values that the state vector takes over a given time interval.
State vector	The complete set of parameters that define the dynamics of a hazard situation, $\mathbf{x}(t)$.
System benefit	The difference between the probability of an incident without the alerting system and the probability of an incident with the alerting system in operation.
Two-stage system	One in which the system transitions from a non-alert state to a Caution state to a Warning state as the measured hazard level crosses two alerting thresholds.
Warning	An alert that indicates that immediate action is needed to avoid an incident.

Nomenclature

(The chapter in which the variable is first used is in parentheses)

Vectors

$\mathbf{a}(t)$	Alert stage (2)
$\mathbf{e}(t)$	External factors affecting the operator (2)
$\mathbf{n}_a(t)$	Noise in alerting system measurements (2)
$\mathbf{n}_n(t)$	Noise in nominal information source measurements (2)
$\mathbf{u}(t)$	Control input vector (2)
$\mathbf{v}(t)$	Velocity (5)
$\mathbf{x}(t)$	Complete state vector of the system (2)
$\mathbf{y}_a(t)$	States observable to the alerting system (2)
$\mathbf{y}_n(t)$	States observable through nominal information sources (2)
$\xi(t)$	Uncertainties in system dynamics (2)
$\mathbf{x}$	State trajectory (2)
$\mathbf{x}_h$	Hazard reference point location (A)
$\mathbf{x}_m$	Measured location of hazard (A)
$\hat{\mathbf{x}}^-$	State estimate before measurement update (C)
$\hat{\mathbf{x}}^+$	State estimate after measurement update (C)

Matrices

$\mathbf{I}$	Identity matrix (C)
$\mathbf{K}$	TCAS alpha-beta filter gain (C)
$\mathbf{T}$	State transition matrix (B)

State-Space Functions

$\mathbf{D}$	Alerting display function (2)
$\mathbf{F}$	System dynamics function (2)
$\mathbf{G}_a$	Alerting system measurement observability function (2)
$\mathbf{G}_n$	Nominal source observability function (2)
$\mathbf{H}$	Operator response function (2)
$\mathbf{T}$	Alert threshold function (2)

Spaces and Sets

A	Avoidance Trajectory (3)
$\mathbf{E}$	Set of points at which exposure to hazard occurs (A)
HS	Hazard Space (2)
MLS	Maneuver Limit Space (2)
N	Nominal Trajectory (3)
$\mathbf{R}$	Space through which the aircraft travels (A)
$\mathbf{S}$	Set of all possible hazard extents σ (A), Set of hazard situations (3)
$\mathbf{T}$	Set of all possible trajectories τ (A)
$\mathbf{X}$	Complete state-space (2)
$\mathbf{Y}$	Observable state-space (2)
$\mathbf{Z}$	Exclusion Zone (A)

Events

CD	Correct Detection (3)
CDR	Correct Detection Rate (3)
CR	Correct Rejection (3)
E	Hazard encounter (2)
FA	False Alarm (3)
FAR	False Alarm Rate (3)
FN	False negative (3)
FP	False positive (3)
H_0	Hypothesis that intent is being followed (5)
H_1	Hypothesis that intent is not being followed (5)
I	Incident (2)
MD	Missed Detection (3)
MDR	Missed Detection Rate (3)
TN	True negative (3)
TP	True positive (3)
$\overline{A}$	Event that A does not occur (3)

Variables

g	Gravitational acceleration (2)
h	Altitude (2)
h_{min}	Minimum altitude attained during pull up maneuver (4)
Δh	Altitude loss during pull up maneuver (4)
N	Total number of trials (A)
n	Noise (3), Load factor (4)
p	Hypothesis test threshold (5), True probability (A)
$\hat{p}$	Estimate of probability (A)
p_{ji}	State transition probability (B)
r	Range (5), radius (A)
S	Total number of successes (A)
s	Observable measurement before noise corruption (3), standard deviation (A), path length (A)
t	Time (2)
v	Velocity (4)
x	True state value (2), horizontal position (4)
y	Observable measurement with noise (3), horizontal position (4)
α	Gamma distribution parameter (4), TCAS estimator parameter (C)
β	Autocorrelation parameter (B), TCAS estimator parameter (C)
$\dot{\gamma}$	Flight path angle rate (4)
ε	Estimation accuracy (A), Small time interval (2)
θ	Gamma distribution parameter (4)
Θ	Mean time (distance) to an incident in a soft hazard (A)
μ	Mean value (A)
ρ	Ownship extent definition (A)
σ	Hazard extent definition (A)

τ	Trajectory definition (A)
τ_R	Response time delay (4)
$\phi_{yy}(\tau)$	Autocorrelation function (B)
ψ	Bearing (5)

Probabilities

P(•)	Probability that • occurs
P(A | B)	Conditional probability that A occurs given that B has occurred
P(I | $\mathbf{x}(t)$)	Probability than an incident occurs when the state is at $\mathbf{x}(t)$ (2)
P_T(I | $\mathbf{x}(t)$)	Probability than an incident will occur along the trajectory T given the current state $\mathbf{x}(t)$. (3)

Functions

C_{FA}	Cost of false alarm (3)
C_{MD}	Cost of missed detection (3)
$E[\bullet]$	Expected value of •
$f_{\bullet}(\bullet)$	Probability density function of •
$(\dot{x})$	Time derivative of x
F	Terrain collision function (4)
$f_{xyz}(\mathbf{x}\text{-}\mathbf{x}_m)$	Probability density function for hazard location (A)
J	Alerting cost function (3)

Acronyms

AGL	Above Ground Level
CFIT	Controlled Flight Into Terrain
GPWS	Ground Proximity Warning System
MSL	Mean Sea Level
PDF	Probability Density Function
RA	Resolution Advisory
ROC	Receiver Operating Characteristic
SOC	System Operating Characteristic
SR	Safety Ratio
TA	Traffic Advisory
TCAS	Traffic Alert and Collision Avoidance System
VMD	Vertical Miss Distance

TCAS Parameters

ALIM	Minimum altitude separation limit (C)
DMOD	Range threshold (5)
τ	Time threshold (5)
H_1	Diverging range threshold (C)
ZTHR	Altitude separation buffer (C)

1. Introduction

A hazard alerting system is one of several safety components typically found in complex human-operated systems such as vehicles, traffic control systems, and process control applications. In these applications, the operator performs a task with the system such as traveling between two locations or maintaining a chemical reaction. This task is managed using a set of operating procedures and informational displays that provide feedback on the state of the system and on the environment in which it operates (Sheridan 1992, Wickens 1992).

Occasionally, hazards to the safe operation of the system may arise. These hazards can be internal to the controlled system (e.g., an engine fire) or may originate from the external environment (e.g., a collision threat). Many of these hazards can be observed and avoided using nominal operating procedures and informational displays. In certain situations, however, an alerting system is needed to actively warn the operator through aural or visual alerts that a hazard exists. Some alerting systems also provide additional resolution information that aids the operator in avoiding the hazard.

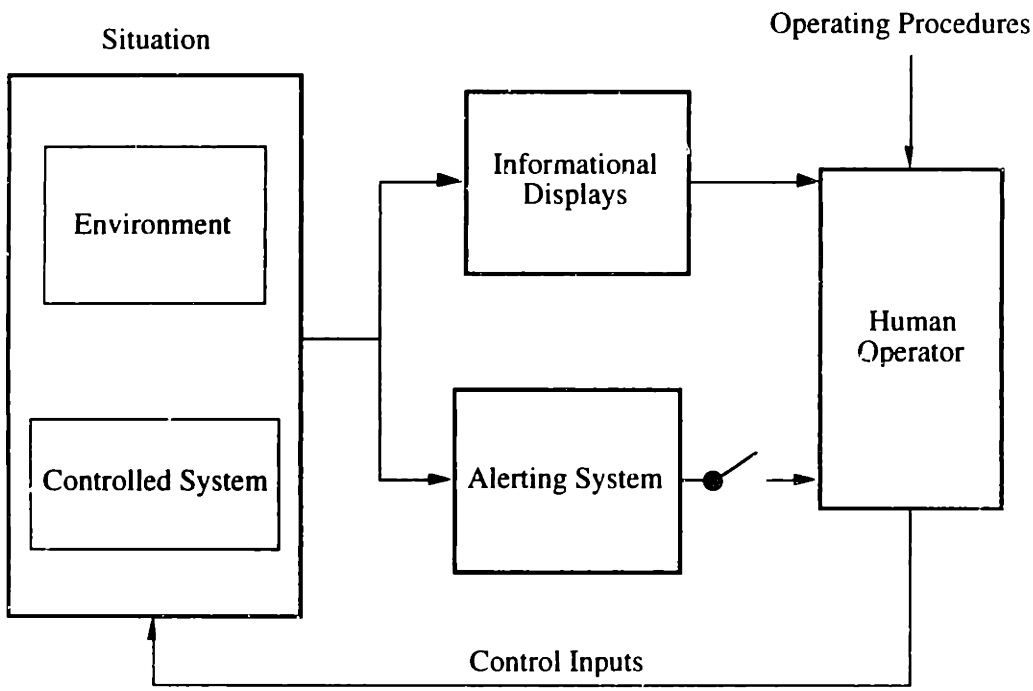


Figure 1.1: System Components

A diagram of the alerting system in relation to the other system components is shown in Figure 1.1. As discussed above, the informational displays and, when

necessary, the alerting system provide the operator with feedback on the controlled system and the environment. The operator uses this information to make control inputs to the system to achieve the desired goal.

Three categories of applications that typically include alerting systems are outlined in Figure 1.2. Alerting systems in process control applications are generally designed to alert the operator that certain parameters of the process are reaching dangerous values. For example, a nuclear power alerting system may produce alerts when coolant levels become low. Other alerts may be generated when critical components such as valves or pumps fail.

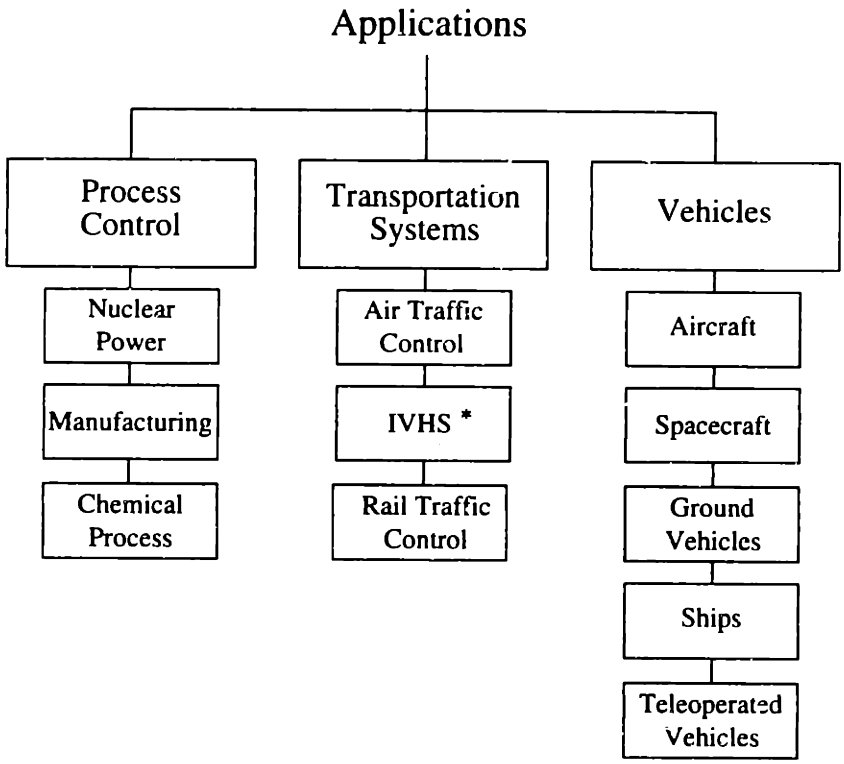


Figure 1.2: Typical Alerting System Applications

Applications in transportation systems typically involve a controller that oversees a number of independent vehicles. Examples include air traffic control, an Intelligent Vehicle Highway System, and rail traffic control. Hazards in these systems generally involve collision threats. Alerts issued to the controller typically result in commands to vehicles to change their trajectories.

Alerting systems in single vehicles are generally designed to warn the vehicle operator that a hazard exists. In response to an alert, the operator typically changes the trajectory or configuration of the vehicle. Aircraft, spacecraft, ground vehicles, and ships use alerting systems to enhance the safety of their operation.

1.1 Alerting System Performance

As the alerting system operates, it essentially makes a discrete decision to either remain silent or to issue an alert. Typically, this decision is based on whether certain states exceed critical values that form an alerting threshold. Due to errors in measurements or limitations in design, faulty alerting decisions can occasionally occur. In particular, a system may fail to alert when necessary (termed a missed detection), or may issue an alert when one is not needed (termed a false alarm).

Missed detections can result in an accident if the operator does not become aware of the hazard through the other informational displays. False alarms have been shown to reduce the operator's confidence in the alerting system (DeCelles 1991, Mellone 1993). Alerting system performance is therefore typically measured in terms of probabilities of missed detection or false alarm. While both types of error are undesirable, they generally cannot be eliminated simultaneously. Rather, some tradeoff between false alarms and missed detections is typically made.

A number of design issues affect alerting system performance. Although many of these issues are common across applications, current design methods typically follow an *ad hoc*, evolutionary process. As alerting systems become more capable and more complex, however, the issues on which to focus design efforts become less evident. A unifying methodology that can be applied over a range of applications is therefore needed to highlight the major design issues and to provide a means for evaluating system performance.

1.2 Overview of the Thesis

This work presents a unified methodology for modeling and evaluating alerting systems. Because of its generalized development, the methodology can be applied to an assortment of hazard types and system platforms such as those shown in Figure 1.2. However, to provide a cohesive foundation on which to discuss alerting systems, the thesis concentrates on applications in civil aviation. These applications involve an aircraft that is controlled by a human operator in the presence of external hazards such as terrain, other

aircraft, or severe weather. Because a wide variety of challenging aeronautical applications exist, this focus also enables the use of several examples from current or proposed alerting systems.

Chapter 2 develops a methodology for modeling alerting systems. The chapter discusses the major components and issues that affect system design and performance from a state-space approach.

In Chapter 3, the issues associated with the selection of an alerting threshold are discussed. Because alerting system performance depends on the effects of uncertainties in measurements and other parameters, a probabilistic analysis of hazard encounter situations is taken. This approach addresses factors such as uncertainties in sensor measurements, the human response to an alert, and knowledge of the character of the hazard (e.g., its size, shape, and severity). The alerting decision is recast as a signal detection problem, enabling the use of classical Signal Detection Theory metrics such as the probabilities of missed detection and false alarm.

The methodology is applied in two aeronautical examples in Chapters 4 and 5. These chapters demonstrate the flexibility of the methodology as it is applied to alerting systems designed for different types of hazards. The methodology is used to evaluate these alerting systems and to quantify the effect of different system designs on performance.

Chapter 6 provides a summary and outlines the major contributions of the thesis. Appendix A derives the probabilistic analysis methodology that is used to calculate the probability of false alarm and missed detection. Appendix B provides a brief overview of Signal Detection Theory. Appendices C and D describe the alerting system applications from Chapters 4 and 5 in more detail.

A number of new terms and concepts are presented in this work. Important terms are identified the first time they are used in italics in the text and are also provided in a list of definitions at the beginning of the thesis.

2. Generalized Model of Alerting Systems

2.1 State-Space Representation of Alerting Systems

The methodology presented in this thesis is based on a state-space representation of the operational system. This approach allows for consideration of all of the issues affecting alerting system design yet also provides generality and allows the methodology to be used in many applications.

The state-space representation is based on multivariable control system theory such as that described in Kirk (1970). The variables $x_1(t)$, $x_2(t)$, ..., $x_n(t)$ are defined as the values of the states of the system at time t. These states represent the complete set of parameters that define the dynamics of a hazard situation. In a collision alerting system application, for example, these states include the relative positions and velocities of the own vehicle and an obstacle. The *state vector*, $\mathbf{x}(t)$, is then defined as:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) & \dots & x_n(t) \end{bmatrix}^T \tag{2.1}$$

In general, $\mathbf{x}(t)$ has n elements and lies within the n-dimensional state-space $\mathbf{X}$.

2.1.1 The State Trajectory

During the operation of the system, the states typically move in state-space. These changes in the state vector occur in accordance with the system dynamics, the environment, and inputs from the human controller. The set of values of $\mathbf{x}(t)$ over a given time interval is the *state trajectory*, denoted $\mathbf{x}$. State trajectories can be viewed in state-space using a State-Space Diagram as shown in Figure 2.1. Although the Figure shows a continuous state trajectory, $\mathbf{x}$ can be discontinuous when states change between discrete values.

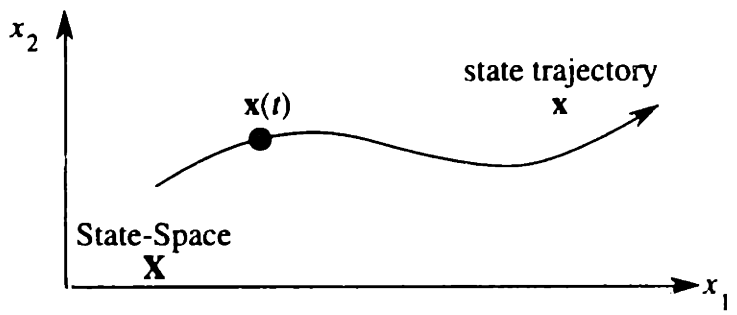


Figure 2.1: Example State-Space Diagram
$\mathbf{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) \end{bmatrix}^T$

Time can be included as an additional dimension in State-Space Diagrams to show the time dependence of **x** explicitly. As shown in Figure 2.2, such a diagram indicates not only what state values occur along the trajectory, but also the time at which those state values occur.

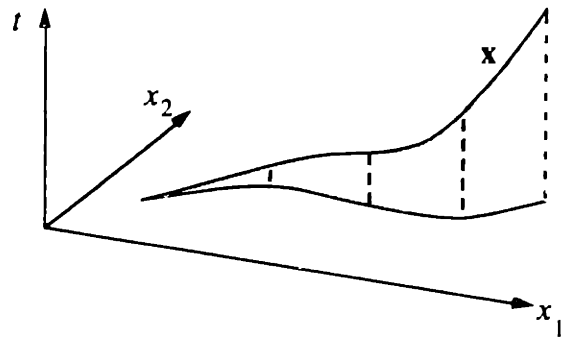


Figure 2.2: Example State-Space Diagram with Time as an Additional Dimension

State trajectories can be constructed both for past state values and for projected, future state values. Because alerting systems are designed to protect against future hazards, only future state trajectories are considered here.

2.1.2 Hazards in State-Space

When **x** enters certain regions of the state-space **X**, an undesirable event termed an *incident* (denoted I) may occur. An incident is defined according to the specific problem under study, and can be defined as the catastrophic loss of the system (such as the collision of a vehicle with terrain), or as a less severe but undesirable event (such as the violation of prohibited airspace). Thus, Hazard Space can be tailored to enable the examination of the potential benefits (e.g., avoiding an accident) and the potential negative effects (e.g., violating an altitude restriction) of using an alerting system.

Those regions in the state-space **X** in which an incident can occur are termed *Hazard Space* (denoted **HS**). In aeronautical applications, for example, Hazard Space consists of terrain, other aircraft, and hazardous weather. Note that some types of Hazard Space may exist such that an incident does not occur with probability one. If an aircraft were to fly through a region of severe weather, for example, it is possible that an incident would not occur. Other types of Hazard Space (e.g., terrain) are such that an incident occurs with probability one.

The relationship between **x** and Hazard Space can be viewed in a *State-Space Hazard Diagram* (Figure 2.3) that combines depictions of the state trajectory and Hazard Space. If Hazard Space changes in size, shape, or location with time, time can be included as an additional dimension as discussed in Section 2.1.1.

In the example in Figure 2.3, the current state, $\mathbf{x}(t)$, is located as shown and the future state trajectory is projected to enter Hazard Space. However, whether an alert is truly warranted at time t in this situation depends on a number of issues that are covered in the remainder of this chapter.

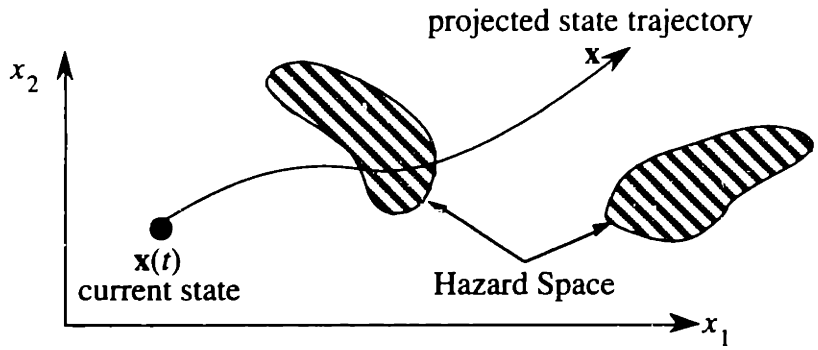


Figure 2.3: Example State-Space Hazard Diagram

At a given location, $\mathbf{x}(t)$, in state-space, the probability that an incident occurs is denoted $P(I \mid \mathbf{x}(t))$. By definition, then, $P(I \mid \mathbf{x}(t))$ is zero outside Hazard Space, and is greater than zero inside Hazard Space:

$$\mathbf{x}(t) \in \mathbf{HS} \quad \Leftrightarrow \quad P(I \mid \mathbf{x}(t)) > 0 \tag{2.2}$$

To distinguish between the event of entering Hazard Space and the event of having an incident, the term *hazard encounter* (denoted E) is used to indicate a situation in which the state vector is in Hazard Space. Recall that it is possible to have a hazard encounter (e.g., to enter a region of severe weather) but not have an incident. Also, by definition an incident cannot occur without a hazard encounter.

$$\mathbf{x}(t) \in \mathbf{HS} \quad \Leftrightarrow \quad E \tag{2.3}$$

2.2 Model of Alerting System Components

Having developed the basic state-space concepts in Section 2.1, it is necessary to determine how $\mathbf{x}(t)$ evolves over time to create a future state trajectory **x**. Continuing the state-space approach, the alerting system is modeled in a control system framework as

shown in Figure 2.4. In this representation, there are four main elements in the control loop: Measurement Sources, Alerting System, Human Operator, and Situation Dynamics.

The Measurement Sources block represents the components that provide estimates of the relevant system states. The Alerting System determines if the situation is such that an alert is warranted. When appropriate, alert information is provided to the Human Operator through alerting displays. The Human Operator uses the information from the alerting system and other sources to make necessary control inputs to the Situation Dynamics. Control inputs and uncertainties in the Situation Dynamics affect the evolving state vector, $\mathbf{x}(t)$, which is fed back to the Measurement Sources as the loop repeats.

Taking a more detailed look at Figure 2.4, recall that the state vector $\mathbf{x}(t)$ contains the complete set of information describing the aircraft and any external threats in its vicinity. The state vector can include internal parameters such as the vehicle's spatial location, velocity, or acceleration, and external parameters describing a hazard's location, size, or severity. As shown in Figure 2.4, these states can be observed by the flight crew through the Nominal Information Sources and the Alerting System Displays.

The Nominal Information Sources block in the Figure represents the information sources that are used during normal operation of the system. Cockpit instruments, air traffic control communications, views through the windscreen, and aeronautical charts are examples of nominal information sources. When a hazard exists, the flight crew may or may not become aware of the hazard through the Nominal Information Sources. The vector $\mathbf{y}_n(t)$ is defined as the subset of $\mathbf{x}(t)$ that is accessible to the flight crew through these nominal information sources. Note that $\mathbf{y}_n(t)$ does not necessarily represent all states in $\mathbf{x}(t)$; some states may be unobservable to the rest of the control loop. Observable states in $\mathbf{y}_n(t)$ may be direct measurements of states in $\mathbf{x}(t)$ (such as reading airspeed or altitude from a cockpit instrument) or may be indirect approximations (such as visually estimating an intruding aircraft's altitude and heading). Thus, each element of $\mathbf{y}_n(t)$ is an estimate (with some associated uncertainty) of a state in $\mathbf{x}(t)$. Following a similar methodology to that taken in multivariable control system design and estimation (e.g., D'Azzo & Houpis (1988) or Gelb (1989)), uncertainties in the estimates are modeled through a noise input vector, $\mathbf{n}_n(t)$. A generalized observability function, $\mathbf{G}_n$, is also defined that maps states in $\mathbf{x}(t)$ to their estimates in $\mathbf{y}_n(t)$ (if such estimates exist).

$$\mathbf{y}_n(t) = \mathbf{G}_n (\mathbf{x}(t), \mathbf{n}_n(t)) \tag{2.4}$$

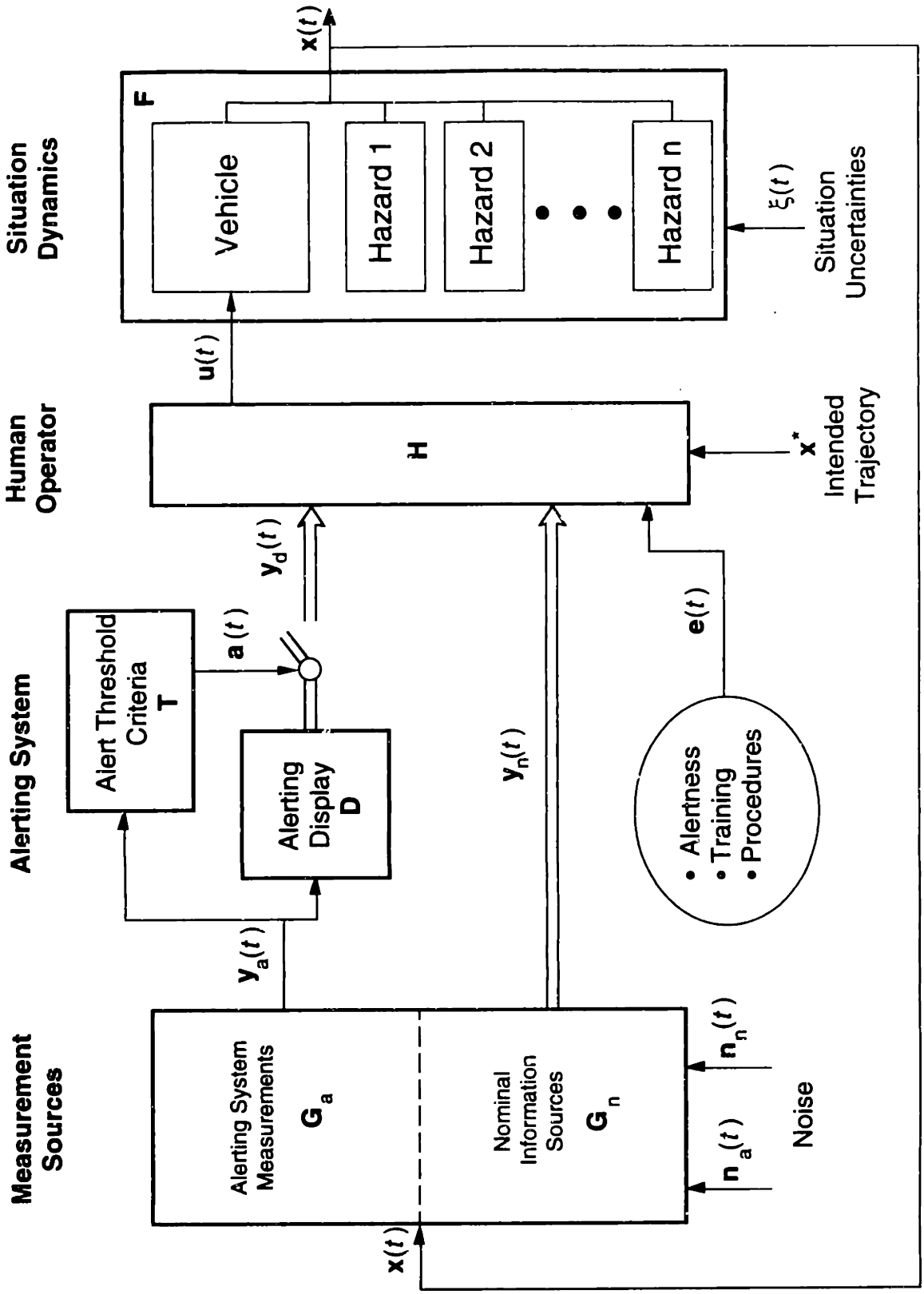


Figure 2.4: Generalized Hazard Situation Block Diagram

The other path by which hazard information is transmitted to the flight crew is through the alerting system. Here, state information is typically measured using sensors, data transmitted from the ground or other aircraft, or an onboard database. The vector $\mathbf{y}_a(t)$ represents estimates of the states of $\mathbf{x}(t)$ that are observable to the alerting system, much like $\mathbf{y}_n(t)$ represented estimated states directly observable by the flight crew. There may or may not be an overlap in the states represented by $\mathbf{y}_n(t)$ and $\mathbf{y}_a(t)$. Also, there may be states of $\mathbf{x}(t)$ that are not represented in either $\mathbf{y}_n(t)$ or $\mathbf{y}_a(t)$. The vector $\mathbf{y}_a(t)$ can also be corrupted by noise, $\mathbf{n}_a(t)$, and will be related to the system state vector through a generalized observability function $\mathbf{G}_a$:

$$\mathbf{y}_a(t) = \mathbf{G}_a (\mathbf{x}(t), \mathbf{n}_a(t)) \tag{2.5}$$

The alerting system compares $\mathbf{y}_a(t)$ against one or more alerting thresholds and categorizes the situation as non-hazardous or as one of possibly several hazard levels, termed *alert stages*. Some alerting systems use one non-hazard stage and one hazard stage. Other systems use several hazard stages to provide additional information to the pilot regarding the urgency with which action is needed. Alert stages are covered in more detail in Section 3.1.

The variable $\mathbf{a}(t)$ specifies the particular alert stage being used. For example, $\mathbf{a}(t) = 0$ might signify that no alert is issued while $\mathbf{a}(t) = 1$ might signify that an alert is issued. As shown in Figure 2.4, a generalized threshold function, $\mathbf{T}$, gives the mapping from the aircraft and hazard state estimates, $\mathbf{y}_a(t)$, to the value of the alert stage variable, $\mathbf{a}(t)$:

$$\mathbf{a}(t) = \mathbf{T} (\mathbf{y}_a(t)) \tag{2.6}$$

Depending on the current alert stage, an alerting display[1] may transmit information to the pilot. The vector $\mathbf{y}_d(t)$ represents the information provided to the flight crew through the alerting system. In general, $\mathbf{y}_d(t)$ contains signals designed to attract the crew's attention and any state information that is needed to resolve the situation. The generalized display function, $\mathbf{D}$, describes the mapping from the state estimates available to the alerting system ($\mathbf{y}_a(t)$) to the states provided to the pilot ($\mathbf{y}_d(t)$) based on the alert stage:

$$\mathbf{y}_d(t) = \mathbf{D} (\mathbf{y}_a(t), \mathbf{a}(t)) \tag{2.7}$$

---

[1] "Display" here indicates any interface that provides information to the operator. Example displays are aural messages and electronic display screens.

As Figure 2.4 shows, the human operator's response, $\mathbf{u}(t)$, is a function of several inputs. The vector $\mathbf{u}(t)$ includes direct control inputs or flight plan changes, and indirect controls such as queries to Air Traffic Control (ATC) for a new heading or altitude. State information from $\mathbf{y}_n(t)$ and $\mathbf{y}_d(t)$ and the desired trajectory, $\mathbf{x}^*$, provide the primary sources for the operator's response. As the system deviates from $\mathbf{x}^*$, the operator typically applies control inputs to return the system to the desired trajectory. Other factors (denoted by $\mathbf{e}(t)$) such as training, alertness, and previous experience modify the operator's actions. The function $\mathbf{H}$ maps the observable states, desired trajectory, and other factors to the control inputs:

$$\mathbf{u}(t) = \mathbf{H}\ (\ \mathbf{y}_n(t),\ \mathbf{y}_d(t),\ \mathbf{x}^*,\ \mathbf{e}(t)\ ) \tag{2.8}$$

In the Situation Dynamics block of Figure 2.4, uncertainties in modeling or system dynamics are represented by $\xi(t)$, accounting for differences between the assumed dynamic model and the actual situation dynamics. Following conventional control system methods, the state derivatives are determined from a generalized function, $\mathbf{F}$, of the current state, pilot inputs, and system uncertainties.

$$\dot{\mathbf{x}}(t) = \mathbf{F}\ (\ \mathbf{x}(t),\ \mathbf{u}(t),\ \xi(t)\ ) \tag{2.9}$$

It can be seen that $\mathbf{x}(t)$ is a function of the different component functions, $\mathbf{F}$, $\mathbf{H}$, $\mathbf{D}$, $\mathbf{T}$, $\mathbf{G}_n$, and $\mathbf{G}_a$, the desired trajectory $\mathbf{x}^*$, the uncertainties $\xi(t)$, $\mathbf{n}_a(t)$, $\mathbf{n}_n(t)$, and the factors contained in $\mathbf{e}(t)$. Thus, the state trajectory, $\mathbf{x}$, also depends on each component in the model shown in Figure 2.4.

2.3 Extrapolation of the State Trajectory

Having defined the primary system components that affect the state vector, it is now possible to discuss the methods by which the future state trajectory can be determined. If the future state trajectory is known, the alerting system will be better able to determine if an alert is necessary. Also, as is discussed in Sections 2.3.2 and 2.3.3, limitations on the system dynamics can constrain the future state trajectory. These constraints are important in determining if adequate space remains in which to perform an avoidance maneuver.

2.3.1 Extrapolation Methods

To illustrate two methods that can be used to extrapolate the state trajectory, consider Figure 2.5. This example shows a situation in which an aircraft with velocity V is

at an altitude of h and is descending toward flat terrain at a descent rate of $-\dot{h}$. A simple, two-state system of altitude and descent rate over flat terrain is used: $\mathbf{x}(t) = [h(t) \quad \dot{h}(t)]^T$.

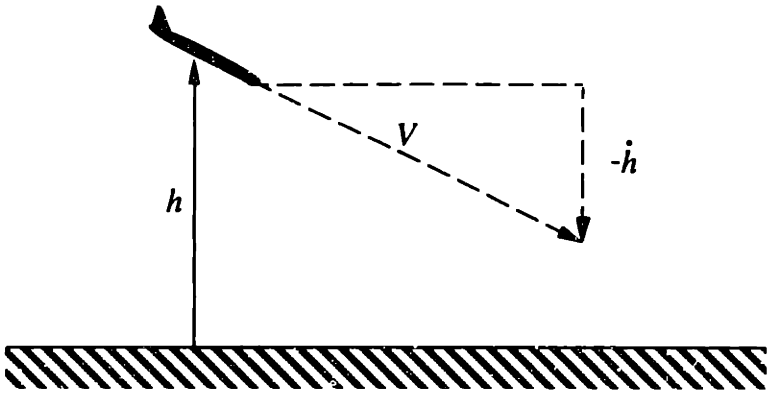


Figure 2.5: **State-Space Example:**
Aircraft descending into flat terrain.
Altitude h, descent rate $-\dot{h}$

Given the current state vector, $\mathbf{x}(t)$, there are two methods by which the future state trajectory can be predicted. The simplest method, termed *Current State Extrapolation* assumes that the highest derivative of each state is constant. In this example, Current State Extrapolation is based on the assumption that $\dot{h}$ is constant. In a State-Space Hazard Diagram such as that shown in Figure 2.6, the extrapolated state follows a straight-line trajectory as shown. While simple, this method cannot anticipate future changes in control inputs or in the environment. As a result, the uncertainties in the future state values typically grow with time.

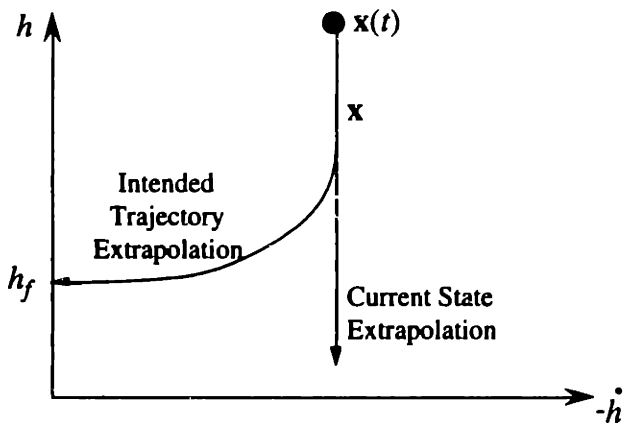


Figure 2.6: **State Extrapolation Methods**
Two-State Example

A more complex method of state extrapolation is possible if additional information is available to the alerting system. If an intended state trajectory is known, it can be used to

improve the trajectory extrapolation, termed *Intended Trajectory Extrapolation*. Returning to Figure 2.6, Intended Trajectory Extrapolation can reflect the *a priori* knowledge that the aircraft will level off at a final altitude h_f above the terrain. Intended Trajectory Extrapolation can be based on a single, target state or on a series of intended state values along the trajectory.

2.3.2 Maneuver Limit Space

Due to maneuvering constraints on the system (e.g., acceleration limits), some trajectories may not be physically realizable. As a result, although the state, $\mathbf{x}(t)$, may be outside Hazard Space, an entrance into Hazard Space may be unavoidable because of maneuvering constraints. Therefore, additional hazardous regions in state-space also exist, termed here as *Maneuver Limit Space*, denoted **MLS**. Maneuver Limit Space therefore defines those regions in state-space where all realizable trajectories starting at $\mathbf{x}(t)$ enter Hazard Space at some point in the future.

$$\mathbf{x}(t) \in \mathbf{MLS} \quad \Leftrightarrow \quad \begin{array}{l}\text{there exists an } \varepsilon \text{ for each possible future} \\ \text{trajectory such that } \mathbf{x}(t+\varepsilon) \in \mathbf{HS}\end{array} \tag{2.10}$$

The outer boundary of Maneuver Limit Space represents the threshold at which a maximum-performance escape maneuver must begin in order to avoid Hazard Space. As discussed later in Section 2.4, the alerting system is intended to alert the pilot of the hazard sufficiently early to allow for a maneuver to begin before entering Maneuver Limit Space.

Recalling the example of the aircraft descending into flat terrain in Figure 2.5, a State-Space Hazard Diagram is shown in Figure 2.7. Hazard Space is defined as the space below the terrain; entry into this space results in an incident. In the example, a maximum acceleration limit is assumed for a pull-up maneuver. Maneuver Limit Space is therefore defined by a parabolic trajectory in state-space as shown. Because the aircraft is descending at a constant rate, the state trajectory follows a straight line as shown in Figure 2.7. If the aircraft does not begin to maneuver before Maneuver Limit Space is reached, an incident cannot be prevented.

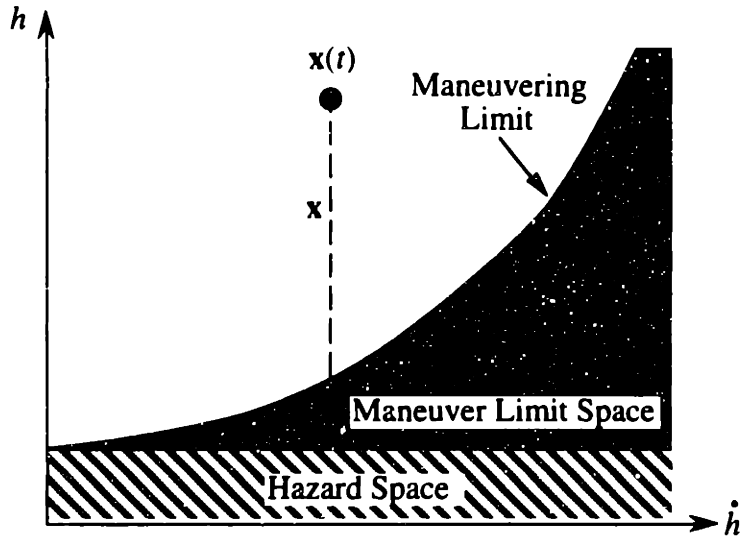


Figure 2.7: State-Space Hazard Diagram for Descent Into Flat Terrain Example

2.3.3 Maneuvering Constraints

Maneuver Limit Space in the previous example is defined based on the maximum-performance maneuver that can be expected of the system. Certain constraints on this maneuver must be considered, such as response latency or the type of maneuver that is performed by the operator. Often these constraints are probabilistic. It is also of interest to consider other maneuvers that may be constrained in some manner. For example, the state trajectory that is followed when an alert is issued is of great interest in Chapter 3.

Constraints on the state trajectory involving acceleration or rate limits like those used to define Maneuver Limit Space are termed *magnitude constraints*. Magnitude constraints may also occur due to the operator's response latency and operating procedures. For example, it may be unrealistic to assume that the operator will respond to an alert with an extreme but possible maneuver. Instead, some limits on response latency and maneuver aggressiveness are typically assumed.

Additional constraints may also exist that are related to the type of maneuver that is performed rather than the magnitude of that maneuver. *Dimension constraints* restrict the number of states that are controlled during the avoidance maneuver. A climb maneuver is an example of a one-state maneuver, and a climbing turn is a two-state maneuver. The number of dimensions used in a maneuver can be linked to the type of information that is presented to the operator. For example, Kuchar & Hansman (1993) showed that the type

of maneuver used by a pilot in response to an alert can be influenced by the type of alerting display that is used.

In Figure 2.8, an example dimension constraint is shown. If only climbing avoidance maneuvers are possible, an avoidance maneuver must begin before reaching location **A** or an incident will result. Thus, the region between **A** and Hazard Space is Maneuver Limit Space. If, however, lateral maneuvers are also possible, the avoidance maneuver can be delayed until reaching location **B** because the added dimension of lateral maneuverability allows the vehicle to avoid the hazard more efficiently. Thus, maneuvering constraints govern the amount of space that can be safely used by the vehicle.

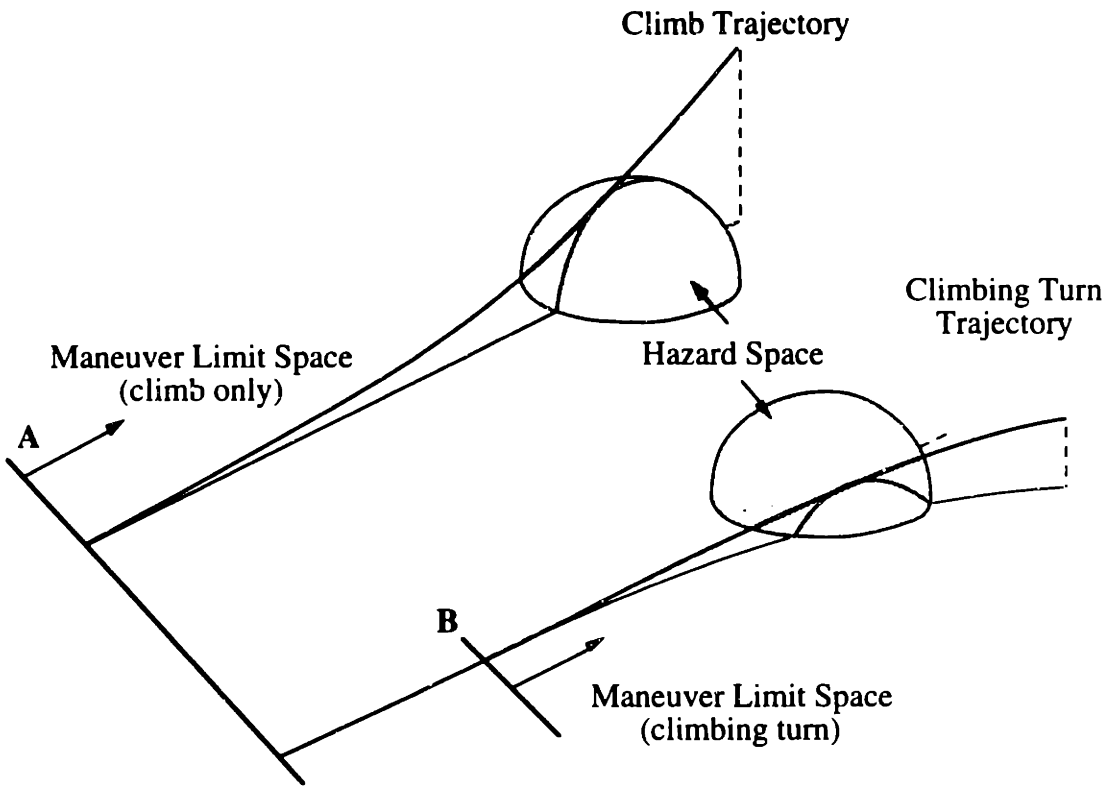


Figure 2.8: Effect of Maneuver Dimensionality on Maneuver Limit Space
Climb vs. Climbing turn

Complexity constraints generally limit the number of changes in control behavior during the avoidance maneuver. An example of a simple maneuver is a right turn, while a complex maneuver might involve a right turn for several seconds followed by a left turn. Complexity constraints are typically present due to operating procedures or limited information available to the operator.

Maneuvering constraints, summarized in Table 2.1, must be considered when extrapolating state trajectories. For example, the definition of Maneuver Limit Space in a certain situation may be dependent upon whether climbs or climbing turns can be performed. As is discussed in Section 2.4, maneuvering constraints also ultimately affect the location of the alerting thresholds.

Table 2.1
Example Maneuvering Constraints

	Example Maneuver Types	
Constraint	Low Performance	High Performance
Magnitude	Low-acceleration	High-acceleration
Dimension	Climb only	Climbing turn
Complexity	Turn right	Turn right, then turn left

Note that a State-Space Hazard Diagram can also be constructed for non-maneuvering actions such as ejecting from the aircraft or issuing an emergency radio call. In these non-maneuvering actions, the pilot does not explicitly change the future state trajectory. However, Maneuver Limit Space can still be defined by factors such as the amount of time the pilot needs to perform the required actions.

2.4 Alert Space

With Hazard Space and Maneuver Limit Space defined from the preceding Sections, it is now necessary to consider the purpose of the alerting system. The utility of an alerting system is that when an alert is issued, it generates a discrete change in the operator's behavior intended to avoid an incident. Specifically, the operator changes the state trajectory from the nominal trajectory that would have been followed had no alert been issued to an avoidance trajectory intended to prevent an incident.

Because it is unrealistic to expect the vehicle to maneuver at its limits when avoiding hazards, an alert should be issued before $\mathbf{x}(t)$ reaches Maneuver Limit Space. The region in space in which alerts are generated is termed Alert Space. Typically, Alert Space allows for a certain response latency and an avoidance maneuver well within performance limitations. The outer boundary of Alert Space is defined by the alerting threshold function $\mathbf{T}$. By definition, no alerts are generated when $\mathbf{x}(t)$ is outside Alert Space.

Returning to the descent into terrain example from Figure 2.5, consider the State-Space Hazard Diagram in Figure 2.9. In Figure 2.9, Alert Space is shown for an example avoidance maneuver based on some response latency and some pull-up acceleration less than that used to define Maneuver Limit Space. Consider an aircraft located in state-space at point **A** with a constant descent rate, $-\dot{h}_0$. The aircraft's state therefore travels down the vertical dashed line. When the state reaches point **B** (crossing the alerting threshold defined by Alert Space), an alert is issued. After some time delay, the pilot begins to respond to the alert at point **C**. Using a constant-acceleration pull-up maneuver, the state then follows the parabolic trajectory from point **C** to point **D**. If the pilot reacts at least as quickly and aggressively as required by the designed Alert Space, the hazard will be avoided.

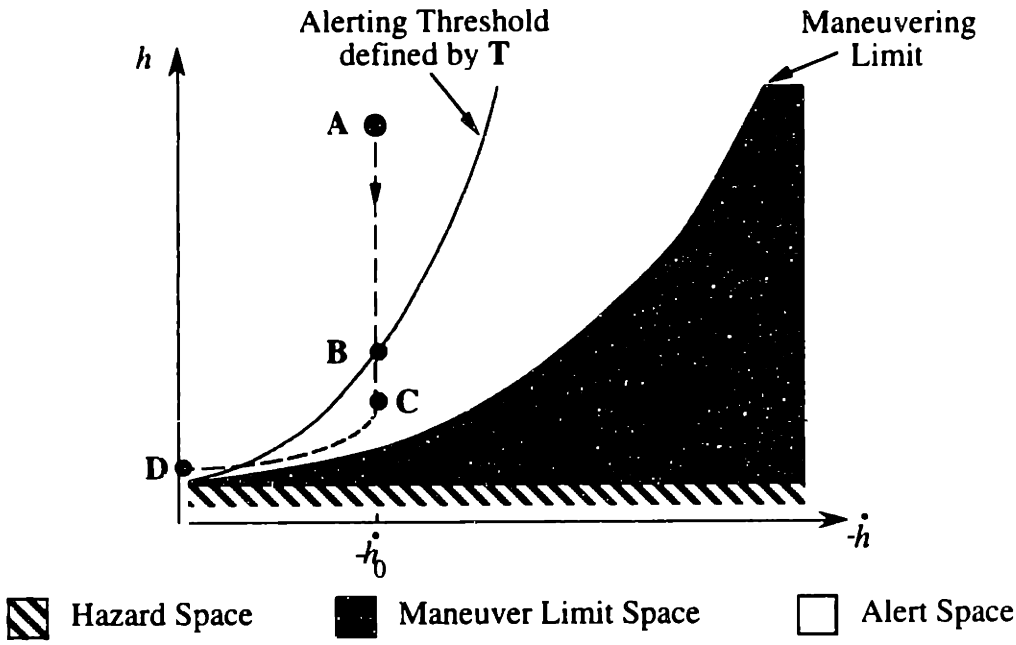


Figure 2.9: State-Space Hazard Diagram for Descent into Flat Terrain Example

2.4.1 Alerting Decision Outcomes: False Alarms and Missed Detections

As the alerting system operates, one of two mutually exclusive events occurs: either an alert is issued or it is not. The true situation is also one of two mutually exclusive events: either an alert is needed (i.e., an incident will occur without an alert) or an alert is not needed. Because the projected occurrence of an incident is probabilistic, however, the alerting system's decision may not reflect the true need for an alert. Specifically, two types of error can occur.

A *false alarm* occurs when the alerting system issues an alert when an alert is not needed to avoid an incident. A *missed detection* occurs when an alert is needed to avoid an incident but an alert is not issued or is too late to prevent an incident. As mentioned in the Introduction, false alarms and missed detections are undesirable and can have a major effect on the acceptability of an alerting system. However, these errors cannot be simultaneously eliminated. The remainder of this chapter discusses the factors that contribute to false alarms and missed detections. Chapter 3 discusses the methods that can be used to determine the probabilities of false alarm and missed detection and how to examine the tradeoffs between them.

2.5 Factors Contributing to False Alarms and Missed Detections

Section 2.4 shows that the alerting decision depends on whether the current state lies within Alert Space. Recall, however, that the alerting system operates using the observable state vector $\mathbf{y}_a(t)$ rather than the full state vector $\mathbf{x}(t)$. Therefore, an analysis of the alerting decision must consider the *observable* state-space, $\mathbf{Y}$, spanned by the vector $\mathbf{y}_a(t)$, rather than the complete state-space, $\mathbf{X}$.

Typical errors in alerting system decisions arise from a combination of factors that can be grouped into three categories discussed below: measurement uncertainties, errors in the state trajectory extrapolation, and unobservable states.

2.5.1 Measurement Uncertainties

A *Measurement-induced* error occurs when measurement errors result in an incorrect alerting decision. This effect occurs due to the noise input vector $\mathbf{n}_a(t)$ in the relation between $\mathbf{y}_a(t)$ and $\mathbf{x}(t)$ in Equation (2.5).

An example is shown in Figure 2.10. A measurement-induced alerting error occurs when the true state is outside the alerting threshold, but due to measurement errors the system observes a state vector that is inside Alert Space (i.e., a false alarm). Alternatively, the true state may be inside the threshold but the system uses information that shows the state to be outside the threshold (i.e., a missed detection).

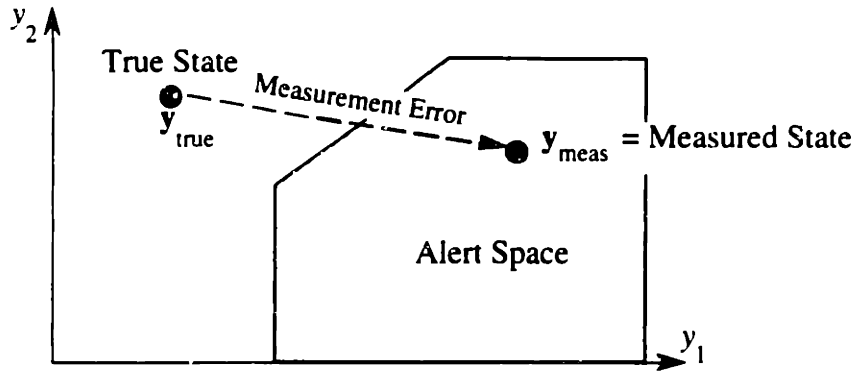


Figure 2.10: Example Measurement-Induced False Alarm
$\mathbf{y}_{meas}(t) = \mathbf{y}_{true}(t) + \mathbf{n}_a(t)$

Depending on the application, measurement errors may or may not be a significant factor in causing false alarms or missed detections. For example, a measurement-induced missed detection may be acceptable if a subsequent measurement results in an alert that still allows the hazard to be avoided.

2.5.2 Uncertainties in the State Trajectory Extrapolation

Uncertainties in the system dynamic model and in the trajectory extrapolation (Section 2.3.1) result in state trajectories that are probabilistic rather than deterministic. In addition, estimates regarding Hazard Space may be in error: the hazard may be larger, more severe, or move along a different trajectory than is assumed. Uncertainties regarding the response of the human to an alert also exist. Failure to follow procedures, high workload situations, and conflicting or misinterpreted information presented to the operator may contribute to a response different from that which is assumed in the design of Alert Space. Delayed reactions or responses that are less aggressive than anticipated may result in the requirement that Alert Space be enlarged to reduce the probability of a missed detection. The effect of these factors is to make the extrapolated trajectory relative to Hazard Space uncertain. It may then be unclear whether an alert is warranted and false alarms or missed detections can occur.

As shown in Figure 2.11, errors can arise from the assumptions that are used when extrapolating the state trajectory when constructing the State-Space Hazard Diagram. For example, the alerting threshold may have been designed assuming a certain response maneuver when an alert is issued as shown in the Figure. In practice, however, some pilots may react differently than assumed. In the example, because the actual avoidance maneuver has a longer latency than anticipated, an incident is not avoided and the situation

can be considered a missed detection. Pilots that are aware of a hazard before an alert is issued may perceive an alert as a false alarm, while other pilots may react too slowly to the alert and an incident may occur.

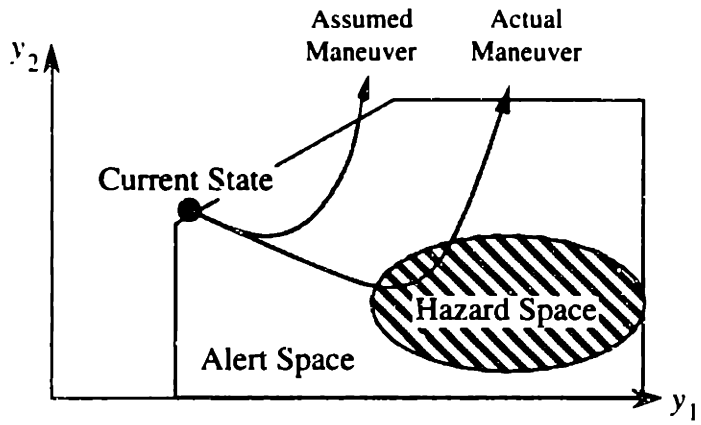


Figure 2.11: Example False Alarm Due to Extrapolation Uncertainty
Alerting threshold based on assumed avoidance maneuver is insufficient to prevent some missed detections.

As a result, extrapolation errors tend to be subjective and may be perceived differently by different operators. Often the term 'nuisance alert' is used instead of 'false alarm' in these cases. The information may be accurate from a measurement standpoint, but is either limited in coverage or the system is too conservative to provide an accurate decision from the standpoint of the pilot's model of the situation.

The method used to extrapolate the state trajectory can also affect false alarms and missed detections. As discussed in Section 2.3.1, some alerting systems may use information about the intended state trajectory. Figure 2.12 shows an example situation that demonstrates the effect of the extrapolation method on the alerting decision. The current state, $\mathbf{y}_a(t)$, is located at point **0** in the Figure.

In Figure 2.12, a projected trajectory based on Current State Extrapolation is a straight line as shown. The true trajectory that is used under Intended Trajectory Extrapolation as discussed in Section 2.3.1 is also shown. For this example it is assumed that there are no additional uncertainties in the extrapolation of the trajectories. If the system is using Current State Extrapolation, an alert is issued at **0** because an incident is projected to occur. Because the true trajectory avoids the hazard, however, the alert could be considered to be a false alarm.

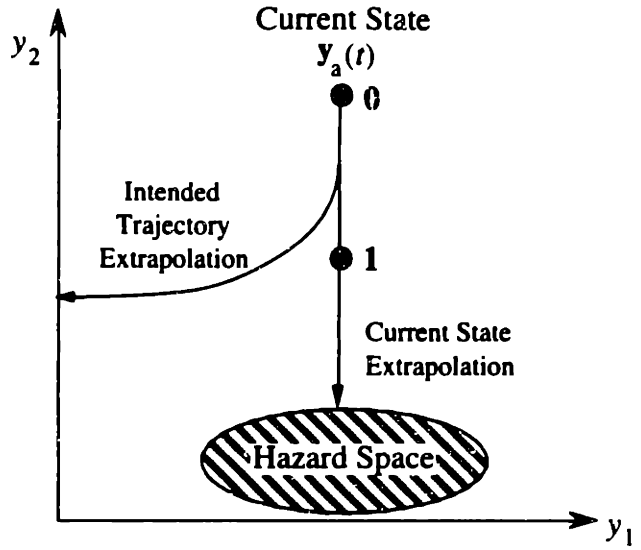


Figure 2.12: Benefit of Using Intended Trajectory Information

If the system is using Intended Trajectory Extrapolation, an alert is not issued because the projected trajectory does not enter Hazard Space. Intended Trajectory Extrapolation therefore prevents a false alarm from occurring.

However, assume that there is some probability that the vehicle does not follow the intended trajectory but may instead follow the straight-line trajectory. When it becomes clear that the state has diverged from the intended path (e.g., at point **1** in the Figure), the system must reject the intended path and revert to Current State Extrapolation. However, the nominal Current State Extrapolation threshold (at point **0**) has already been passed. There may not be sufficient time or space in which to perform an adequate avoidance maneuver, and the situation may be considered to be a missed detection. Therefore, the probability that the assumed trajectory extrapolation is being followed will have an effect on false alarms and missed detections.

Figure 2.13 shows a similar example, but in which Hazard Space lies along the intended trajectory. If using Current State Extrapolation, an alert is not issued at the current time because the extrapolated trajectory does not enter Hazard Space. An alert will be issued in the future, however, as the current state nears the hazard. If the system is using Intended Trajectory Extrapolation, however, the fact that Hazard Space will be encountered is known immediately. In such a case, an alert can be issued at **0**, providing more time and space to the operator to plan for the required avoidance maneuver.

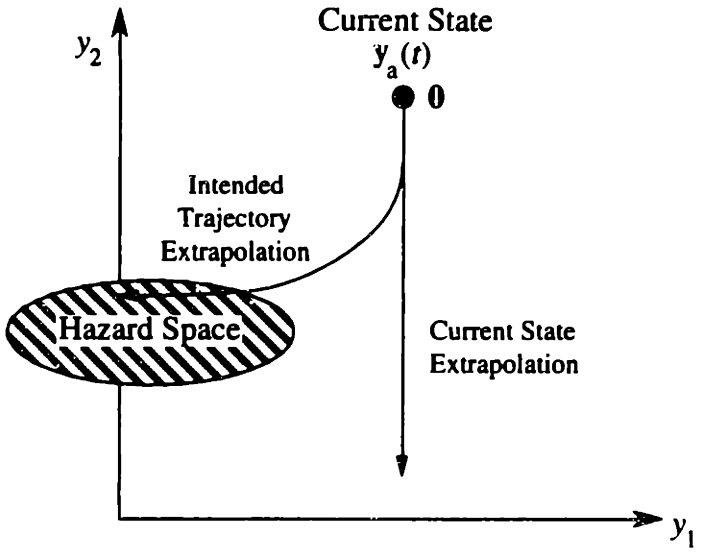


Figure 2.13: Use of Intent Information to Issue an Early Alert

2.5.3 Unobservable States

The processing components of the alerting system (as defined by $\mathbf{G_a}$) may not monitor some of the states that define the complete dynamics of the situation, $\mathbf{x}(t)$. As in control system design, unobservable states may have a large impact on the performance of the system. As confidence in the observed value of one or more states in $\mathbf{x}(t)$ decreases (to the limit that no information about that state is available), the likelihood of false alarms or missed detections generally increases.

Figure 2.14 shows an example State-Space Hazard Diagram for a two-state system. Assume that state $x_2(t)$ cannot be observed by the alerting system: observable state-space is therefore one-dimensional:

$$\mathbf{y_a}(t) = x_1(t) \tag{2.11}$$

Thus, based only on a measurement of $x_1(t)$, the system must correctly determine whether an alert should be issued. A measurement of $x_1(t)$ that is greater than $\mathbf{x}_{1A}$ may or may not indicate a situation leading to an incident. Whether the hazard will be encountered also depends on the value of $x_2(t)$. For example, movement along either trajectory **A** or **B** in Figure 2.14 produces the same measurements of $x_1(t)$; only the values of $x_2(t)$ are different. Trajectory **A** does not lead to a hazardous situation, while trajectory **B** does.

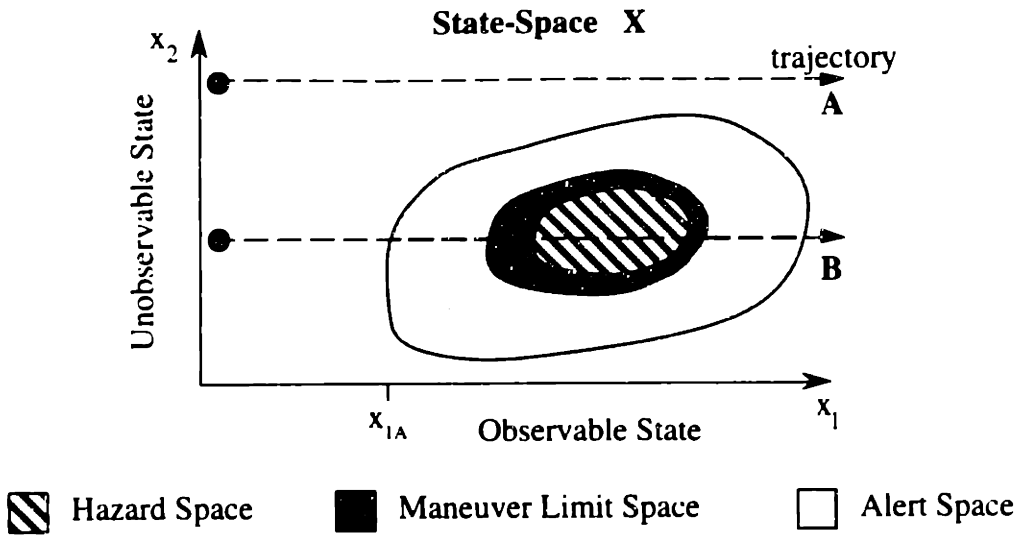


Figure 2.14: Unobservable State Example: Complete State-Space (X)

Because the alerting system operates using only those states that it can observe, the alerting threshold must be considered in observable state-space (**Y**) instead of the total state-space (**X**) shown in Figure 2.14. If the system is designed conservatively (i.e., an encounter with Hazard Space is very undesirable), the alerting threshold would be set at $\mathbf{y}_a(t) = x_{1A}$ as shown on the left in Figure 2.15. With this threshold, when the observable state, $\mathbf{y}_a(t)$, has a value larger than x_{1A}, an alert is issued. This leads to false alarms if $x_2(t)$ is large or small enough that the true state trajectory lies outside Hazard Space. Thus, an unnecessary alarm will occur on trajectory **A**, though the system protects against the hazard along trajectory **B**.

Alternatively, if false alarms are more undesirable than missed detections, the system may be designed such that no alerting thresholds are set at all (right side of Figure 2.15). In this case, there are no alarms regardless of the value of $\mathbf{y}_a(t)$. No alert is issued along trajectory **B**, resulting in a missed detection -- the system does not protect against the hazard.

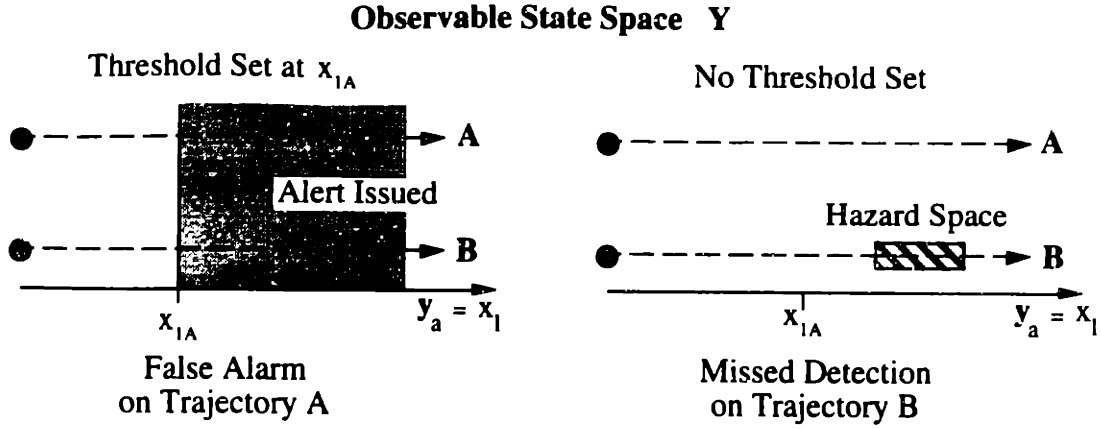


Figure 2.15: Example Threshold Options: Observable State-Space (Based on $y_a(t)$)
Unobservable State $x_2(t)$

2.5.4 Summary

Sections 2.5.1 to 2.5.3 serve to outline the major sources of false alarms and missed detections in the alerting system. The frequency of false alarms and missed detections can generally be reduced by increasing the accuracy and breadth of the data provided to the alerting system. Also, by providing additional information about the dynamics of the encounter (such as using intended trajectory information), the tendency to produce some alerting errors can be reduced.

The effects of uncertainties are such that the extent of Alert Space must be managed to allow for the possibility of overestimating or underestimating the need for an alert. If Alert Space is too large, then false alarms will be produced. If Alert Space is too small, then adequate space will not be available to avoid an incident and missed detections will result. In order to make this tradeoff, however, it is first necessary to be able to quantify the probability of false alarm or missed detection for a given alerting system. It is then possible to examine how the location of the alerting threshold affects the balance between alerting decision errors. To provide the tools needed to perform these evaluations, Chapter 3 discusses alerting thresholds in more detail, and provides a methodology for quantifying the probability of false alarm and missed detection.

3. Probabilistic Analysis of Alerting Systems

Chapter 2 provided a state-space foundation on which to build a model of alerting systems. This chapter takes a probabilistic approach to evaluate the performance of alerting systems in terms of probabilities of false alarm or missed detection. Together, the two chapters provide a methodology to design and evaluate alerting systems.

3.1 Alert Stages

The simplest form of an alerting system uses a *single-stage* process. Single-stage alerting systems transition between one non-hazard state and one *Warning* hazard state. An example single-stage alerting system is shown in Figure 3.1. If the measured hazard level, $y_a(t)$, is below the threshold, no alert is issued. When the measured hazard level is greater than the threshold, a warning is issued that directs the pilot's attention to the threat.

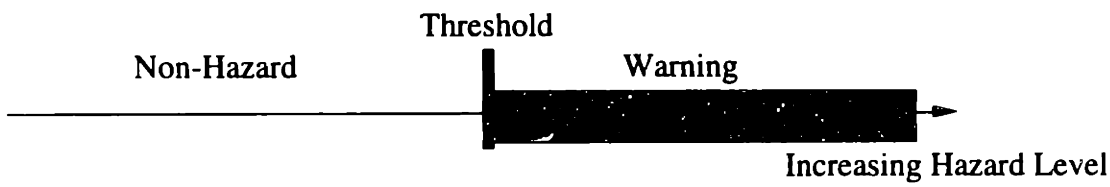


Figure 3.1: Single-Stage Alerting System

In some applications, it is desirable to use a *two-stage* approach. Instead of transitioning directly from a non-hazard state to a warning state, two hazard stages are used: *Caution* and *Warning*. Caution alerts bring the hazard to the attention of the pilot when the situation requires monitoring but may not be sufficiently hazardous to warrant immediate action. Warnings indicate that immediate action is required to avoid a hazard. In a two-stage system (Figure 3.2), a threat generally changes from non-hazard to Caution to Warning as the situation becomes more dangerous. Thus, two-stage systems need two sets of thresholds that differentiate those situations that result in a Caution from those resulting in a Warning. Additional alert stages are possible, but are typically not used to avoid potential confusion among hazard states.

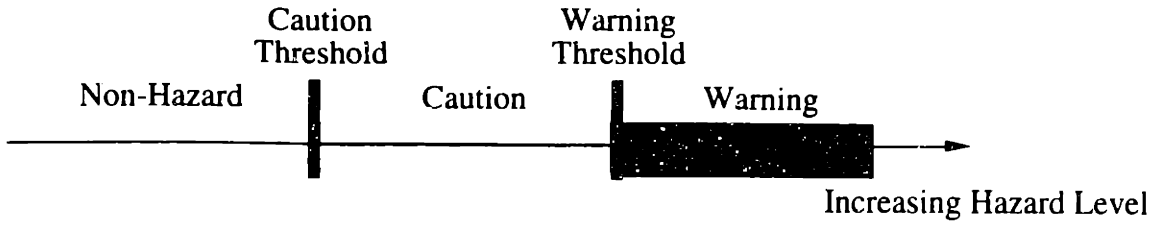


Figure 3.2: Two-Stage Alerting System

Multi-stage alerting systems can be used effectively to manage nuisance alerts. It has been observed that flight crews are less tolerant of false alarms at the Warning stage, while false alarms at the Caution stage are more acceptable (Boucek 1994). Because a Caution is intended only as advisory information, pilots appear to be willing to accept the alert even if it brings attention to a situation that is not truly hazardous. A Warning, however, is a high-stress event requiring immediate action; false Warnings tend to leave a negative impression on the crew, leading to possible distrust of the system in the future.

3.2 Alerting Threshold Issues

3.2.1 The Alerting Threshold

In a typical alerting system, the hazard-related data provided to the system are compared against the set of alerting thresholds that define Alert Space to determine which alert stage should be used. Because the alerting system must make a discrete decision when categorizing a situation into an alert stage, distinct thresholds must be placed to separate those situations that are of one hazard level from those in another. This decision requires measurements[1] of one or more observable states (e.g., distance from Hazard Space) whose values correlate in some way with the true hazard level of the situation.

Figure 3.3 shows an example of a single-stage system. In this example, a large value of the observable state, $\mathbf{y}_a(t)$, corresponds to a high probability that an incident will occur. Therefore, alerts are generated when the state value is higher than the threshold value. Two different measurements of $\mathbf{y}_a(t)$ are shown that were taken at different times. Measurement 1, because its state value is less than the threshold, is not considered indicative of a threat. Therefore, no alert is issued. Measurement 2, however, has a state value higher than the threshold and an alert is issued.

[1] "Measurement" here may mean a direct measurement or the output of an estimation system.

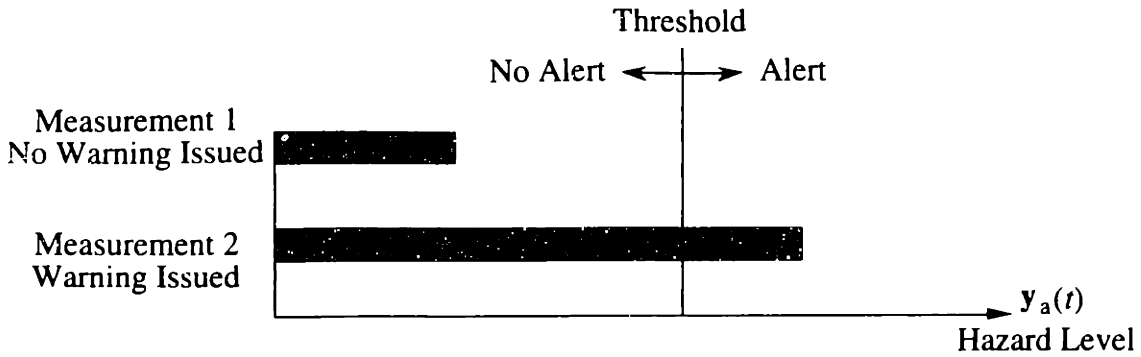


Figure 3.3: Alerting System Threshold

In general, alerting thresholds may be multidimensional, involving several states. In such cases, the threshold is a curve or surface in state-space. For simplicity, one-dimensional alerting thresholds are considered here, though the same issues apply to multidimensional thresholds as well.

3.2.2 The Effect of Uncertainties

Figure 3.4 shows a current state measurement, $\mathbf{y}_a(t)$, that includes some uncertainty. Given $\mathbf{y}_a(t)$, an error ellipse is shown that describes the region in state-space in which $\mathbf{x}(t)$ truly lies with some probability. A region of Hazard Space is also shown in the Figure.

By itself, a state measurement does not indicate whether an incident will occur in the future. Instead, some extrapolated trajectory must be used to determine if an incident will occur. In Figure 3.4, a projected state trajectory, T, is also shown. The extrapolation used to determine T assumes a certain future control input history from the operator and may account for expected changes in the size or location of Hazard Space. The trajectory T is also probabilistic, depending on uncertainties in $\mathbf{y}_a(t)$, in the future control inputs, and in the future behavior of Hazard Space. A probability region is shown in the Figure for T that describes the region in state-space that the true trajectory $\mathbf{x}$ will lie with some probability.

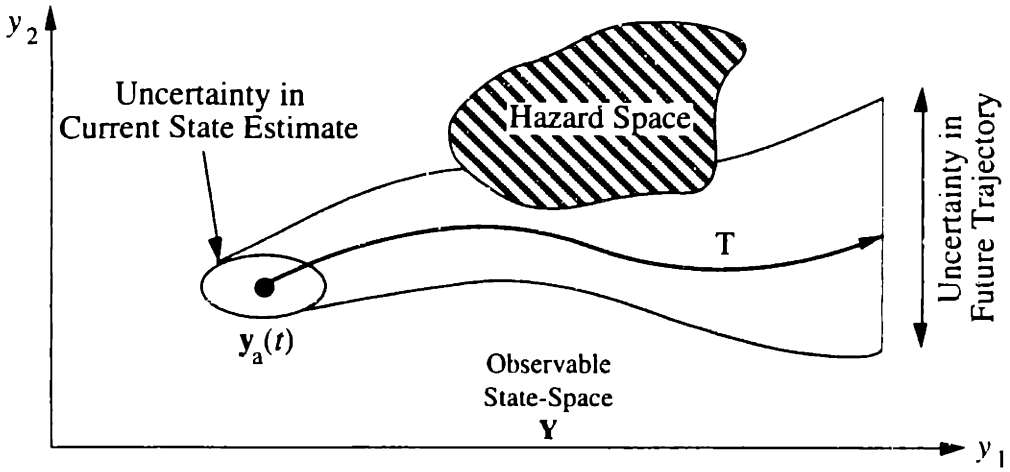


Figure 3.4: Example Probabilistic State Measurement and Projected Trajectory

Having defined T, there is some probability that an incident will occur in the future and some probability that an incident will not occur. The probability that an incident occurs in this situation is denoted $P_T(I \mid \mathbf{y}_a(t))$. This notation explicitly shows that the probability is evaluated along the trajectory T, which is itself a function of $\mathbf{y}_a(t)$ and the uncertainties in the situation.

For simplicity, a one-dimensional situation is used in the following discussion. Given an observable state measurement, $\mathbf{y}_a(t) = y$, the output of the alerting system is considered to be a deterministic event: either an alert is issued or it is not.[2] Because the alerting system's decision may or may not reflect the true situation (that an alert is indeed needed), there are four possible outcomes of any single-stage alerting decision, as shown in Table 3.1. If the situation is truly hazardous (i.e., an incident will occur along T), an alert may be necessary and is termed a true-positive, denoted TP.[3] Similarly, a failure to alert when the situation is non-hazardous (no incident occurs along T) is a true-negative, denoted TN. An alert that is issued in a non-hazardous situation is a false-positive, denoted FP, and likewise, failure to alert to a truly hazardous situation is a false-negative, denoted FN.

---

[2] Note that in signal detection theory, the true situation is typically regarded as deterministic and the measurement is considered to be a random variable (see Appendix B). In this discussion, however, the measurement is considered to be known deterministically and the true situation is treated as a random variable.

[3] "True" because the correct alerting decision is made, and "positive" because an alert is issued. "False" indicates that the alerting decision is wrong, and "negative" indicates that an alert is not issued.

Table 3.1
Alerting Decision Outcomes

	True Situation	
	Non-Hazardous (Incident does not occur along T)	Hazardous (Incident occurs along T)
Alert Not Issued	**TN:** True-Negative	**FN:** False-Negative
Alert Issued	**FP:** False-Positive	**TP:** True-Positive

Figure 3.5 shows a situation where the state $\mathbf{y}_a(t)$ is shown in a State-Space Hazard Diagram in two different locations, marked **1** and **2**. A probabilistic future state trajectory, T, is also shown for each state location. It is then possible to construct a plot of $P_T(I \mid y)$ as $\mathbf{y}_a(t)$ takes on different values, y. As the Figure shows, as the state moves from **1** to **2**, the probability that an incident will occur along T increases.

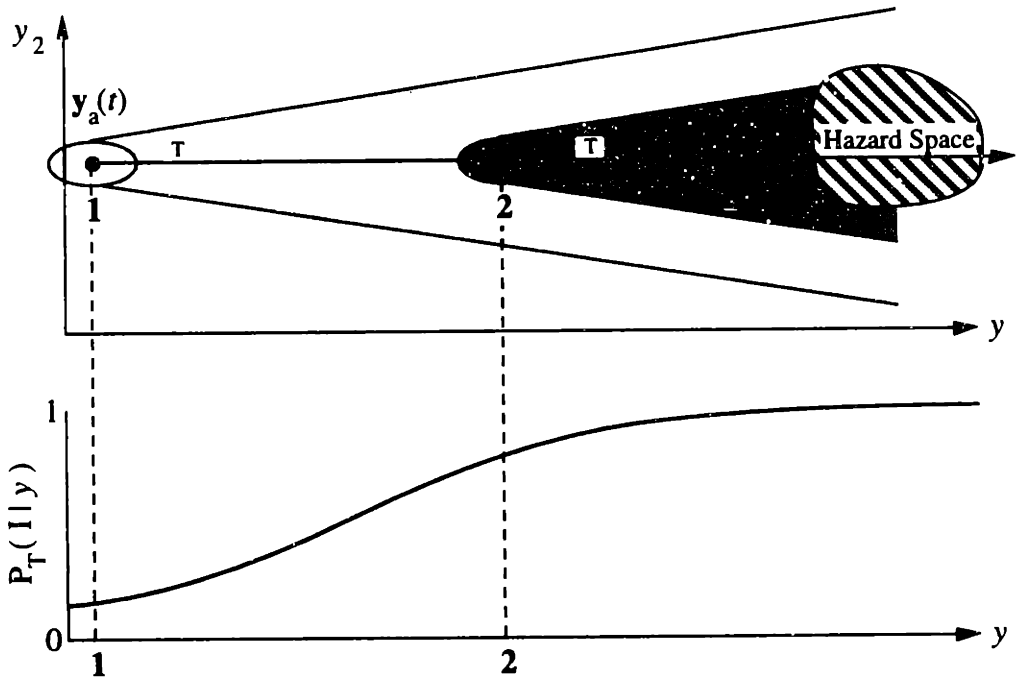
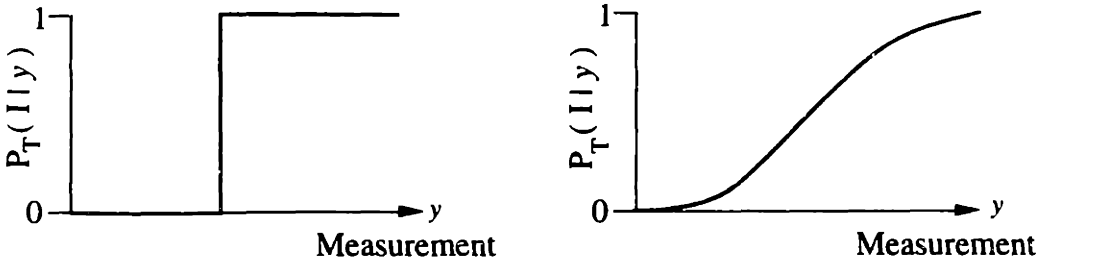


Figure 3.5: Example Evaluation of $P_T(I \mid y)$

In an ideal system where there is a clear distinction between hazardous and non-hazardous situations, $P_T(I \mid y)$ is a sharp step function (shown in Figure 3.6a). In actual systems, uncertainties may hinder the system's ability to accurately determine if an incident

will occur along the projected trajectory. A curve similar to Figure 3.6b results, reflecting the uncertain knowledge of the true situation.



Ideal Alerting System
a

Alerting system with uncertainties
b

Figure 3.6: Effect of Uncertainties on Probability of an Incident

The connection between curves of $P_T(I \mid y)$ and the four alerting decision outcomes is shown in Figure 3.7. The Figure also overlays an alerting threshold: measured hazard levels greater than the threshold produce alerts, and measurements less than the threshold do not. The value of $P_T(I \mid y)$ at a given measurement, y, shows the probability that the situation warrants an alert. The four outcome regions (TP, TN, FP, FN) are also shown.

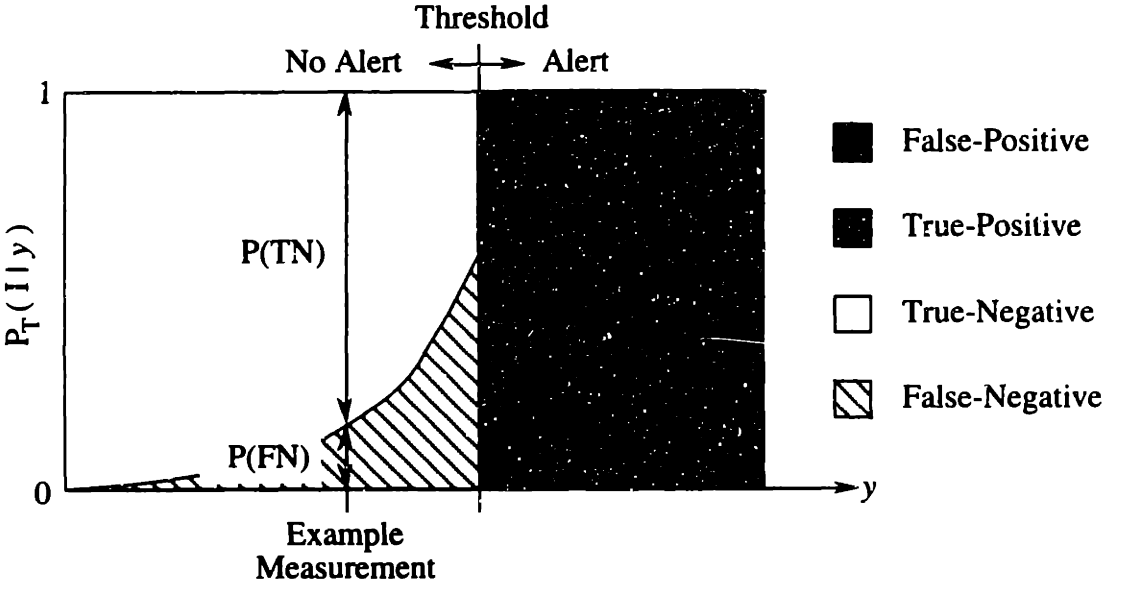


Figure 3.7: Alerting Decision Outcomes
Example curve of $P_T(I \mid y)$ with alerting threshold.

For the example measurement in the Figure, no alert is issued because the measurement is less than the alert threshold. The value of $P_T(I \mid y)$ at that value of y is

therefore the probability of a false-negative, P(FN). The value above the $P_T(I \mid y)$ curve is the probability that the situation is truly not hazardous, P(TN).

Similar behavior occurs when the measurement is greater than the threshold. There exists a probability of a true-positive, P(TP), as well as a probability of a false-positive, P(FP).

Note that $P_T(I \mid y)$ is analogous to a cumulative probability distribution and is not a probability density. It is therefore incorrect to integrate over the four hazard outcome regions to determine system performance. The overall probability of a false-positive, for example, cannot yet be determined because the probability that a given measurement will occur has not yet been accounted for.

Figure 3.8 shows how a change in the alerting threshold location may affect these probabilities. Moving the threshold trades off false-positive probabilities against false-negative probabilities. In Figure 3.8a, an alerting system is shown that has a threshold located at a low hazard measurement value. This threshold selection is representative of systems that have been designed to have a low probability of a false-negative. Relatively small hazard measurements will produce alerts, resulting in higher false-positive probabilities but lower false-negative probabilities than cases in which the threshold is more centrally located on the $P_T(I \mid y)$ curve.

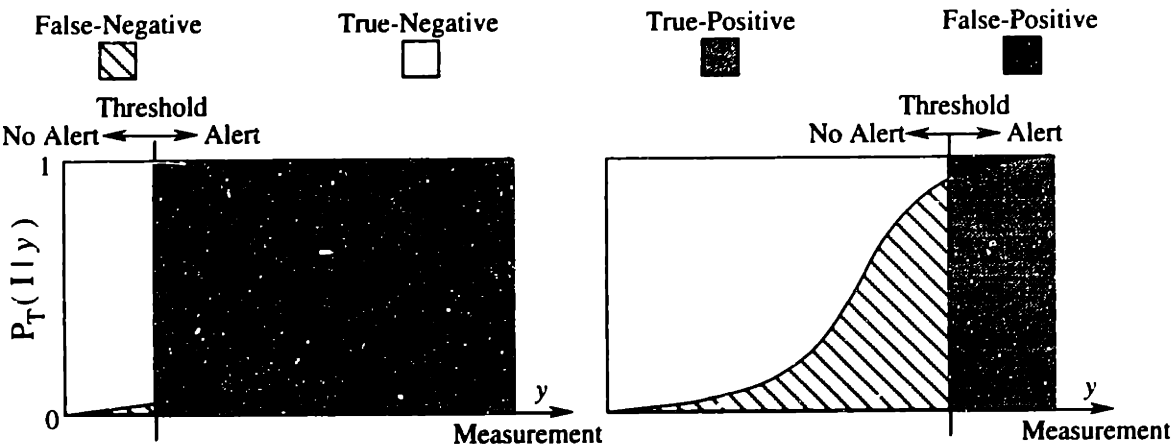
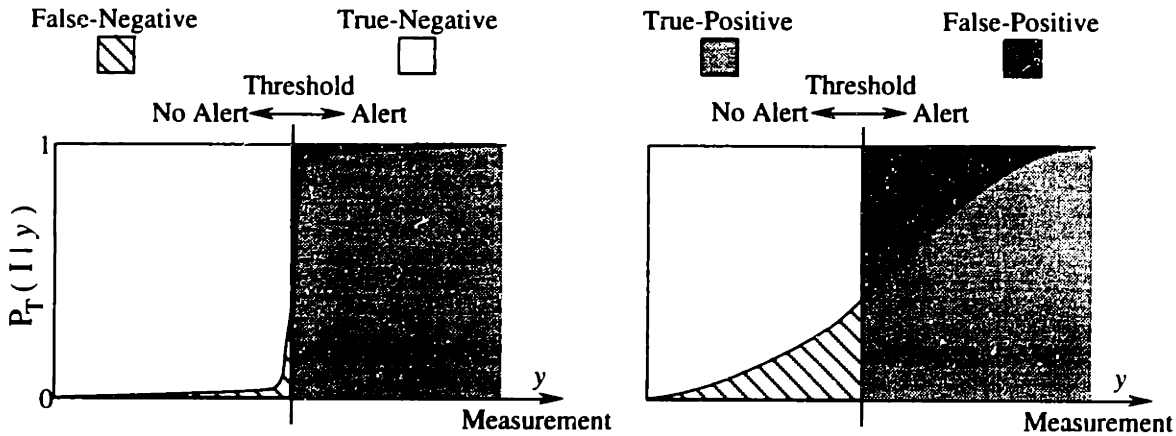


Figure 3.8: Effect of Threshold Placement on P(FP) and P(FN)

In other applications, it may be desirable to alert only when hazard measurements are relatively large and the probability of a false-positive is relatively small. Unfortunately,

this allows for a large probability of a false-negative. Such a system is shown in Figure 3.8b. Thus, the threshold location always results in tradeoffs between probabilities of false-positive and false-negative.

One measure of alerting system performance is the system's tendency to produce false alarms and missed detections. An ideal system produces no false-positives or false-negatives, but actual systems must deal with uncertainties. In Figure 3.9a, a $P_T(I \mid y)$ curve is shown that rapidly changes from 0 to 1 over a small range of measurement values. For most measurement values, the system will perform well with low probabilities of false-positive or false-negative. In contrast, Figure 3.9b shows a system in which $P_T(I \mid y)$ changes gradually over a large range of measurement values. False-positives and false-negatives will occur with some likelihood for a large range of measurement values.



Low-Uncertainty System
a

High-Uncertainty System
b

Figure 3.9: Relationship Between $P_T(I \mid y)$ and System Performance

In general, as more information is made available regarding the hazard, the curve of $P_T(I \mid y)$ will change. Some alerting systems may be designed so that alerts are inhibited until a certain probability of a correct decision can be assured. This is essentially the same as setting an alerting threshold based on a fixed value of P(TP), for example, rather than locating the threshold at a fixed value of y.[4]

[4] Similar logic is followed in methods such as Wald's Sequential Probability Ratio Test (SPRT) (Wald 1957, Tantaratana 1986).

3.2.3 Improvements Through Multi-Stage Alerting

A multi-stage alerting system can be tailored to provide acceptable performance even with a gradually changing $P_T(I \mid y)$ curve by using several alerting stages. Figure 3.10 shows an example of a two-stage system. The left threshold is used to provide Caution alerts, and the right threshold produces Warning alerts. The Caution threshold is used to identify potential threats, possibly even those with a low probability of occurring. A high false-caution probability at this stage may be acceptable if the Caution alerts are treated as awareness or advisory information. The Warning threshold is used specifically for situations in which there is a high probability of an incident. Thus, when a Warning is issued, it typically indicates that immediate action is required to avoid an incident.

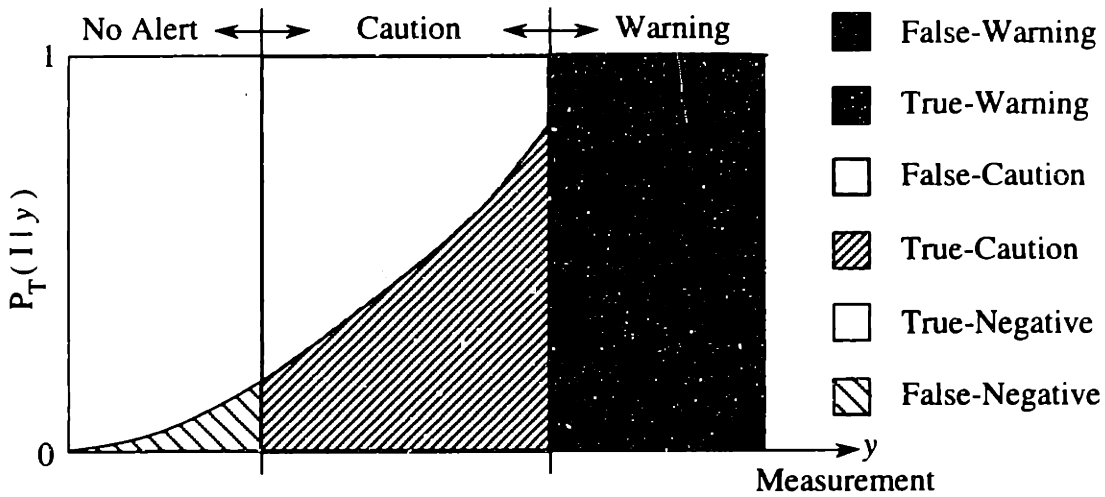


Figure 3.10: Two-Stage System Management of False-Positives and False-Negatives
Caution threshold: high but acceptable false-caution probability.
Warning threshold: low false-warning probability.

3.2.4 Factors Affecting $P_T(I \mid y)$

In addition to using a two-stage alerting method, alerting system performance may also be improved by changing the shape of the $P_T(I \mid y)$ curve. As discussed earlier, a system with a rapidly changing curve provides better performance than one with a slowly changing curve. The factors contributing to the shape of $P_T(I \mid y)$ provide some insight into the places where effort will be productive in improving system performance.

The shape of the $P_T(I \mid y)$ curve is governed by uncertainties from two sources, shown in Figure 3.11. First, the alerting system may have access to only a subset of the total states describing the encounter. As discussed in Section 2.5.3, unobservable states

can produce poor performance because the system must rely upon a few states to describe a more complex situation. The *observable state*, s, used by the alerting system therefore provides only a probabilistic measure of the true state, x. Second, sensor inaccuracies, finite sampling rates, or other error sources produce measurement uncertainties that affect the accuracy with which the *measured state*, y, represents the observable state.

To calculate $P_T(I \mid y)$, it is first necessary to find the probability that a certain value of s produced the given measurement, y. Next, given s, the probability that an incident will occur can be found. With these two steps, $P_T(I \mid y)$ can be calculated for any measurement provided to the alerting system.

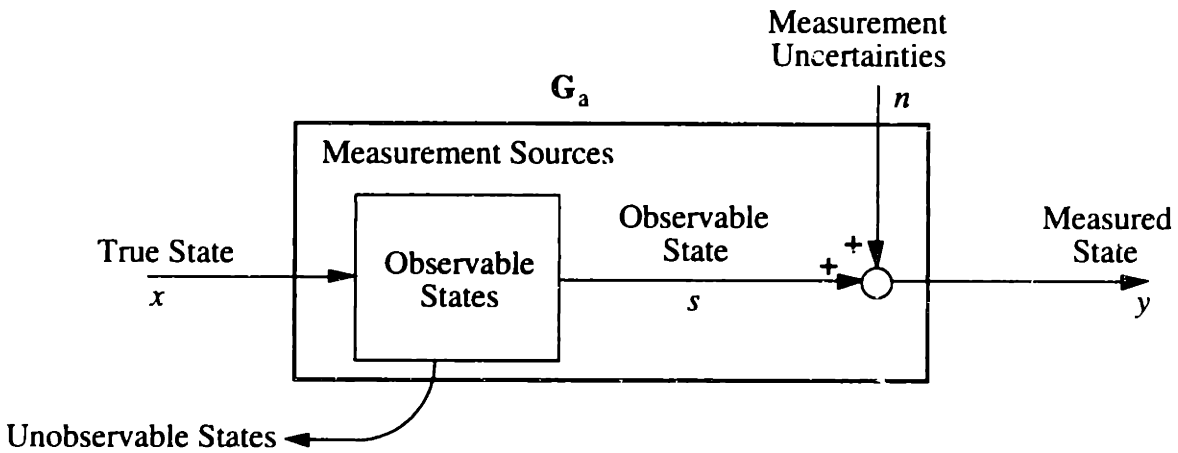


Figure 3.11: Sources of Uncertainty in the Measured State

3.2.4.1 Effect of Measurement Uncertainties

Recall that each state measurement, y, is considered to be a deterministic value. The observable state, s, and measurement noise, n are treated as random variables. Given a measurement, it is then of interest to find the probability that the observable state, s, takes on certain values. Referring to Figure 3.11, s can be calculated as:

$$s = y - n \tag{3.1}$$

Because s is treated as a random variable, a probability density function (PDF), $f_{s|y}(s|y)$, describing the observable state must be found. First, the PDF of the measurement noise, $f_n(n)$, must be specified (an example is shown in Figure 3.12).

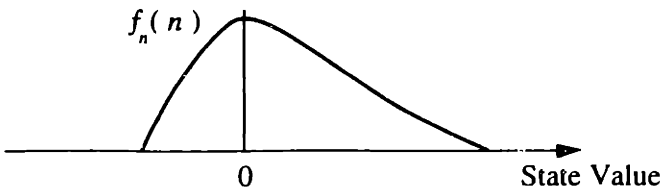


Figure 3.12: Example Measurement Noise Probability Density Function

Because s is the difference of two random variables (Equation 3.1), its PDF is given by a convolution integral:

$$f_{s|y}(s|y) = \int_{-\infty}^{\infty} f_y(y)\, f_n(y-s)\, dy \tag{3.2}$$

However, because y is considered to be deterministic, $f_y(y)$ is a delta function centered at y, and Equation (3.2) reduces to:

$$f_{s|y}(s|y) = \int_{-\infty}^{\infty} \delta(y)\, f_n(y-s)\, dy = f_n(y-s) \tag{3.3}$$

Thus, the PDF for s is given by a reflection of the noise PDF shifted such that it centers upon the measured value. Figure 3.13 shows the PDF of the observed state given a certain measurement and using the same noise PDF as shown in Figure 3.12. Note that the curve in Figure 3.13 is the same as the negative of the noise PDF, shifted to the estimate.[5]

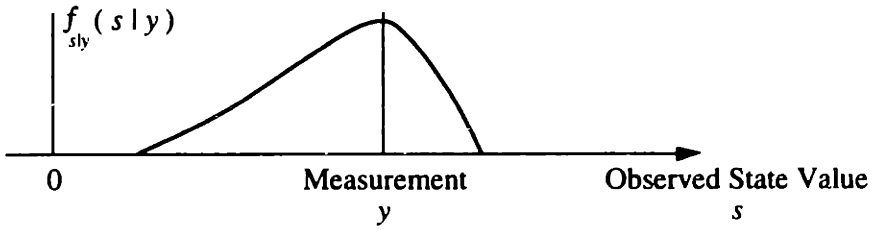


Figure 3.13: Example PDF of Observable State
Same noise PDF as in Figure 3.12.

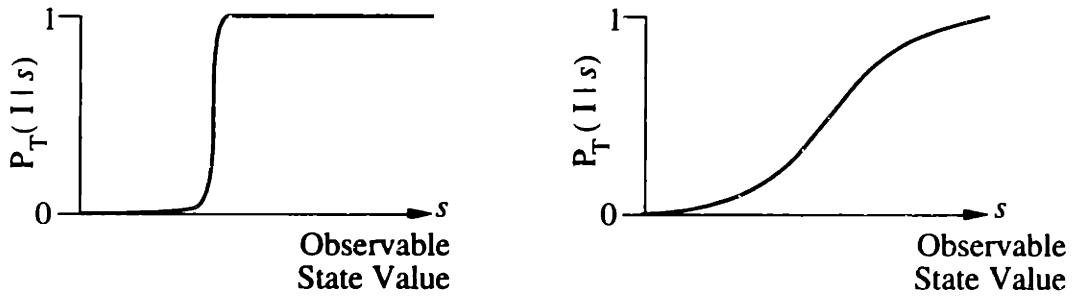
Therefore, given a PDF for the measurement noise, the conditional PDF of the observed hazard level can be found.

[5] The negative mapping is due to the $-n$ term in Eq. (3.1) and the $-s$ term in Eq. (3.3).

3.2.4.2 Effect of State Observability

Because of unobservable states and uncertainties, the hazard metric (e.g., time to impact) that is used may not accurately correlate with the true probability that an incident will occur, further impairing effectiveness. Even if a perfect measurement were available (i.e., $y = s$), the state being used may be such that s does not correspond well with the true severity of the situation. A probability function, denoted $P_T(I \mid s)$, is constructed that describes the probability that an incident will occur given an observable hazard level value. Note that this is similar to but not the same as $P_T(I \mid y)$. $P_T(I \mid y)$ describes the probability of an incident occurring given a state measurement that includes uncertainties. $P_T(I \mid s)$ describes the probability of an incident occurring given the observable state before measurement uncertainties have been included.

Two examples are shown in Figure 3.14 for different observable states. Figure 3.14a shows an observable state that clearly indicates whether an incident will occur: $P_T(I \mid s)$ changes rapidly across a small range in s. The right plot shows a weaker correlation between s and I; only for extreme values of s can the status of the situation be known with certainty.



High-Accuracy Observable State
a

Low-Accuracy Observable State
b

Figure 3.14: Example $P_T(I \mid s)$ Curves

Using the two descriptions of uncertainty given by $f_{s|y}(s|y)$ and $P_T(I \mid s)$, the probability, $P_T(I \mid y)$, that a certain situation is hazardous, given a measurement, may be found. Because each possible observable hazard level, s, is a mutually exclusive event, $P_T(I \mid y)$ is calculated as the integral of the probability of s occurring given y, times the

probability of I occurring given s:

$$P_T(I \mid y) = \int_{-\infty}^{\infty} f_{s|y}(s \mid y)\ P_T(I \mid s)\, ds$$
$$= \int_{-\infty}^{\infty} f_n(y - s)\ P_T(I \mid s)\, ds \qquad (3.4)$$

Thus, the probability that a situation is hazardous based on a measurement of a state is the convolution integral of the measurement noise PDF with $P_T(I \mid s)$. Due to the convolution integral, $P_T(I \mid y)$ may be thought of as a stretched-out version of $P_T(I \mid s)$ where the degree of stretching depends on the measurement noise PDF.

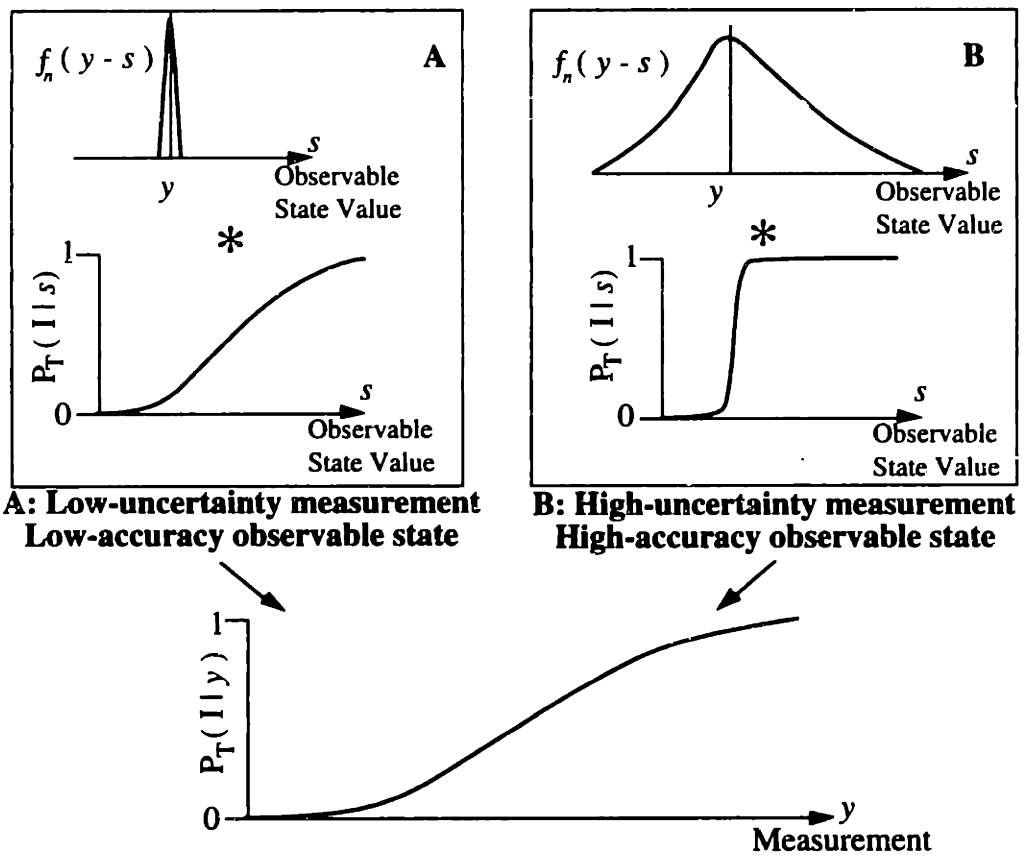


Figure 3.15: Construction of $P_T(I \mid y)$ Curves

Figure 3.15 demonstrates how different combinations of state measurement uncertainty and observable state accuracy can produce similar $P_T(I \mid y)$ curves. On the left side of the Figure (marked **A**), a situation is shown in which measurement uncertainties are small but the observable hazard metric is not accurate in distinguishing a hazard from a non-hazard. On the right side (marked **B**), a situation is shown in which measurement

uncertainties are large but the observable hazard metric can accurately distinguish a hazard from a non-hazard. Both situations produce similar $P_T(I \mid y)$ curves. In each case, however, the area in which effort needs to be focused to improve system performance may be different. In situation **A**, effort might best be made in the improvement of the observability of the state x. Additional sensors or other information may provide this improvement. In situation **B**, effort could be spent on improving the accuracy of the measurements being used by the system.

3.2.5 Calculation of $P_T(I \mid y)$

The previous discussions have assumed that some means is available with which to calculate $P_T(I \mid y)$ given a probabilistic trajectory T in a specific situation described by y. A generalized method for obtaining $P_T(I \mid y)$ has been developed as part of the methodology and is discussed in detail in Appendix A.

The basic methodology for calculating $P_T(I \mid y)$ is shown in Figure 3.16. One requirement is a set of equations that describe the dynamics of the aircraft when following a given trajectory. Probability density functions describing the uncertainties in the parameters that define the trajectory are also needed. For example, measurement uncertainties, the response time delay, and the aggressiveness of the avoidance maneuver can each be described by appropriate PDFs.

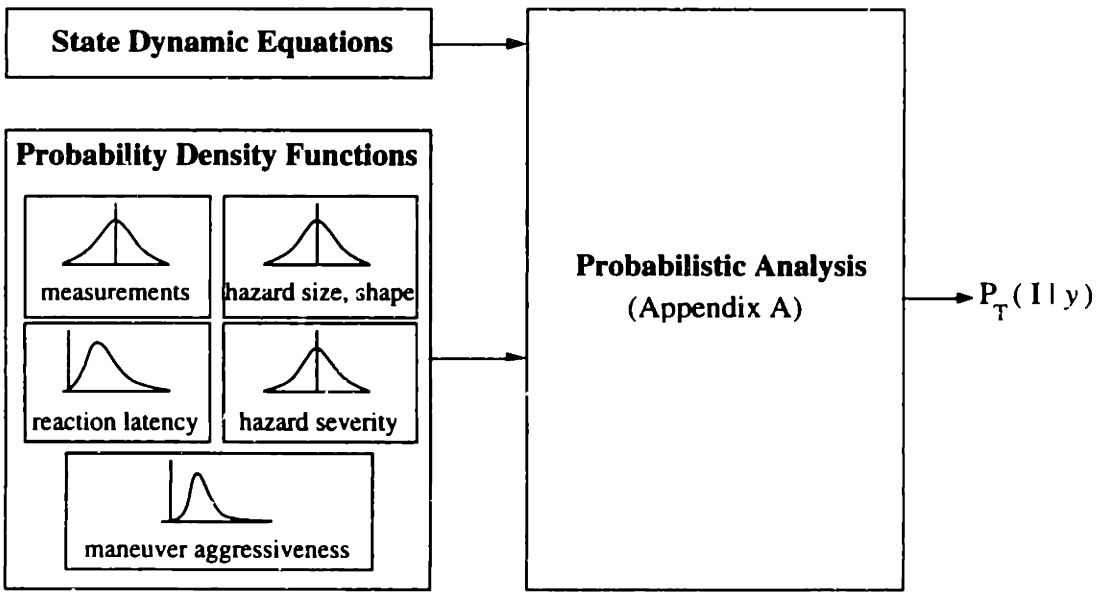


Figure 3.16: Probabilistic Analysis Methodology (Appendix A)

The procedure in Appendix A takes the equations of the dynamics and the PDFs and, through numerical integration or Monte Carlo evaluation, produces the probability of an incident. The flexibility of the procedure lies in its ability to treat the PDFs as generic modules, allowing for a rapid comparison between the effects of different PDFs on $P_T(I \mid y)$.

The PDFs describing the various parameters that affect the situation can be obtained through hardware specifications or through a statistical study of the situation. Actual events or simulation results, for example, can be used to build the statistics of a PDF.

3.3 Evaluation of Alerting System Performance

Section 3.2 discussed the probability of a false-positive or false-negative for a certain trajectory T based on a single measurement $\mathbf{y}_a(t) = y$. Typical situations involving hazards, however, provide a sequence of measurements of the hazard level to the alerting system over a period of time. The important metric in this case is not whether a single measurement results in an erroneous decision, but whether the alerting system, over time, provides the correct decision based on the sequence of measurements. For example, a false-negative as defined in Section 3.2 is a case where the alerting system fails to notify the crew about a hazard at that moment. It says nothing about whether the crew will eventually get an alert and still avoid the hazard.

The alerting decision must balance the need to alert the operator sufficiently early that an incident can be avoided against the desire to only alert when absolutely necessary. To determine if an alert is warranted in a given situation, it is therefore necessary to examine the hypothetical outcomes of the alert / do not alert decision. As is shown in this Section, the alert / do not alert decision is analogous to the signal detection problem of determining if a signal is present in background noise. A brief summary of Signal Detection Theory is provided in Appendix B.

3.3.1 Nominal and Avoidance Trajectories

If no alert is issued, the state continues along what is termed the projected *Nominal Trajectory*, denoted N. In response to an alert, there is, in general, a discrete change in the actions of the operator. When an alert is issued, the state follows a different trajectory that may avoid an incident, termed the *Avoidance Trajectory*, denoted A. Both N and A are, in general, probabilistic just as T was in Section 3.2. In particular, A may include some

probability that no action is taken in response to the alert. The uncertainties in N and A are functions of the factors discussed in Section 2.5.

Following the notation used earlier, the probability that an incident will occur given a measurement y and assuming that the Nominal Trajectory is followed is denoted $P_N(I \mid y)$. Similarly, $P_A(I \mid y)$ describes the probability that an incident occurs along the Avoidance Trajectory (i.e., that the avoidance maneuver fails to avoid the hazard). Note that the Avoidance Trajectory may also encounter other hazards that would not have been encountered had no alert been issued. It is therefore possible to examine both the positive and negative effects of the alerting system using this method.

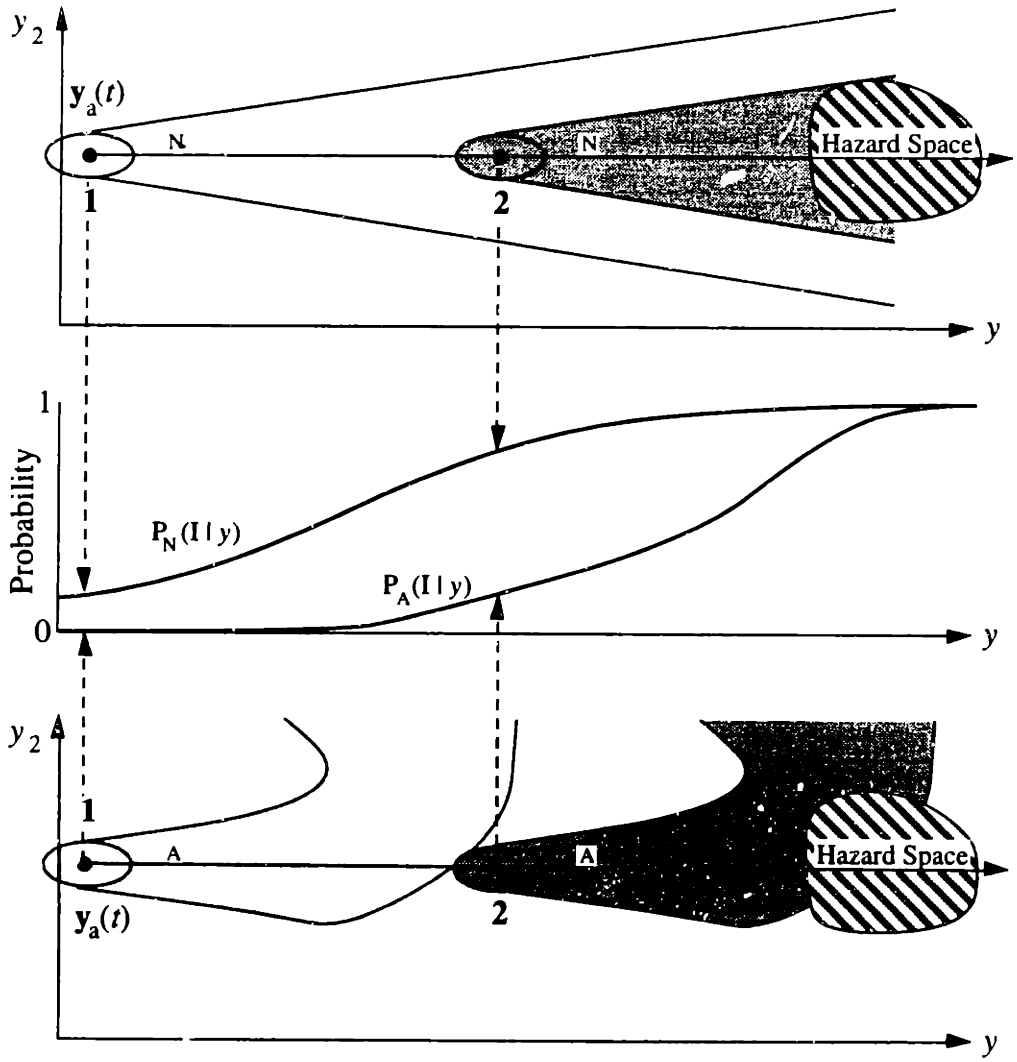


Figure 3.17: Example Construction of $P_N(I \mid y)$ and $P_A(I \mid y)$

Figure 3.17 shows how a combined plot of $P_N(I \mid y)$ and $P_A(I \mid y)$ can be constructed. $P_N(I \mid y)$ is obtained using Nominal Trajectories in which it is assumed that an

alert is not issued in the same manner in which $P_T(I \mid y)$ was constructed in Figure 3.5. $P_A(I \mid y)$ is calculated using Avoidance Trajectories that assume that an alert is issued at the current state value. In this example, the Avoidance Maneuvers include a response time delay before maneuvering begins. As shown in Figure 3.17, if an alert is issued at location **1**, the probability of an incident along A is essentially zero. If an alert is delayed until **2**, there is some probability that Hazard Space will not be avoided, but there still is some improvement over the Nominal Trajectory case at **2**.

A second example plot of $P_N(I \mid y)$ and $P_A(I \mid y)$ is shown in Figure 3.18. In this example, as the measurement value increases the probability that an incident will occur along either trajectory also increases. When an alert is not issued, the state follows trajectory N and the probability that an incident will occur is shown by $P_N(I \mid y)$ -- the value of the upper curve. When an alert is issued, trajectory A is followed and the probability that an incident will occur is given by $P_A(I \mid y)$ -- the value of the lower curve. The difference in probability values between the upper and lower curves at the alerting threshold location represents the effective *System Benefit*. System Benefit is the decrease in the probability of an incident that is possible because the alerting system is operating.

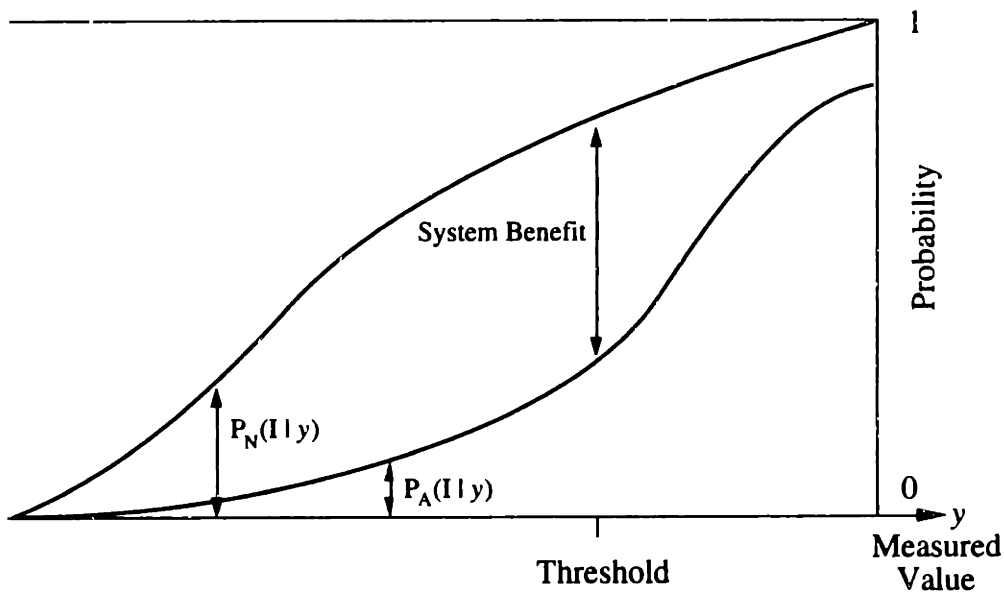


Figure 3.18: Example Plot of $P_N(I \mid y)$ and $P_A(I \mid y)$
Upper curve: Nominal Trajectory (no alert).
Lower curve: Avoidance Trajectory (alert is issued).

Another measure of system performance is the *Safety Ratio*, denoted *SR*. The Safety Ratio is the ratio between the probability of an incident that would occur with the

alerting system in operation to the probability of an incident that would occur if the system was not in operation.

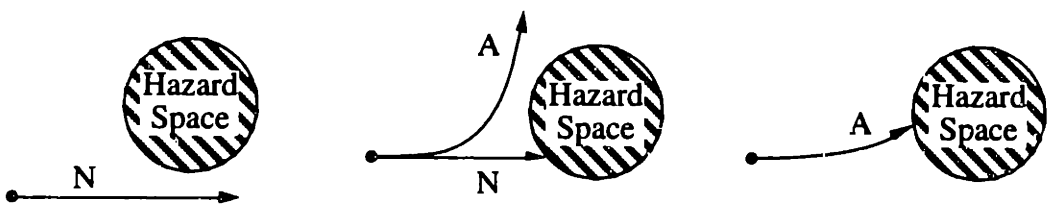
$$SR = \frac{P_A(I \mid y)}{P_N(I \mid y)} \tag{3.5}$$

A Safety Ratio of 0 indicates a system that provides perfect protection against incidents. A Safety Ratio of 1 indicates a system that is of no additional benefit in protecting against hazards. Safety Ratios above 1 are possible, but signify alerting systems that actually increase the probability of an incident. For example, an escape maneuver could avoid one hazard but increase the probability of encountering a second.

3.3.2 Alerting Decision Outcomes

In Figure 3.19, three situations are shown where the state measurement is in the vicinity of Hazard Space. It is assumed for the purposes of this example that the trajectories in the Figure are known perfectly. Following discussions relax this assumption. Therefore, the state follows the Nominal Trajectory exactly if no alert is issued, and follows the Avoidance Trajectory exactly if an alert is issued.

In Figure 3.19a, a projected Nominal Trajectory is shown that avoids Hazard Space. Because an incident is avoided without any action from the alerting system, an alert should not be issued in this situation. Correctly refraining from issuing an alert in this situation is a *correct rejection*, CR. If an alert is issued, however, it is a *false alarm*, FA.



No alert is needed **a**

Alert may be needed **b**

Missed Detection **c**

Figure 3.19: Potential Alerting Decision Outcomes

In Figure 3.19b, a situation is shown in which N leads to Hazard Space but A does not. In this case, an alert allows the pilot sufficient time to avoid an incident. If an alert is issued in this situation, it is termed a *correct detection*, CD. However, it may be true that the alert could be delayed and still provide time to avoid an incident. If an alert is not issued in this case, it is termed *deferred action*, DA.

In Figure 3.19c, a third situation is shown in which A leads to Hazard Space. In this case, the alert is issued too late to allow the pilot to avoid an incident, and the situation is a *missed detection*, MD. Note that a hazard encountered along A may be different from a hazard encountered along N.

These potential alerting decision outcomes are summarized in Figure 3.20.

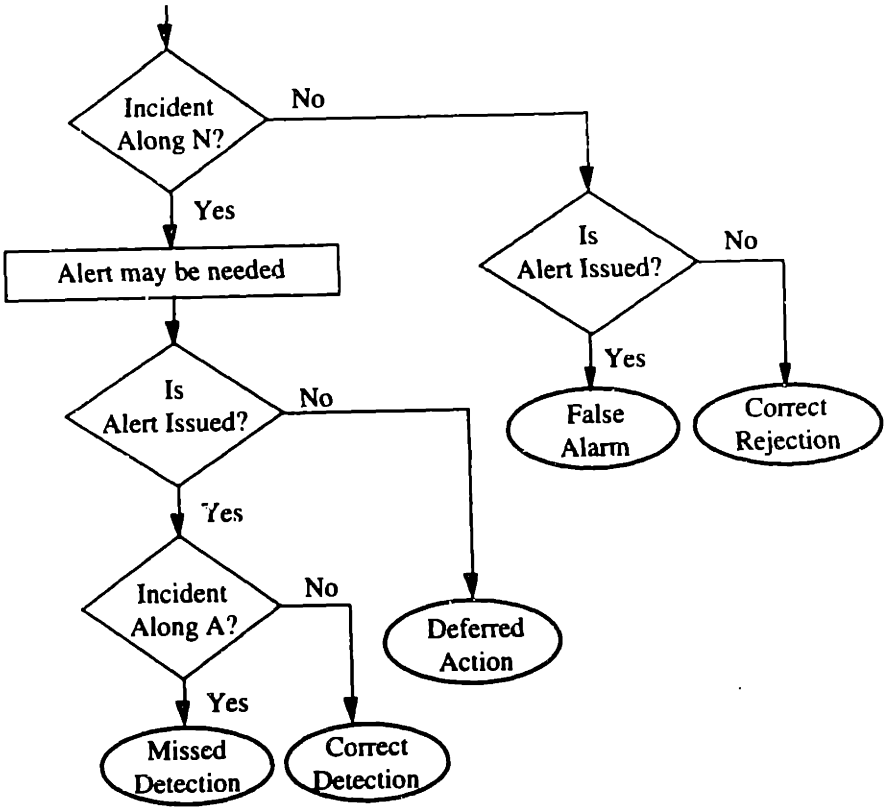


Figure 3.20: Alerting Decision Outcome Flowchart

Following the definitions of the alerting decision outcomes from above, the probability of a false alarm, P(FA), is the probability that a hazard will not be encountered along N:

$$P(FA) = P_N(\overline{I} \mid y) = 1 - P_N(I \mid y) \tag{3.6}$$

Note that P(FA) is implicitly a function of the state measurement, y, and the assumed Nominal Trajectory, N.

The probability of a missed detection, P(MD), is similarly given by:

$$P(MD) = P_A(I \mid y) \tag{3.7}$$

The probability of a correct detection is the probability that an incident does not occur along trajectory A:

$$P(CD) = P_A(\overline{I} \mid y) = 1 - P_A(I \mid y) = 1 - Pr(MD) \qquad (3.8)$$

When an alert is issued, the three probabilities, P(FA), P(MD), and P(CD), provide a measure of system performance in a manner analogous to the same measures used in Signal Detection Theory (see Appendix B). It is desirable to minimize P(FA) and P(MD) to obtain an acceptable alerting system. These objectives cannot be met simultaneously but must be managed to produce a system that meets specifications for acceptable false alarm or missed detection probabilities.

Plots of $P_N(I \mid y)$ and $P_A(I \mid y)$ can also be used to examine the effect of different threshold locations on false alarms and missed detections, as shown in Figure 3.21. For example, P(MD) is given by the value of $P_A(I \mid y)$ evaluated at the location of the alerting threshold. The probability value above the curve of $P_N(I \mid y)$ represents P(FA).

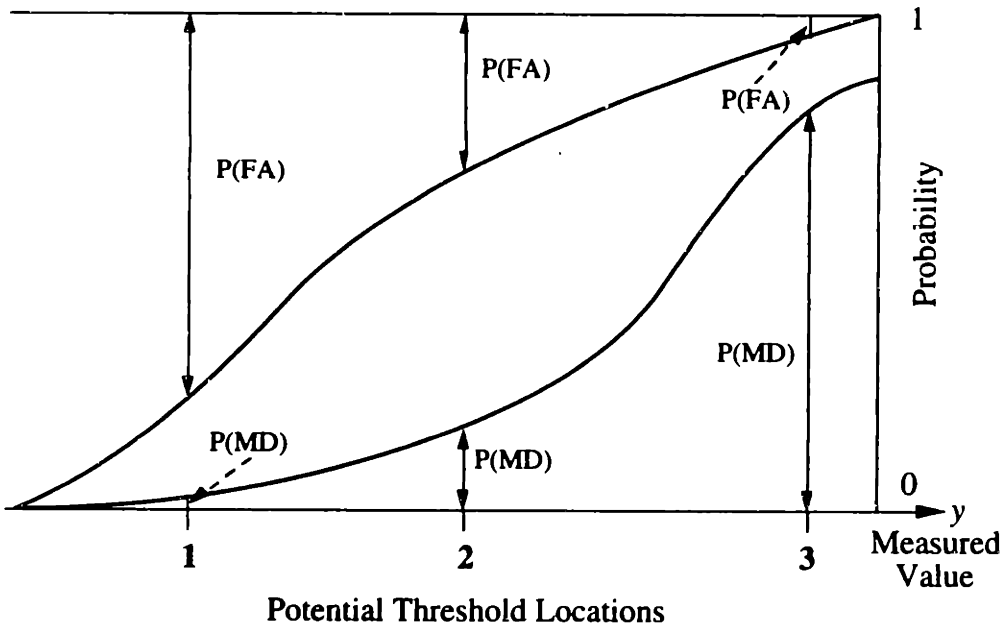


Figure 3.21: Relationship Between Alerting Threshold Location and P(FA) and P(MD)

Figure 3.21 shows three potential alerting threshold locations for the same curves that were shown in Figure 3.18. Recall that it is desired to minimize P(MD) while also minimizing P(FA). Minimizing P(MD) suggests alerting early, at low measured hazard levels, such as threshold 1 in Figure 3.21. However, P(FA) will increase as the threshold is moved to the left, contrary to the desire to minimize P(FA). A threshold far to the left

reduces the System Benefit because the high false alarm probability reduces the effectiveness of the system. Threshold **2** in the Figure is more moderate and produces a value of P(MD) similar to P(FA). If the threshold were moved to location **3**, there would be a small probability of false alarm but a large probability of missed detection. A placement of the alerting threshold far to the right in Figure 3.21 also results in a reduction in the System Benefit: the system waits too long before alerting and missed detections occur. Thus, the graph provides a view of the tradeoff between missed detections and false alarms.

3.3.3 System Operating Characteristic Curves

Now that the probabilities of missed detection and false alarm can be found, methods from Signal Detection Theory can be used. In particular, System Operating Characteristic (SOC) curves are plots of P(CD) vs. P(FA) and are analogs of Receiver Operating Characteristic Curves (ROC) in Signal Detection Theory (Appendix B).

Recall from Equation (3.8) that P(CD) = 1 - P(MD). Thus, SOC curves can be generated directly from plots of $P_A(I \mid y)$ and $P_N(I \mid y)$ such as those used in Figure 3.21. Figure 3.22 is an example SOC curve for the same system shown in Figure 3.21.

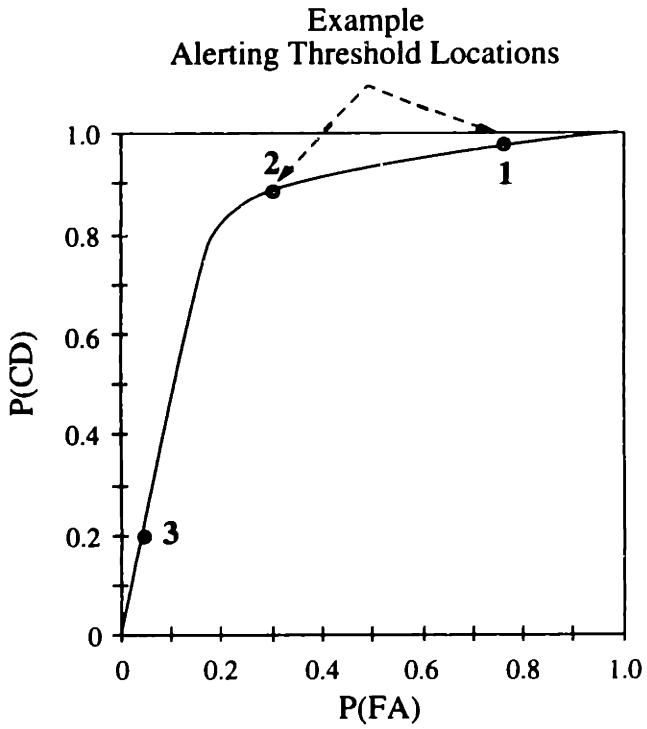


Figure 3.22: Example SOC Curve

SOC curves show the tradeoff between P(FA) and P(CD) as a function of the alerting threshold location. Each choice of an alerting threshold maps onto a single point along the SOC curve. In Figure 3.22, the three example alerting threshold locations from Figure 3.21 are shown.

Thus, given a definition of the system that produces an SOC curve, the threshold location can be chosen according to the relative desirability of false alarms and correct detections. However, unless changes are made in the system design or in the accuracy of the information that is available, the system's performance will be constrained to follow the SOC curve.

Given a certain system design in terms of sensor accuracy, for example, the tradeoff between P(FA) and P(CD) is constrained to lie on a single SOC curve, as shown in Figure 3.23. P(FA) can then be balanced against P(CD) by changing the alerting threshold location, which changes the operating point on the SOC curve.

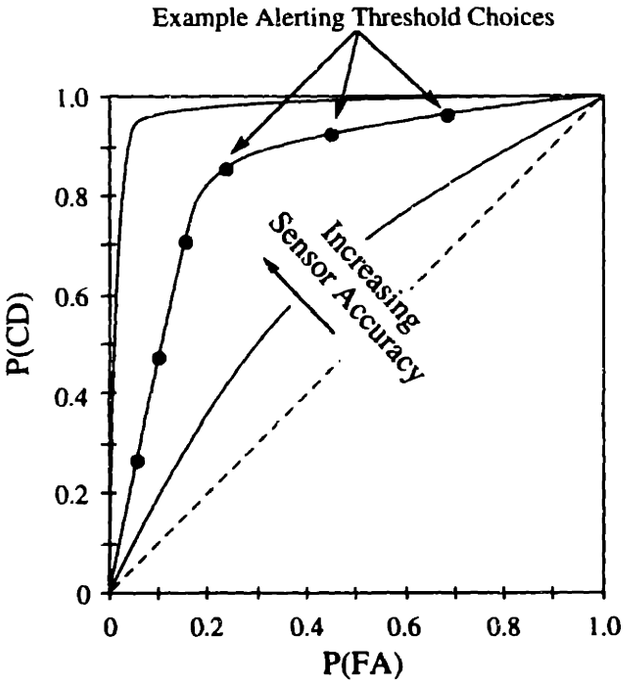


Figure 3.23: Control Over P(FA) and P(CD) with SOC Curve
System design sets the SOC curve, and the alerting threshold location results in an operating point on that curve.

Note, however, that the SOC curve only provides an indication of system performance for the specific situation used to calculate P(FA) and P(CD). A more complete

description of system performance over all operating conditions requires consideration of the exposure to various situations (this is covered in Section 3.4).

As was the case in Section 3.2, system performance can be improved through the use of a two-stage alerting system or by changing the shape of the $P_A(I \mid y)$ and $P_N(I \mid y)$ curves. It should also be noted that an actual alerting system may have some probability that it will not operate properly (i.e., no alert is issued for a measurement greater than the threshold). In such a case, the additional probability of an incident that results from a failure to operate also needs to be accounted for.

3.3.4 Threshold Placement

Once an SOC curve has been developed for an alerting system, it can be used to determine where the alerting threshold should be set. If a constraint on the maximum acceptable value of P(FA) or P(MD) is available, the threshold location can be set directly on the SOC diagram.

If the relative costs of false alarms (C_{FA}) and missed detections (C_{MD}) can be quantified, then an alerting cost function, J, can be defined that weighs these costs by the probability of each undesirable outcome.

$$J = P(FA)\, C_{FA} + P(MD)\, C_{MD} \tag{3.9}$$

Rewriting Equation (3.9) in terms of P(CD),

$$J = P(FA)\, C_{FA} + (1 - P(CD))\, C_{MD} \tag{3.10}$$

To minimize the cost of alerting (and to have the optimal alerting threshold), the derivative of the cost function with respect to the alerting threshold location is set equal to zero:

$$dJ = dP(FA)\, C_{FA} - dP(CD)\, C_{MD} = 0 \tag{3.11}$$

Rearranging,

$$\frac{dP(CD)}{dP(FA)} = \frac{C_{FA}}{C_{MD}} \tag{3.12}$$

Note that the left side of Equation (3.12) is the slope of the SOC curve. Therefore, the alerting threshold is set to that point on the SOC curve that has a slope equal to the ratio of the costs associated with alerting.

Figure 3.24 shows an example SOC curve with an alerting threshold placed for a system in which the costs of false alarm and missed detection are equal. If instead the relative cost of false alarms were high, the threshold would be moved left in the Figure, toward the high-slope region of the SOC curve.

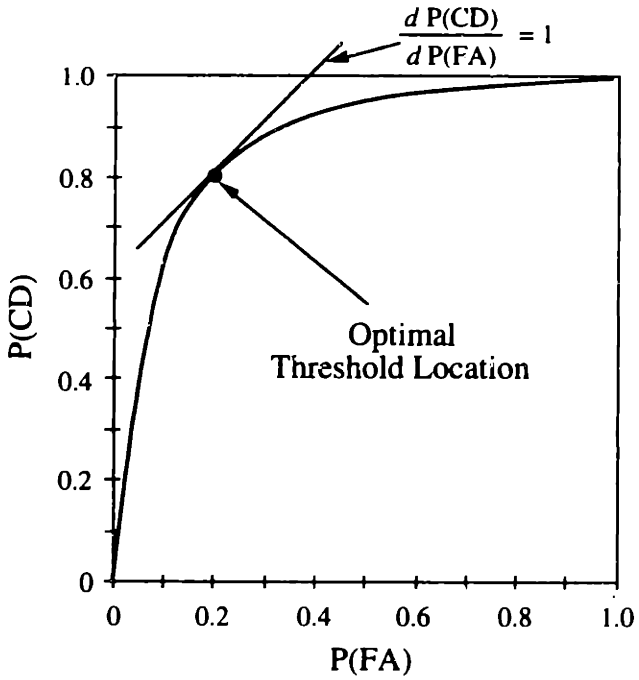


Figure 3.24: Optimal Threshold Placement Using Cost Ratio

3.3.5 Negative Effects of an Alert

Using the methodology covered in this thesis, it is also possible to examine the potentially negative effects of alerting. For example, an alert may avoid one hazard but result in an encounter with a different hazard. Figure 3.25 shows an example situation in which an alerting system is designed to avoid Hazard A (representing another aircraft, for example). The Avoidance Trajectory followed when an alert is issued is shown in gray. If an alert is issued at point **1**, however, the aircraft instead encounters Hazard B (representing terrain, for example). The alert should be delayed until the state lies near point **2**, in which case both hazards are avoided.

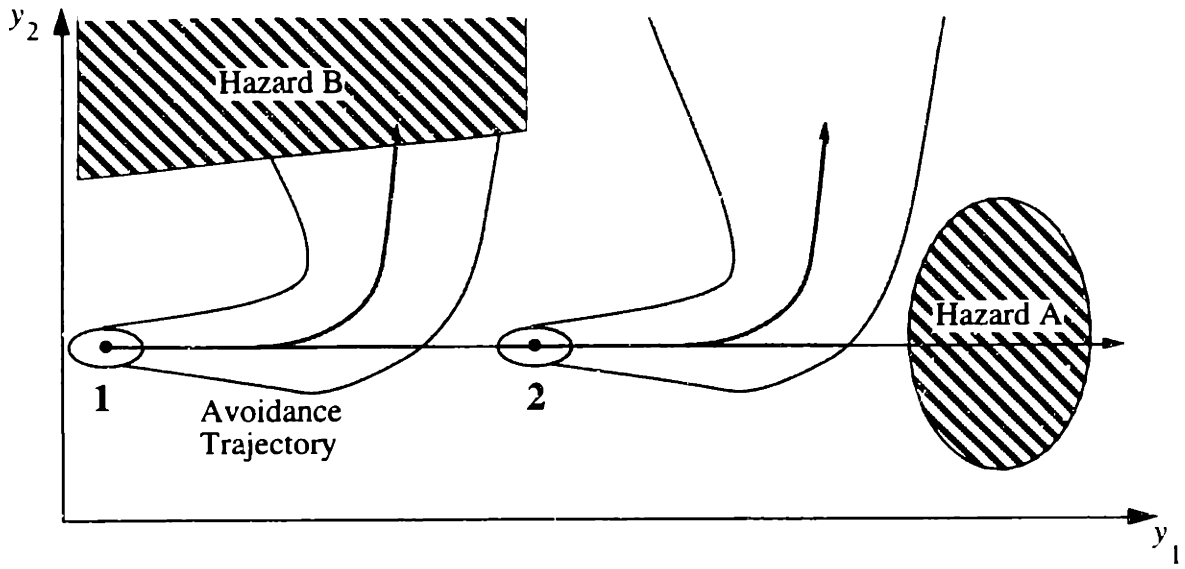


Figure 3.25: Example of Negative Effect of Alerting

The same analysis techniques described above can be used in this type of situation. Figure 3.26 shows an example plot of $P_A(I \mid y)$ and $P_N(I \mid y)$ for the situation shown in Figure 3.25. Note that the curves of $P_A(I \mid y)$ and $P_N(I \mid y)$ need not be monotonically increasing. In addition, there is a region of negative system benefit, centered on state value **1**, corresponding to the region in which Hazard B is encountered. If the alert is delayed to **2**, both hazards are avoided (i.e., P(MD) = 0).

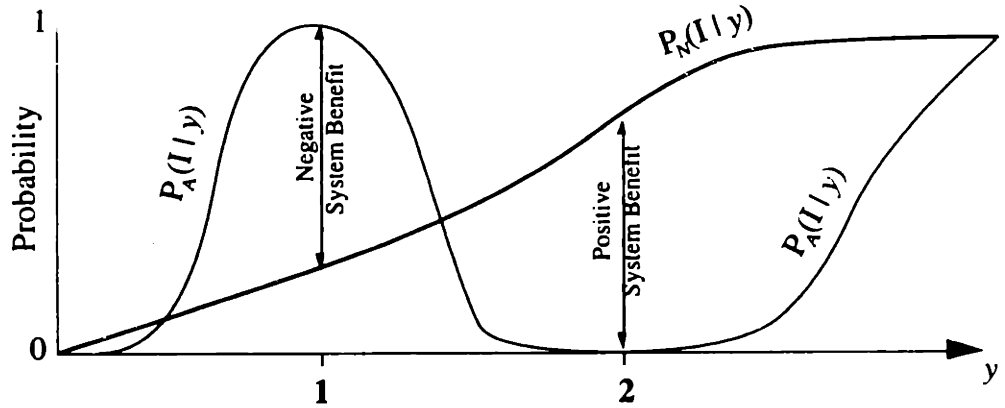


Figure 3.26: Example of Negative Effect of Alerting

As Figure 3.27 shows, an SOC curve for this example also has a different shape from the SOC curves discussed earlier. The region of negative system benefit corresponds to that region on the SOC curve that is below the diagonal line (P(FA) = P(CD)) in the Figure -- that is, those operating points for which the probability of a false alarm is greater

than the probability of a correct detection. At those points on the SOC curve below the diagonal line, it is more advantageous to not alert than to alert.

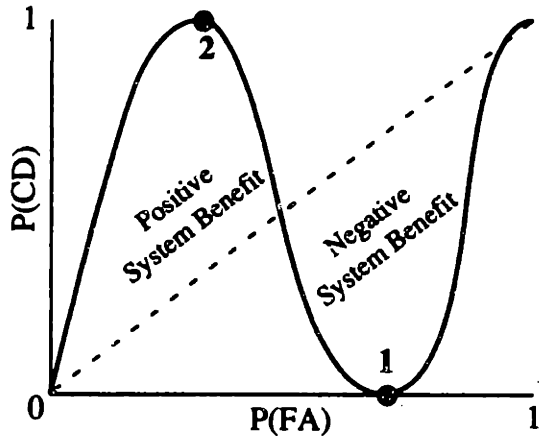


Figure 3.27: SOC Curve: Negative System Benefit Example

3.4 Exposure to Hazards

It is often desirable to consider the alerting system in terms of the overall rates of missed detections or false alarms that can be expected over a range of operating conditions. The example trajectories used in Section 3.3 represented one possible *Hazard Situation.* A Hazard Situation is defined as a set of initial conditions for $\mathbf{y}_a(t)$ and a set of assumptions regarding the maneuvers that occur along the trajectories N and A. For example, the Hazard Situation used in Section 3.3 can be represented by case A in Figure 3.28. Other Hazard Situations (B for example) are possible and may produce different values of P(FA) and P(MD).

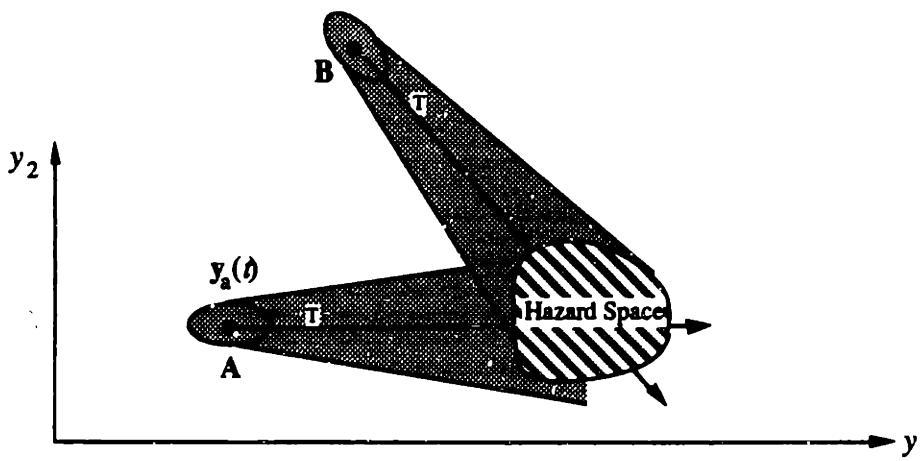


Figure 3.28: Example Hazard Situations

When interested in missed detection or false alarm rates over a wide range of Hazard Situations, it is necessary to consider the probability that each Situation can occur. Consider a set S of Hazard Situations, S. Each Situation in S describes a mutually exclusive set of initial conditions and maneuvering behavior. The probability that a certain Situation occurs is given by the PDF $f_S(S)$. For each Situation, S, the probability of a missed detection, $P_S(MD)$, can be found as was discussed in Section 3.3. The overall rate of missed detections can then be found through Bayes Rule (Brown & Hwang 1992). For example, the expected *Missed Detection Rate* (MDR) is given by:

$$MDR = \int_S P_S(MD)\, f_S(S)\, dS \tag{3.13}$$

MDR represents the overall safety level of the system -- the rate at which incidents will occur with the alerting system in operation. Similarly, the expected *False Alarm Rate* (FAR) is given by:

$$FAR = \int_S P_S(FA)\, f_S(S)\, dS \tag{3.14}$$

The *Correct Detection Rate*, (CDR) can also be defined:

$$CDR = 1 - MDR \tag{3.15}$$

The set of Situations over which Equations (3.13) and (3.14) are integrated depends on the application. For example, it may be of interest to find the FAR for a collision alerting system based on the expected flight paths of aircraft arriving at a particular airport or set of airports. Each flight path is considered to be a Situation. FAR could then give a measure of the expected false alarm rate per flight. Alternatively, FAR could be found as a measure of alerts per year or other unit of time.

In general, $f_S(S)$ may be difficult to estimate. Several methods are possible, however, including the generation of the statistics of S from actual flight data or through a Monte Carlo simulation of the system under study.

Given MDR and FAR as functions of the alerting threshold, SOC curves may also be constructed that show CDR vs. FAR. These SOC curves are similar to those shown in Section 3.3.3, but instead describe rates over a set of Situations rather than probabilities in a single Situation.

4. Example Application: Evaluation of the Ground Proximity Warning System

This chapter provides an example application of the methodology developed in Chapters 2 and 3 to the Ground Proximity Warning System (GPWS) used on civil transport aircraft. Following a brief background on the alerting system, the architecture of GPWS is discussed using the framework developed in Chapter 2. The performance of GPWS is then evaluated in two situations: descent into a flat terrain field, and descent into a probabilistic terrain field representative of actual terrain. These examples illustrate the use of a probabilistic model of the human response to the alert, and show the effect of unobservable states on the performance of the system.

This chapter is focused on only a few of the many operating points over which GPWS is designed. Therefore, it is not intended to represent a comprehensive examination of GPWS, but rather to provide illustrative examples of how the framework can be applied to an actual alerting system.

4.1 Background

The Ground Proximity Warning System (GPWS) provides an alert to civil transport flight crews when the aircraft is in danger of impacting terrain. GPWS was mandated on jet transports in the U.S. following several Controlled Flight Into Terrain (CFIT) accidents, notably the crash of a Trans World Airlines Boeing 727 into hills west of Dulles International Airport on December 1, 1974 (NTSB 1975, FAA 1994: §121.360).[1] Continued CFIT accidents worldwide have motivated renewed efforts to implement GPWS on foreign carriers and smaller domestic aircraft (Nordwall 1993, Hughes 1994).

Since its introduction, GPWS has played a part in reducing the CFIT accident rate in the U.S., as shown in Figure 4.1. However, CFIT accidents still occur, even on aircraft with functioning GPWS equipment, and CFIT remains the leading cause of aircraft fatalities (Figure 4.2a). In the six year period between 1988 and 1993, 1,200 fatalities were attributable to CFIT out of a total of 2,150 fatalities in civil jet transports (56%).

[1] CFIT is defined as an accident in which an aircraft, in controllable flight, collides with terrain.

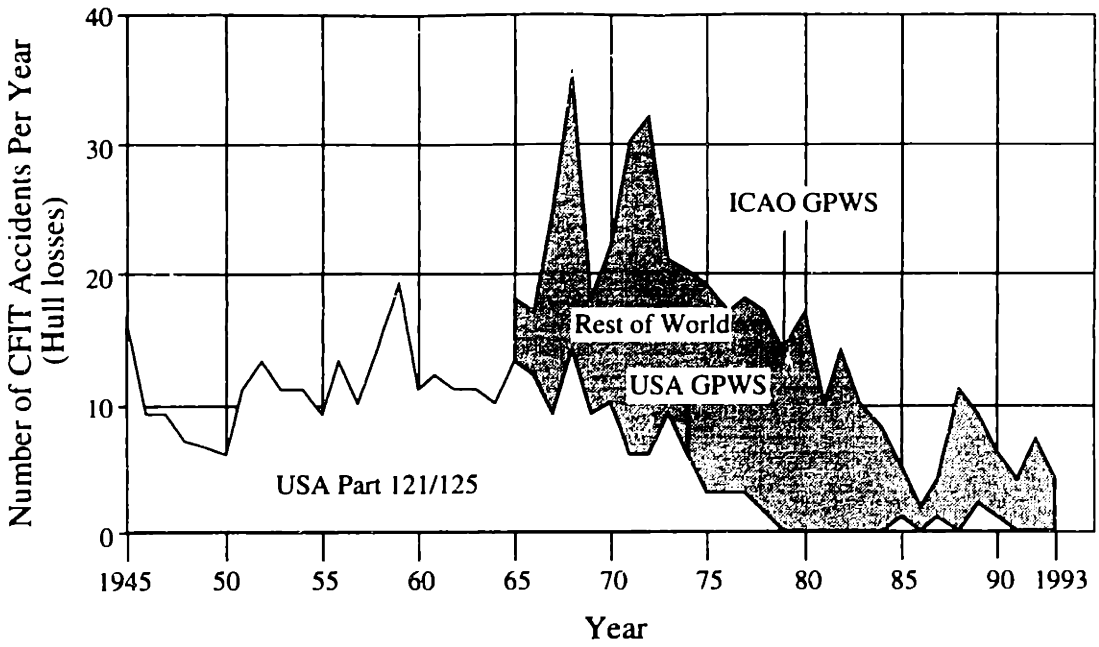
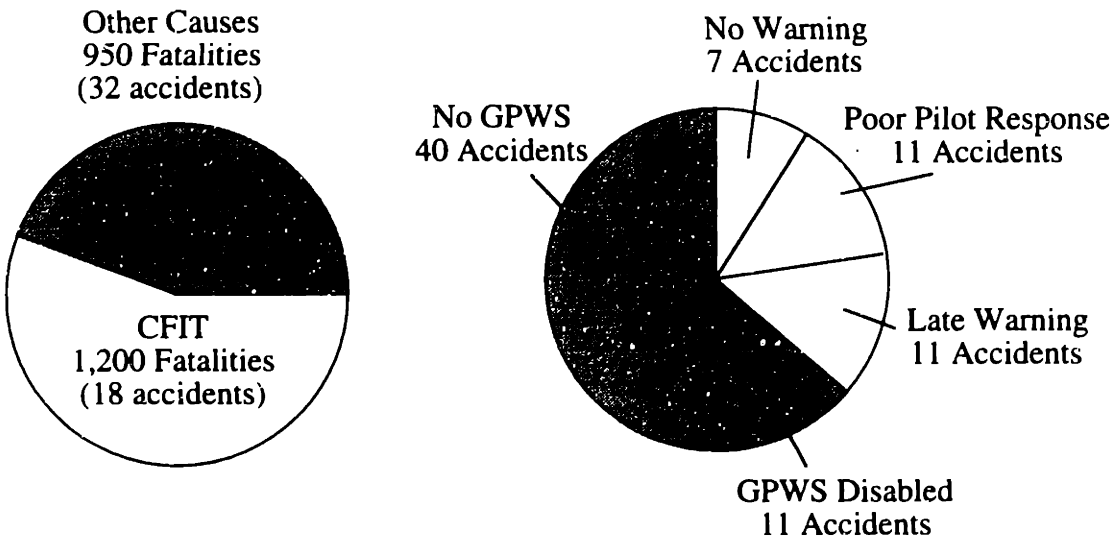


Figure 4.1: Civil Transport CFIT Accidents, 1945-1993
Source: Bateman, 1994a.
ICAO = International Civil Aviation Organization



Fatalities Attributable to CFIT
(1988-1993)
a

GPWS status in CFIT accidents
(1975-1992)
b

Figure 4.2: CFIT Statistics: Jet Transport Aircraft Worldwide
Source: Bateman, 1993.

As shown in Figure 4.2b, the majority of CFIT accidents occurred with aircraft that either did not have a GPWS system or had one that had been disabled. Aircraft with a

functioning GPWS system accounted for approximately 35% of the CFIT accidents between 1975 and 1992. On aircraft that had a working GPWS system, CFIT accidents occurred in one of three categories:

1) No warning of impending impact was given in 24% of CFIT accidents with a functioning GPWS system. These cases generally involve an aircraft descending while in landing configuration. Because false alarms are undesirable during final approach to an airport, GPWS is designed to become less sensitive to terrain hazards when the aircraft is in the landing configuration (i.e , landing gear is down and flaps are extended).

2) Late warnings, in which the pilot had too little time to respond to an alert contributed to 38% of the CFIT accidents with a functioning GPWS system. As is discussed in Section 3.2, GPWS has limited information regarding terrain. To prevent an excessive number of false alarms, the alerting thresholds are set such that alerts may be late if terrain rises extremely rapidly.

3) Poor pilot response was also found to contribute to 38% of the CFIT accidents with a functioning GPWS system. In these cases, GPWS provided adequate time to react to a hazard, but the crew delayed their response or made an inadequate avoidance maneuver. In some cases, the delayed response may be due in part to previously-experienced false alarms with GPWS (DeCelles 1991).

4.2 Structure of Typical GPWS System

GPWS systems on U.S. jet transport aircraft must comply with a set of performance specifications (FAA 1994: §121.360, RTCA 1976). These specifications are designed to protect against a variety of CFIT accident types. The examples in this chapter, however, concern only those situations in which the aircraft is descending toward a terrain field while not in landing configuration (i.e., landing gear is up and flaps are retracted).

GPWS systems currently in place on transport aircraft rely upon both barometric and radar altitude measurements to determine if an alert needs to be issued (Figure 4.3). The barometric altimeter measures altitude above Mean Sea Level (MSL), h_b, and does not by itself provide a measure of altitude above terrain. The radar altimeter provides a measurement of the Above Ground Level (AGL) altitude of the aircraft, h_r. AGL altitude is calculated from the time delay in a radar pulse beamed at the ground directly below the aircraft. The rate of change in MSL altitude or AGL altitude is then extrapolated to estimate

the danger posed by proximity to terrain. No information is available to GPWS regarding the terrain ahead or to the side of the flight path.

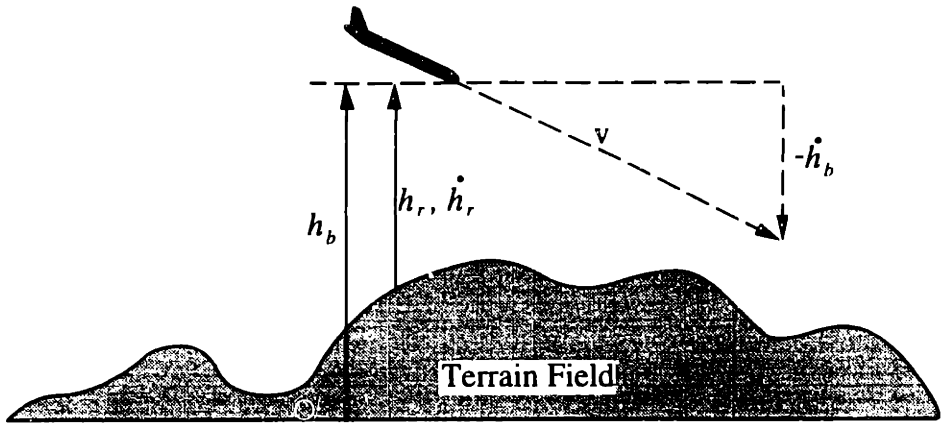


Figure 4.3: Observable States for GPWS

The full state vector, $\mathbf{x}(t)$, for a terrain hazard situation is made of several components, shown in Equation (4.1). The three-dimensional position (x, y, h_b) of the aircraft relative to a reference point on the terrain field determines the proximity of the aircraft to the terrain. The velocity components in each of the three axes ($\dot{x}$, $\dot{y}$, $\dot{h}_b$) describe the rate at which the aircraft approaches the terrain. The AGL height above terrain (h_r) and its rate of change ($\dot{h}_r$) are also included in $\mathbf{x}(t)$ because they provide a direct measure of the proximity to terrain. Other states such as gear and flap position, while important in other terrain situations, are not important in the examples used here and are omitted for simplicity.

$$\mathbf{x}(t) = [x\ \dot{x}\ y\ \dot{y}\ h_b\ \dot{h}_b\ h_r\ \dot{h}_r]^T \tag{4.1}$$

The GPWS system, because it only uses barometric (h_b) and radar altitude (h_r) measurements, cannot observe the lateral states (x, y, $\dot{x}$, $\dot{y}$) in $\mathbf{x}(t)$. Velocity (v) information is available, however, through airspeed measurements. The observable state vector for the alerting system is therefore given by Equation (4.2). These states are also shown schematically in Figure 4.3.

$$\mathbf{y}_a(t) = [h_b\ \dot{h}_b\ h_r\ \dot{h}_r\ \mathrm{v}]^T \tag{4.2}$$

4.2.1 Unobservable States

One drawback of current GPWS systems is that no information is available to the system regarding the terrain ahead or to the side of the flight path. The system must, in

effect, extrapolate the terrain altitude ahead of the aircraft. Two undesirable effects may result due to this limitation, shown in Figures 4.4a and 4.4b. In Figure 4.4a, false alarms may occur when flying over small, sheer terrain rises because the extrapolated terrain slope appears to be hazardous. Missed detections are also possible when flying into rapidly rising terrain if the terrain slope is underestimated (Figure 4.4b). In order to be effective, GPWS alerting logic must attempt to minimize both of these effects.

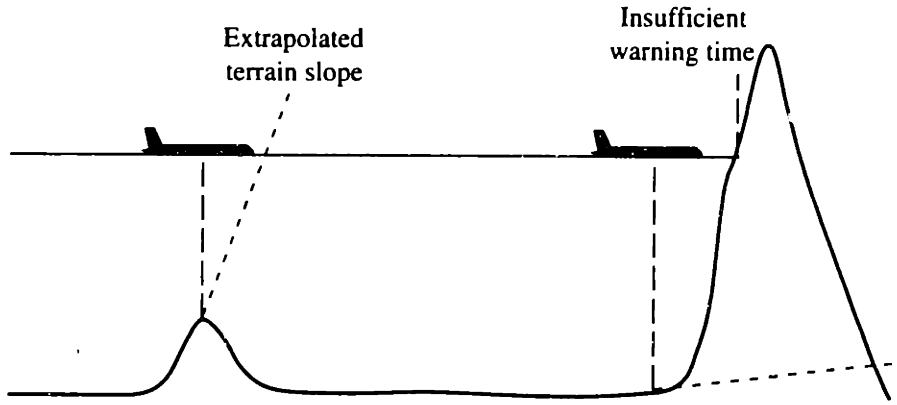
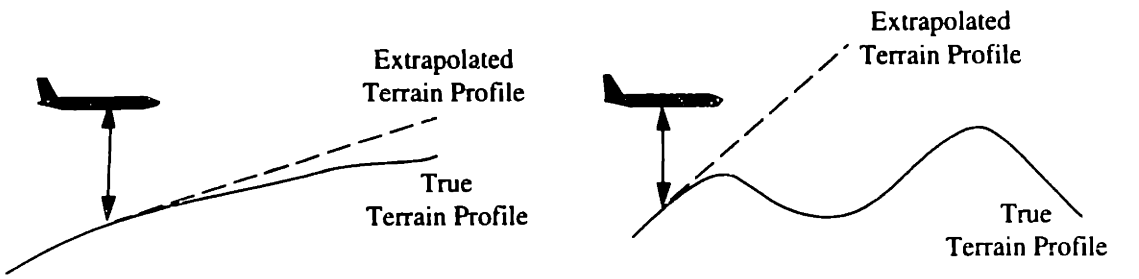


Figure 4.4: Example Erroneous GPWS Alerting Decisions

The information available to GPWS represents only one dimension of the terrain: its altitude directly below the aircraft. Additional dimensions are possible through additional sensors or an on-board terrain database. A two-dimensional system might provide terrain altitude as a function of along-track distance ahead of the aircraft. A full three-dimensional system could use a complete set of terrain altitude as a function of cross- and along-track distance. These concepts are shown in Figure 4.5. Current GPWS, because it is based on limited one-dimensional information, operates with unobserved states and its potential performance is inherently constrained.

Figure 4.5: Potential GPWS Terrain Information Dimensions

The impact of this unobservable-state limitation depends on the type of terrain around the aircraft (Figure 4.6). In regions of slowly varying terrain altitude (Figure 4.6a), the altitude and altitude rate information available to GPWS may be acceptable in providing an accurate extrapolation of future terrain altitude. In areas where terrain altitude changes rapidly (Figure 4.6b), a one-dimensional GPWS system will not be able to extrapolate future terrain altitude with much accuracy.



Slowly Varying Terrain
Accurate Terrain Extrapolation
a

Rapidly Varying Terrain
Inaccurate Terrain Extrapolation
b

Figure 4.6: Effect of Frequency of Terrain Altitude Variation on the Accuracy of Predicted Terrain Profile

4.2.2 Alerting Thresholds

The Minimum Performance Specifications for GPWS provide alerting thresholds for the system based on the observable states in $\mathbf{y}_a(t)$ (RTCA 1976). The minimum specifications only require a one-stage alerting method. In practice, however, GPWS systems use a two-stage system of Caution (aural alerts of "Sink Rate" or "Terrain") and Warning (aural alert of "Pull Up"). GPWS uses Current State Extrapolation and does not use intended path information when extrapolating the state trajectory.

Figure 4.7 shows an example alerting threshold diagram from the Boeing 767 "Excessive Barometric Descent Rate" category (Boeing 1983). For example, an aircraft that is descending at 4,000 ft/min generates a "Sink Rate" Caution at an altitude of approximately 2,000 ft above the terrain, and a "Pull Up" Warning at an altitude of approximately 1,200 ft above the terrain. If the aircraft is not in landing configuration, a cautionary "Too Low Gear" alert is issued if the measured AGL altitude is below 500 ft. Note that although barometric altitude, h_b, is available to the system, MSL altitude by itself provides no insight into the danger posed to the aircraft and is not used in the alerting thresholds. Barometric altitude rate, however, is used by GPWS.

Figure 4.8 shows the alerting threshold diagram for the same aircraft for alerts triggered due to a rapid decrease in radar altitude (AGL), termed "Excessive Terrain Closure Rate".

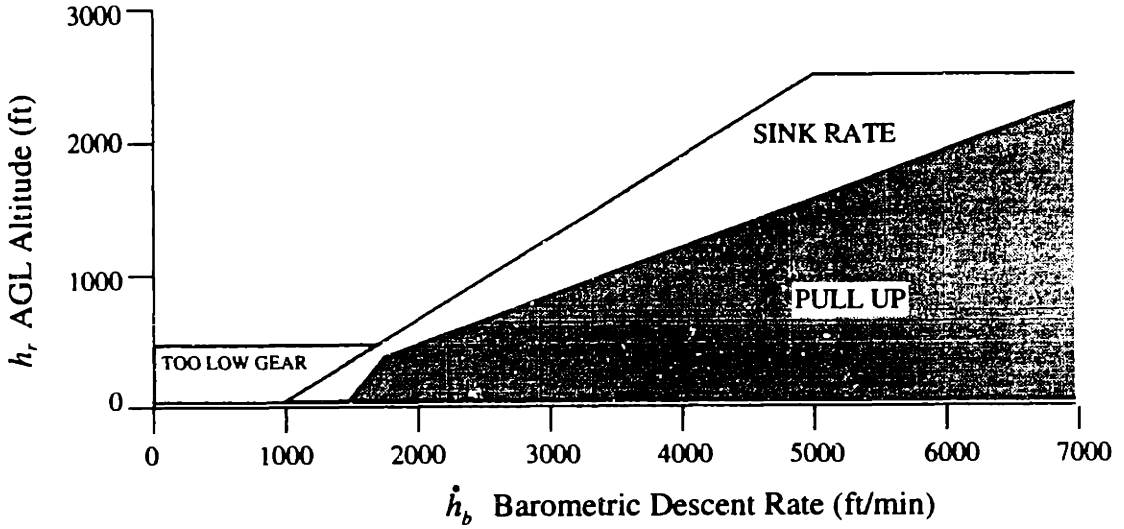


Figure 4.7: GPWS Alerting Threshold: Excessive Descent Rate
Aircraft not in landing configuration.

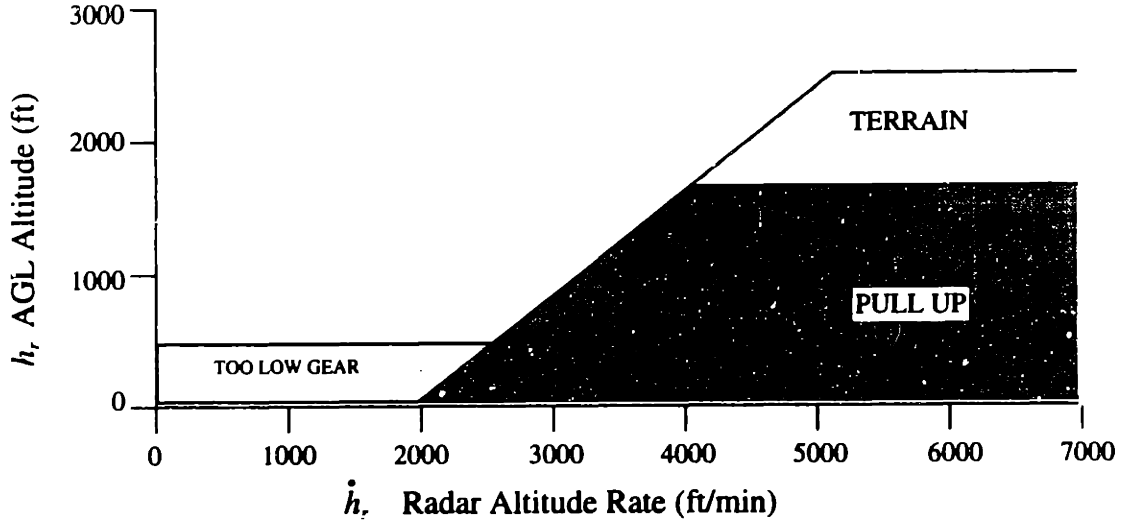


Figure 4.8: GPWS Alerting Threshold: Excessive Terrain Closure Rate
Aircraft not in landing configuration.

4.2.3 Alerting Displays

GPWS uses an aural alarm and simple visual displays to attract the attention of the flight crew. Cautions provide some indication of the cause for the situation and its

severity. For example, the Caution "Sink Rate" tells the pilot that the descent rate is too high. "Terrain" suggests that rising terrain may pose a hazard. "Pull Up" is a Warning that indicates that immediate, positive action is required, regardless of the cause.

The visual displays for GPWS augment the aural alerts. An amber "GND PROX" light illuminates for Cautions. A red "PULL UP" light illuminates for "Pull Up" Warnings. Master Caution or Warning lights in the cockpit also illuminate when a GPWS alert is triggered. However, no other supporting information is available except for the barometric and radar altimeters and the vertical speed indicator. The lack of supporting information is thought to contribute to delayed response to GPWS alerts (Kuchar & Hansman 1993).

4.2.4 Crew Response

Procedures call for an immediate reaction to a "Pull Up" warning if visual contact with terrain has not been made (DeCelles 1991). The recommended course of action is to apply maximum thrust and pitch up at 2°-3° per second until a 15°-20° pitch attitude is attained (Poole 1992, Lewis 1994). The pilot should then hold this attitude until the alert ends.

Because the alerting displays are such that little information is available to the pilot regarding the nature of the terrain hazard, pilots are instructed that they should not attempt to turn away from terrain when they receive a GPWS alert. Rather, procedures stipulate that a wings-level pull-up maneuver should be used.

Figure 4.9 shows a State-Space Hazard Diagram for GPWS for the situation in which the aircraft is descending into flat terrain. Hazard Space is defined by the terrain itself: any $h < 0$ results in an incident. A reasonable upper limit on the flight path angle rate of a jet transport aircraft is approximately 5°/sec. Accordingly, Maneuver Limit Space in this example is defined by a wings-level pull-up maneuver at a constant 5°/sec flight path angle rate. For reference, a 2°/sec pull-up trajectory is also shown in the Figure. Alert Space is shown for the same alerting thresholds discussed previously.

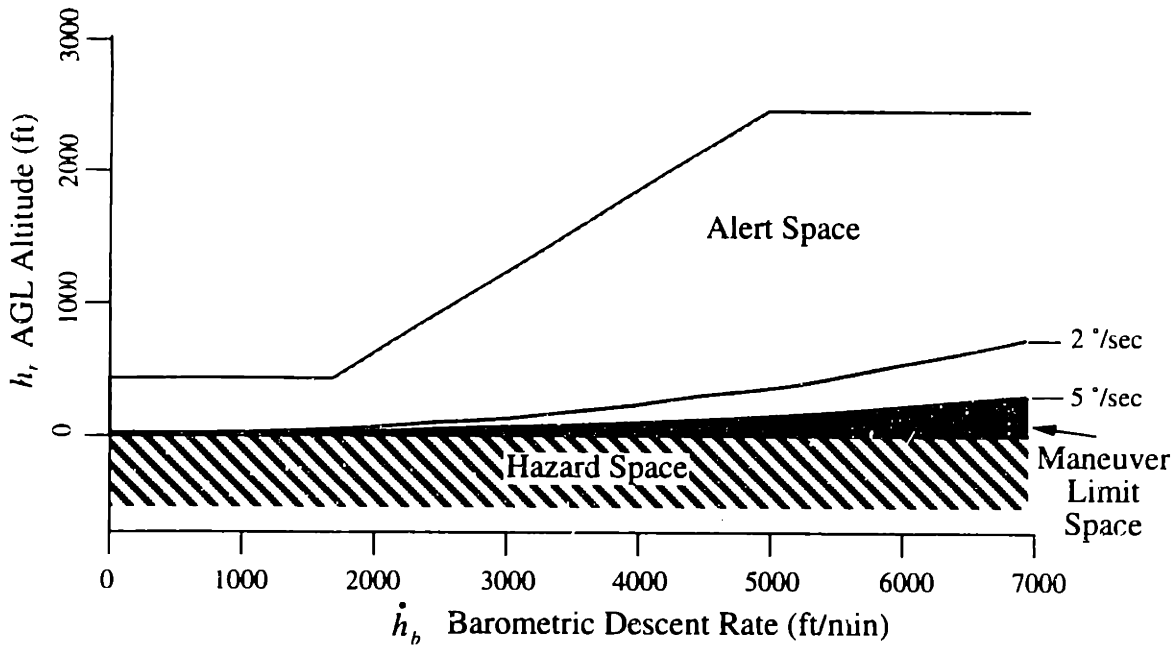


Figure 4.9: GPWS State-Space Hazard Diagram
Descent into flat terrain.

In the Figure, the GPWS alerting system design appears to be quite conservative: Alert Space is much larger than Maneuver Limit Space. However, it must be remembered that when an alert is issued, there will be some latency in the response and the avoidance maneuver may be such that the pitch rate is less than 5°/sec. In addition, the terrain may rise while the aircraft is arresting its descent, increasing the effective descent rate. A more detailed examination of the GPWS alerting thresholds is covered in the following sections.

4.3 Example GPWS Performance Evaluation: Flat Terrain with a Probabilistic Avoidance Maneuver

4.3.1 Equations of Dynamics

As a simplified starting point, the performance of GPWS when descending into a flat terrain field is examined. The basic problem statement at this point is shown in Figure 4.10. Measurements of AGL altitude (h_0), altitude rate ($\dot{h}_0$), and velocity (v) are used by the alerting criteria to determine if an alert should be issued.

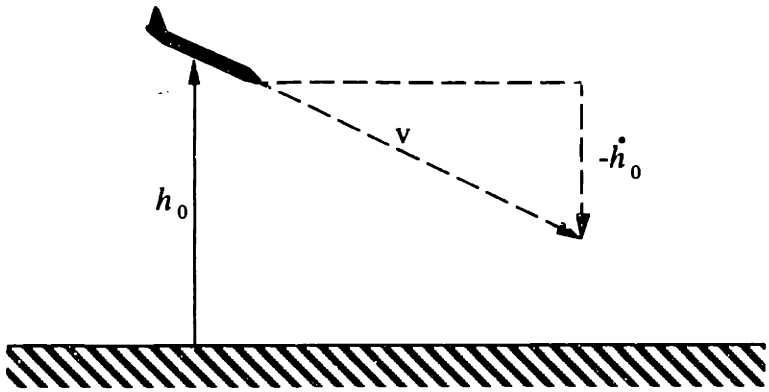


Figure 4.10: Descent Into Flat Terrain Field

The aircraft is located a known distance h_0 above the terrain, with a known vertical rate $\dot{h}_0$ (< 0). In response to an alert, there is a response time delay (τ_R), followed by a pull-up maneuver performed at a flight path angle rate of $\dot{\gamma}$ radians per second (assumed constant throughout the pull up maneuver). Airspeed, v, is assumed to be 170 knots throughout this chapter and is considered to be constant throughout the maneuver. This airspeed is typical of situations of maneuvering for final approach where terrain hazards may be present.

Figure 4.11 shows an example Avoidance Trajectory, A. During the maneuver, the aircraft loses an altitude of Δh and attains a minimum altitude of h_{min}. If $h_{min} \leq 0$, a terrain impact incident would occur.

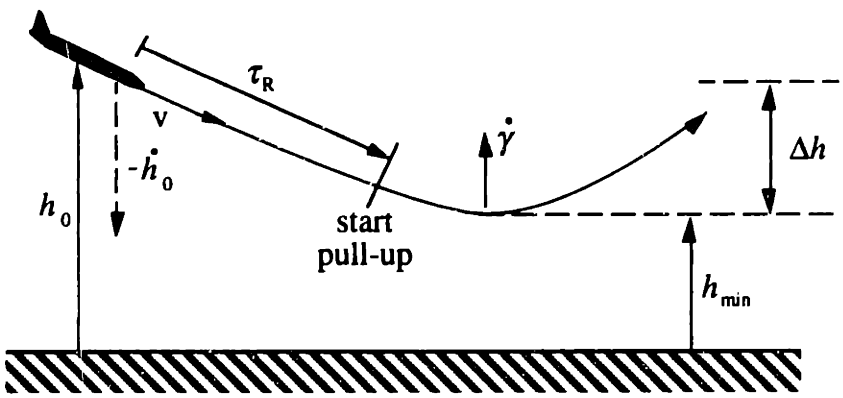


Figure 4.11: Avoidance Maneuver Dynamics

As derived in Appendix C, the minimum altitude of the trajectory is given by:

$$h_{min} = h_0 + \dot{h}_0 \tau_R - \frac{v}{\dot{\gamma}} + \frac{\sqrt{v^2 - \dot{h}_0^2}}{\dot{\gamma}} \tag{4.3}$$

Also, the load factor, n, in g's, experienced during the pull-up is given by:

$$n = \frac{v\dot{\gamma}}{g} + 1 \tag{4.4}$$

The altitude lost during the maneuver, Δh, is shown as a function of response time and descent rate in Figure 4.12, and as a function of flight path angle rate and descent rate in Figure 4.13.

$$\Delta h = h_0 - h_{min} \tag{4.5}$$

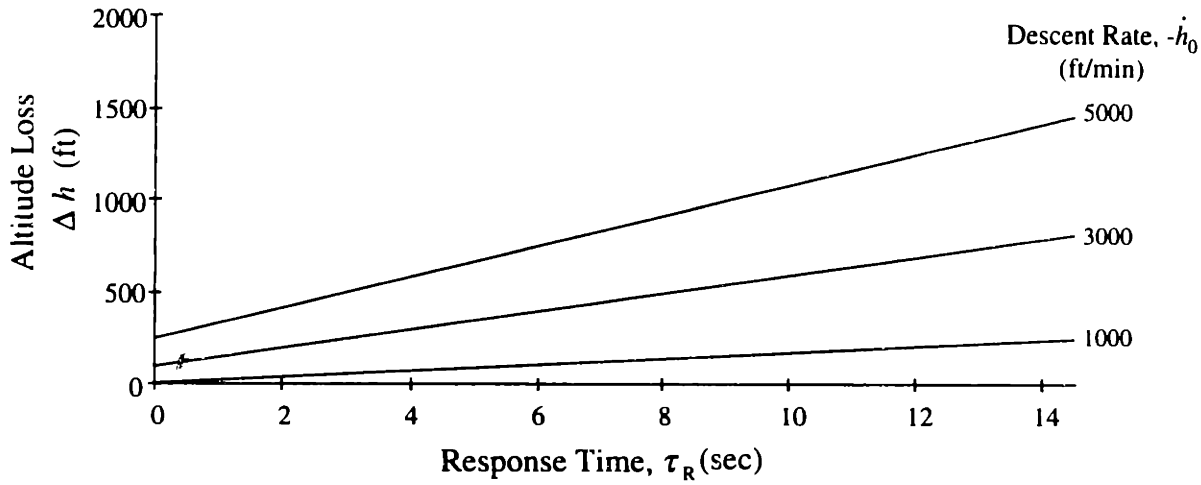


Figure 4.12: **Δh as a Function of Response Time**
170 kts, 3°/sec flight path angle rate.

An increase in the response time delay produces a linear increase in the altitude required to complete the maneuver. Figure 4.13 illustrates the diminishing returns of increased flight path angle rate, $\dot{\gamma}$. There is little additional benefit to performing the pull-up maneuver at a flight path angle rate greater than 3°/sec because at high flight path angle rates the altitude lost during the pull-up maneuver is small compared to the altitude lost due to the time delay. The choice of a standard 2°/sec or 3°/sec pull-up maneuver therefore appears well founded, providing a compromise between load factor and altitude loss.

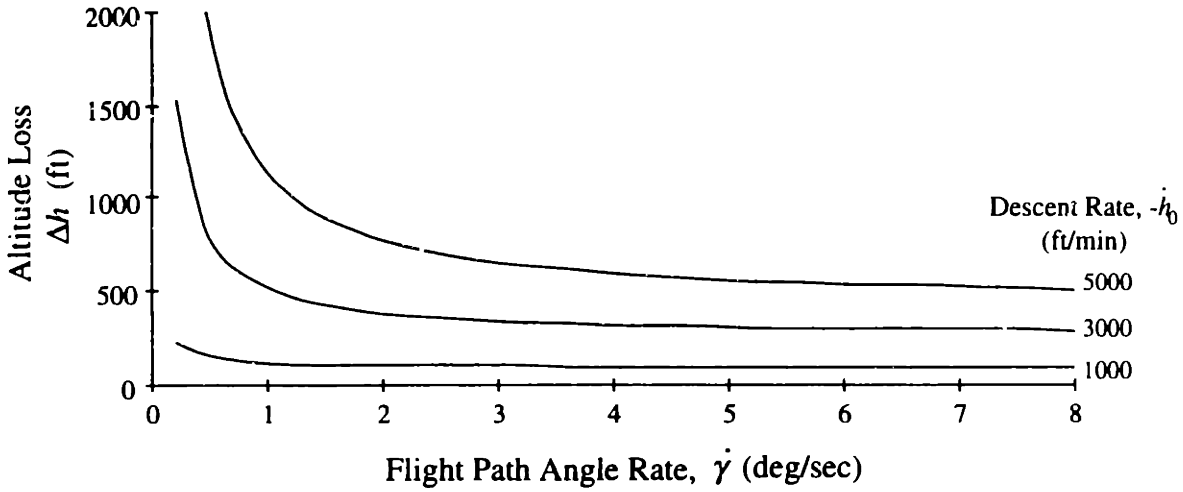


Figure 4.13: **Δh as a Function of Flight Path Angle Rate**
170 kts, 5 second response time delay.

4.3.2 Probability Density Functions

When an alert is issued, the methodology treats the situation as a combination of deterministic and random variables. In particular, response time and flight path angle rate are treated as random variables, and altitude, descent rate, and velocity are treated as known, deterministic values. For this example, representative time delay and flight path angle rate probability density functions (PDFs) $f_\tau(\tau)$ and $f_{\dot\gamma}(\dot\gamma)$ are needed.

Gamma distributions were chosen to model the PDFs (shown in Figures 4.14 and 4.15) because they provide a smooth, skewed distribution similar to that which might be expected (Hogg & Tanis 1988). These PDFs also provide a non-zero probability of very long response times representative of a flight crew that disregarded or misunderstood an alert. A more detailed study of GPWS could use statistics from actual events or flight simulations to build PDFs of response time delay and flight path angle rate.

The value of a gamma distribution is given by:

$$f_x(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta} \tag{4.6}$$

$$\text{where } \Gamma(\alpha) = (\alpha - 1)!$$

The mean, $\bar{x}$, of the gamma distribution is given by:

$$\bar{x} = \alpha\theta \tag{4.7}$$

Three separate PDFs were chosen for response time and four were chosen for pitch rate to provide insight into the sensitivity of GPWS to changes in the aggressiveness of an avoidance maneuver. The parameters α and θ in the PDFs were chosen to provide mean response times and pitch rates that are comparable to a maneuver of a 5 second delay and 2°/sec pull up maneuver. A 5 second response delay is typical of hazard alerting systems on aircraft (Poole 1992, RTCA 1976). As shown in Figure 4.14, the PDFs for response time vary in mean from 3 to 7 seconds. The PDFs in Figure 4.15 have a range in mean flight path angle rate between 1°/sec and 4°/sec.

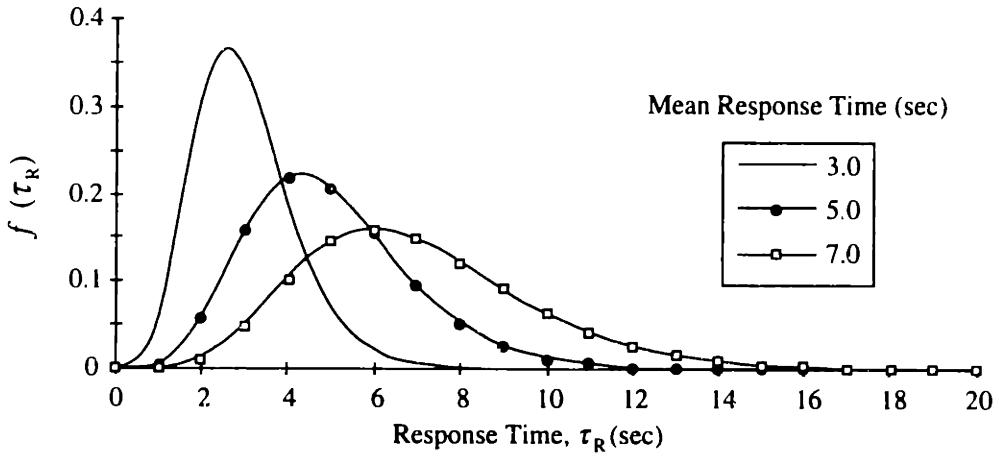


Figure 4.14: Response Time PDFs Used in Flat Terrain Example[2]

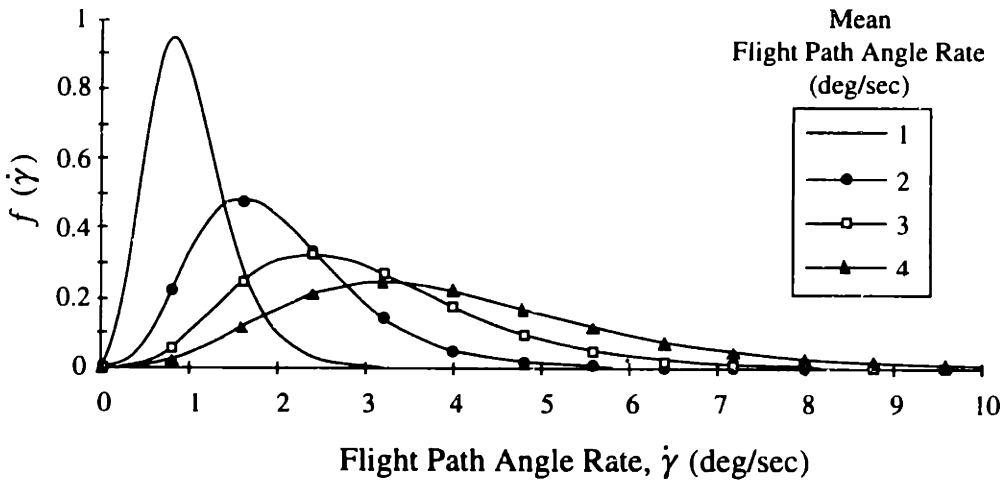


Figure 4.15: Flight Path Angle Rate PDFs Used in Flat Terrain Example[3]

[2] Gamma distribution $\alpha = 7$; $\theta = 0.43, 0.71, 1.0$.
[3] Gamma distribution $\alpha = 5$; $\theta = 0.2, 0.4, 0.6, 0.8$.

4.3.3 Calculation of Probability of an Incident

From Chapter 3, the probability that an incident will occur along the Avoidance Trajectory, $P_A(I \mid y)$, is a function of the observable state measurement and the extrapolated state trajectory. For simplicity, this probability is written $P_A(I)$ in the following discussion. From Appendix A, the general solution for the probability of an incident evaluated over the probabilistic Avoidance Trajectory A is:

$$P_A(I) = \int_{\mathbf{T}} \int_{\mathbf{S}} \int_{\mathbf{Z}(\sigma,\tau)} f_{xyz}(\mathbf{x} - \mathbf{x}_m)\, f_\sigma(\sigma)\, f_\tau(\tau)\, d\mathbf{x}\, d\sigma\, d\tau \tag{4.8}$$

where $\tau \in \mathbf{T}$ is a possible trajectory flown by the aircraft, $\sigma \in \mathbf{S}$ is a possible hazard characteristic (size and shape), and $\mathbf{Z}$ is the region in space over which the integration of spatial location uncertainty is performed (see Appendix A for a complete description).

Given a current measurement of altitude and descent rate, the potential trajectories that the aircraft could follow can be computed. The terrain field in this example is assumed to be flat and is therefore fully determined. Thus, the size and shape of the hazard is deterministic and no integral over $\mathbf{S}$ is needed. Also, because the aircraft's initial position relative to the terrain is known, the probability of an encounter, given a particular trajectory, is either 0 or 1 as defined by the function $F\left(h_{\min}(v, h_0, \dot{h}_0, \dot{\gamma}, \tau_R)\right)$ in Equation (4.9). The integral over $\mathbf{Z}$ is reduced to the following:

$$\int_{\mathbf{Z}(\sigma,\tau)} f_{xyz}(\mathbf{x} - \mathbf{x}_m)\, d\mathbf{x} \equiv F\left(h_{\min}(v, h_0, \dot{h}_0, \dot{\gamma}, \tau_R)\right) = \begin{Bmatrix} 0, & \text{if } h_{\min} > 0 \\ 1, & \text{if } h_{\min} \le 0 \end{Bmatrix} \tag{4.9}$$

The probabilistic nature of the trajectory A is defined by the time delay and the flight path angle rate. The integral over $\mathbf{T}$ in Equation (4.8) is therefore a double integral over τ_R, and $\dot{\gamma}$. The resultant equation (for a given h_0 and $\dot{h}_0$) for $P_A(I)$ is then:

$$P_A(I) = \int_0^\infty \int_0^\infty f_\tau(\tau_R)\, f_{\dot{\gamma}}(\dot{\gamma})\, F\left(h_{\min}(v, h_0, \dot{h}_0, \dot{\gamma}, \tau_R)\right) d\dot{\gamma}\, d\tau_R \tag{4.10}$$

The integral in Equation (4.10) was solved using numerical integration for the different PDFs shown in Figures 4.14 and 4.15. The velocity, v, used in the example is 170 knots and the descent rate is 4,000 ft/min. The results are shown in Figures 4.16 through 4.18.

Figure 4.16 shows $P_A(I)$ as a function of measured AGL altitude (h_0) and mean response time from the gamma distributions in Figure 4.14. All curves in the Figure are shown for the flight path angle rate PDF that has a mean of 3°/sec. It can be seen that longer response times result in a more gradually-varying $P_A(I)$ curve. As discussed in Chapter 3, it is therefore expected that the system's performance will be lower (in terms of false alarms and missed detections) as the pilot's mean response time increases.

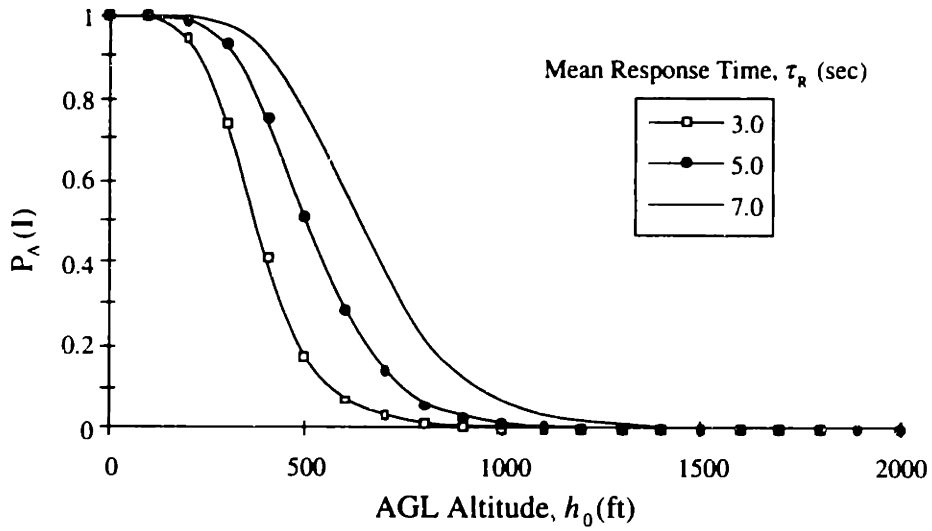


Figure 4.16: **Flat Terrain Example: $P_A(I)$ vs. h_0 and τ_R**
Flight path angle rate PDF mean: $\dot{\gamma}$ = 3°/sec
170 kts, 4,000 ft/min descent

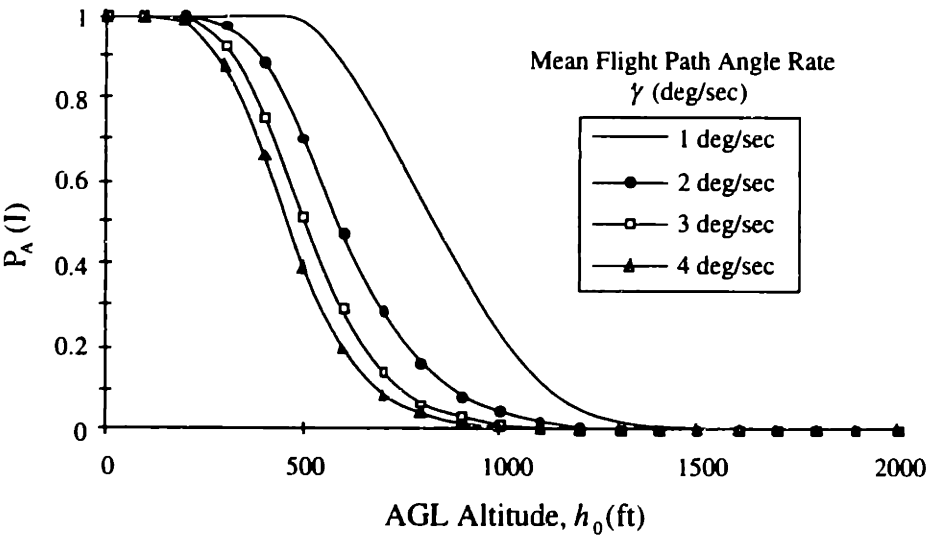


Figure 4.17: **Flat Terrain Example: $P_A(I)$ vs. h_0 and $\dot{\gamma}$**
Response time PDF mean: τ_R = 5 sec
170 kts, 4,000 ft/min descent

In Figure 4.17, $P_A(I)$ curves are shown as a function of AGL altitude and mean pitch rate, for the response time PDF that has a mean of 5 seconds. Again, the diminishing benefit of larger pitch rates can be seen above approximately 3°/sec.

Contours of constant $P_A(I)$ are shown in Figure 4.18 as a function of AGL altitude and descent rate at 170 knots for a 5 second mean time delay and a mean flight path angle rate of 3°/sec. For comparison, the GPWS excessive descent rate alerting thresholds are also shown. With the assumed PDFs for response time and flight path angle rate, a GPWS "Sink Rate" Caution provides a probability of an incident of approximately $1x10^{-7}$ after the alert is issued when descending into flat terrain. If the crew delays their response until a "Pull Up" Warning is received, the probability that an incident will occur increases to approximately $1x10^{-3}$. It should be repeated that these results are based on the assumed PDFs for response time and pitch rate, which may not be representative of actual pilot behavior.

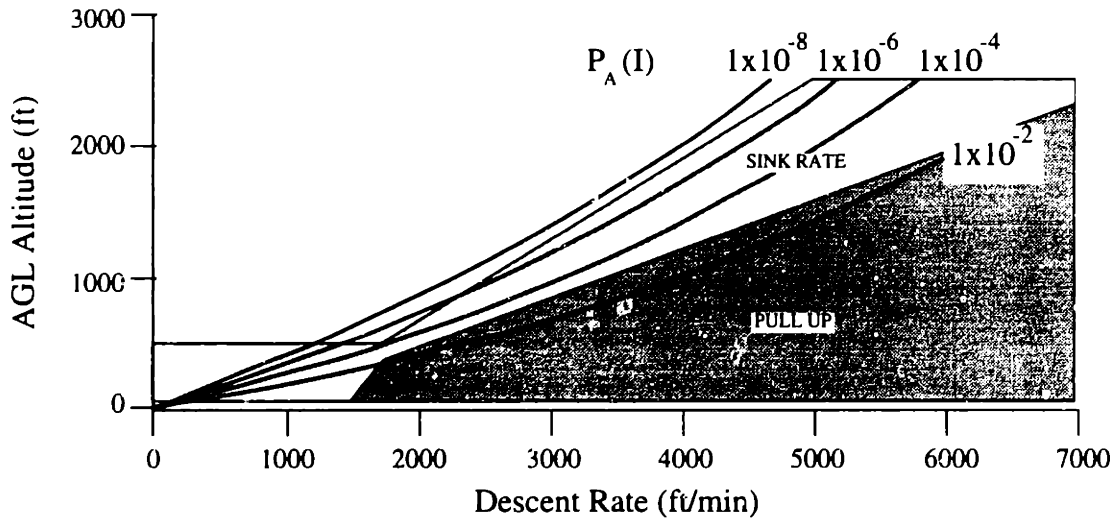


Figure 4.18: Flat Terrain Example: Contours of $P_A(I)$ vs. h_0 and $\dot{h}_0$
170 knots, not in landing configuration.
Probabilistic reaction to an alert is assumed with mean response latency of 5 sec and mean flight path angle rate of 3°/sec.

4.4 Example GPWS Performance Evaluation: Probabilistic Terrain Field

Section 4.3 provided an example evaluation of GPWS for the relatively simple situation of descent into a known, flat terrain field. As discussed earlier, however, one of the major limitations of GPWS is that it may not accurately predict the altitude of the terrain

in front of the aircraft. To probabilistically examine the effectiveness of GPWS in different types of terrain first requires a statistical model of the variations in terrain altitude.

Returning to Equation (4.8), it is no longer possible to ignore the potential size or shape of the hazard given by the set **S**. Instead, some means of calculating the probability that the terrain takes on certain shapes is required. The probability that an incident will occur then depends on the probability that a terrain profile will exist that conflicts with the future trajectory of the aircraft (Figure 4.19).

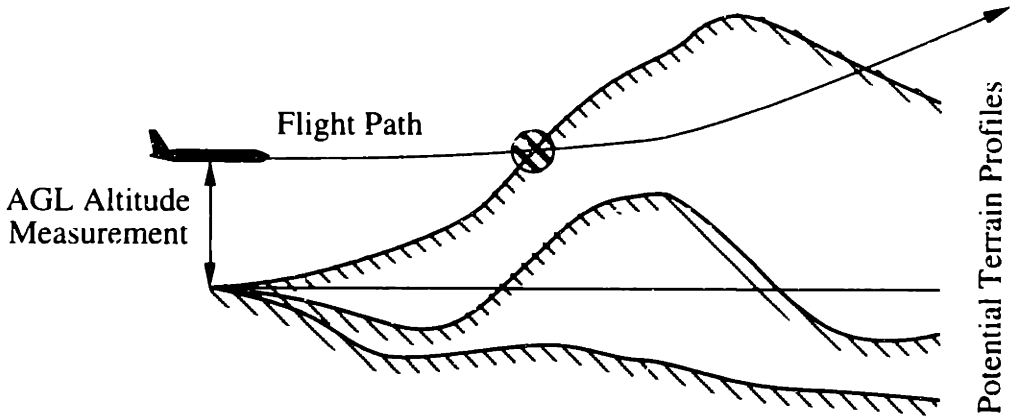


Figure 4.19: Uncertainty in Terrain Altitude Along Flight Path

Appendix C details the procedure that was used to calculate the probability that the terrain is a certain height at a given distance in front of the aircraft. In summary, a first-order Gauss-Markov model was developed as part of the thesis that described the statistics of terrain. A Markov Chain was then used to calculate the probability of terrain altitude as a function of distance ahead of the aircraft.

Five categories of terrain were defined, ranging from slowly-varying smooth terrain to rapidly-varying steep terrain based on the parameters presented in Table 4.1. For example, Smooth terrain was defined as terrain with a 99% probability that the terrain altitude does not deviate more than 500m (1,640 ft) over a 60 nmi distance. The probabilistic model of the terrain was developed empirically using a terrain database covering the Great Plains and Rocky Mountain region of the U.S. (RMC 1991). To simplify the example, only the Smooth, Moderate, and Steep terrain categories are considered here.

Table 4.1
Terrain Categories

Terrain Category	Altitude Variance (m^2)	99% Confidence Interval Altitude Span over 60 nmi
Smooth	< 7000	< 500 m
Moderately Smooth	7,000 - 27,778	500 - 1000 m
Moderate	27,778 - 62,500	1000 - 1500 m
Moderately Steep	62,500 - 111,111	1500 - 2000 m
Steep	> 111,111	> 2000 m

4.4.1 Level Flight

First, the probability that the aircraft will impact terrain was calculated for situations in which the aircraft flies at a constant altitude. Given an initial AGL altitude measurement, the probability that the aircraft will be involved in an incident as a function of distance flown can be calculated. An example is presented in Figure 4.20 which shows curves of the cumulative probability of an incident as a function of distance traveled for the three terrain categories when h_0 is 1,000 ft. For example, when the aircraft is flying level with a current radar altimeter measurement that indicates that h_0 = 1,000 ft above Moderate terrain, the probability that an incident will occur before the aircraft flies 3 nmi is approximately 0.01.

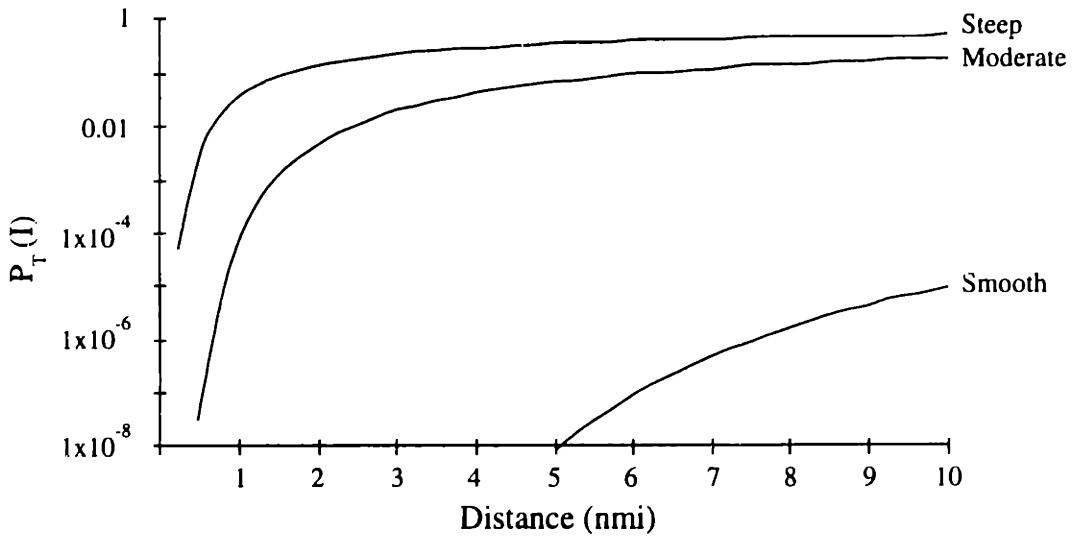


Figure 4.20: **$P_T(I)$ vs. Distance Flown**
Level flight trajectory initially at 1,000 ft AGL
Smooth, Moderate, and Steep Terrain.

Note that Figure 4.20 shows the cumulative probability of an incident based on a single measurement. As the aircraft flies, however, it will obtain new altitude measurements that may modify the risk to the aircraft.

4.4.2 Performance During Descent

Next, the performance of GPWS in terms of P(FA) and P(MD) was evaluated for a range of descent rates. Following the methodology presented in Chapter 3, Nominal (N) and Avoidance (A) Trajectories need to be defined in order to estimate P(FA) and P(MD). It is assumed that if an alert is issued, the Avoidance Trajectory is followed, and if an alert is not issued, the Nominal Trajectory is followed. A probabilistic Nominal Trajectory for this situation can be difficult to estimate because it is unclear what the intentions of the pilot are. To provide some measure of the need for an alert, however, the Nominal Trajectory is assumed to be a 3-nmi-long straight-line extrapolation of the aircraft's flight path. Under this definition, if terrain will not be encountered within 3 nmi at the current descent rate, then an alert is not needed. The 3 nmi length of N was chosen because it is representative of the length scale over which a typical terrain proximity incident would occur. This distance also corresponds to the maximum advance warning time typically used in GPWS systems (approximately 60 seconds at a speed of 170 knots) (Bateman 1994b).

The Avoidance trajectory assumed a standard maneuver of a 5 second delay followed by a 2°/sec pull up to a 15° flight path angle. This maneuver corresponds to typical response times and the suggested flight path angle rate when a GPWS alert is issued (Poole 1992). Figure 4.21 shows the Nominal (N) and Avoidance (A) trajectories.

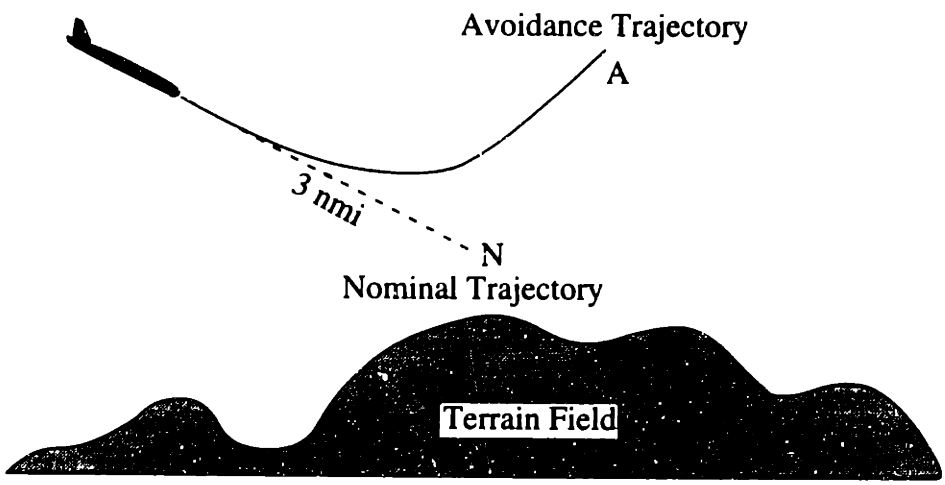


Figure 4.21: Definition of Nominal and Avoidance Trajectories

4.4.2.1 Descent Rate Example: 3,000 ft/min

The probabilities of an incident along the Nominal and Avoidance Trajectories, $P_N(I)$ and $P_A(I)$, were evaluated over a range of conditions that specify AGL altitude, descent rate, and velocity using the Markov Chain model described in Appendix C. One example evaluation is shown in Figure 4.22 for an aircraft at 170 knots with a descent rate of 3,000 ft/min. To see the effect of different types of terrain, curves of $P_N(I)$ and $P_A(I)$ are shown for Smooth and Steep terrain categories. The GPWS excessive descent rate alerting threshold (located at approximately 1,275 ft) is also shown.

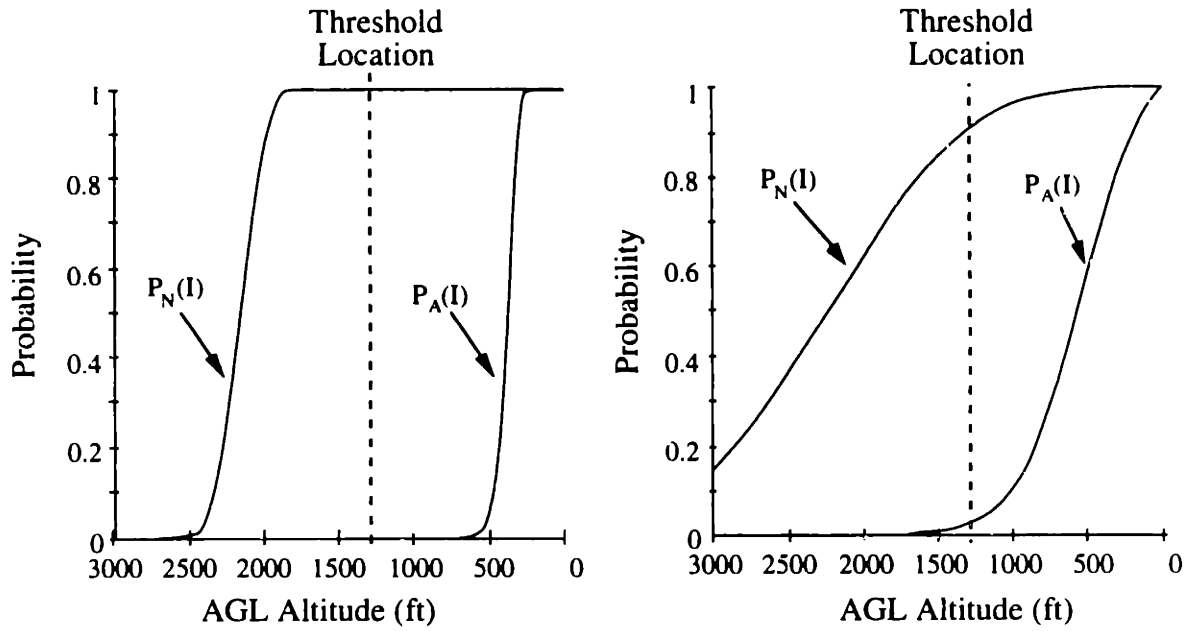


Figure 4.22: $P_N(I)$ and $P_A(I)$: 3,000 ft/min Descent Smooth and Steep Terrain Categories
170 knots.

As can be seen in Figure 4.22, the curves of $P_N(I)$ and $P_A(I)$ for Smooth terrain rise more rapidly than the corresponding curves for Steep terrain. It is expected, therefore, that system performance will be higher in Smooth terrain than in Steep terrain.

Figure 4.23 shows the resulting SOC curves, again for the Smooth and Steep terrain cases. The location of the GPWS alerting threshold for each case is also shown by a circle on the SOC curve. The Figure illustrates that GPWS is able to operate at a higher P(CD) and a lower P(FA) in Smooth terrain than in Steep terrain.

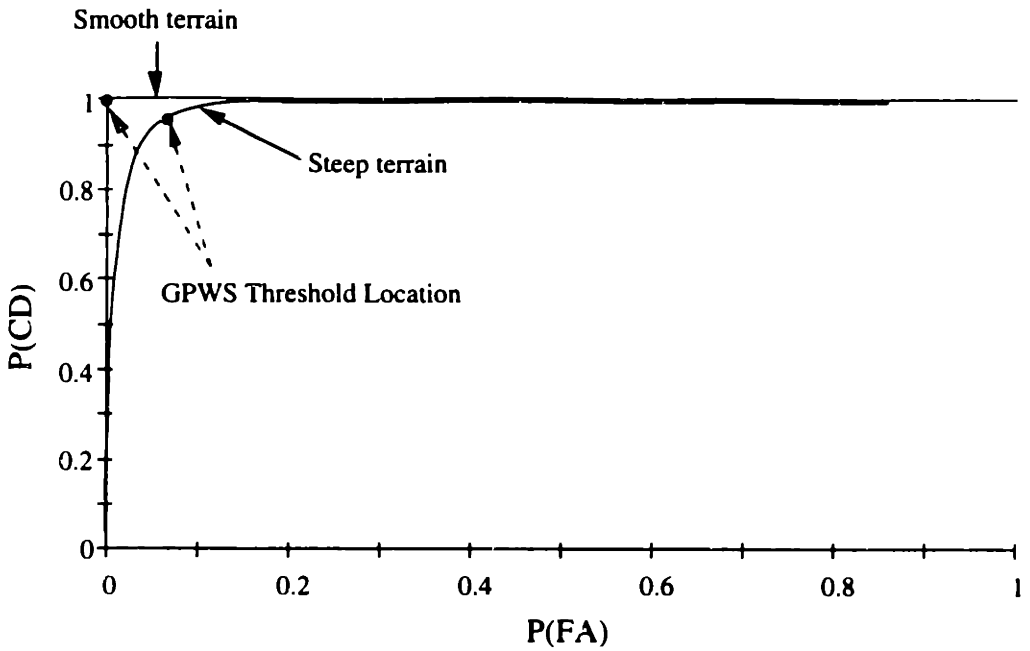


Figure 4.23: SOC Curve: 3,000 ft/min Descent
Smooth and Steep Terrain Categories
170 knots

As summarized in Table 4.2, the probability that an alert is a false alarm increases from $1x10^{-4}$ to 0.08 between the Smooth and Steep terrain categories. The probability that a missed detection occurs increases from $<1x10^{-16}$ to 0.03 between the same terrain categories. Thus, there is a decrease in performance both in terms of false alarms and missed detections in Steep terrain.

Table 4.2
Summary of GPWS Performance at 3,000 ft/min

Terrain Category	P(FA)	P(MD)
Smooth	$1x10^{-4}$	$<1x10^{-16}$
Steep	0.08	0.03

4.4.2.2 Evaluation at all Descent Rates

Next, it is of interest to view how current GPWS thresholds compare to the probability of an incident over a range of AGL altitudes and descent rates. Contours of constant $P_A(I)$ are shown in Figure 4.24 as a function of altitude and descent rate assuming the same standard avoidance maneuver used earlier (5 sec delay, 2°/sec pull up). Also shown in the Figure are the GPWS alerting thresholds.

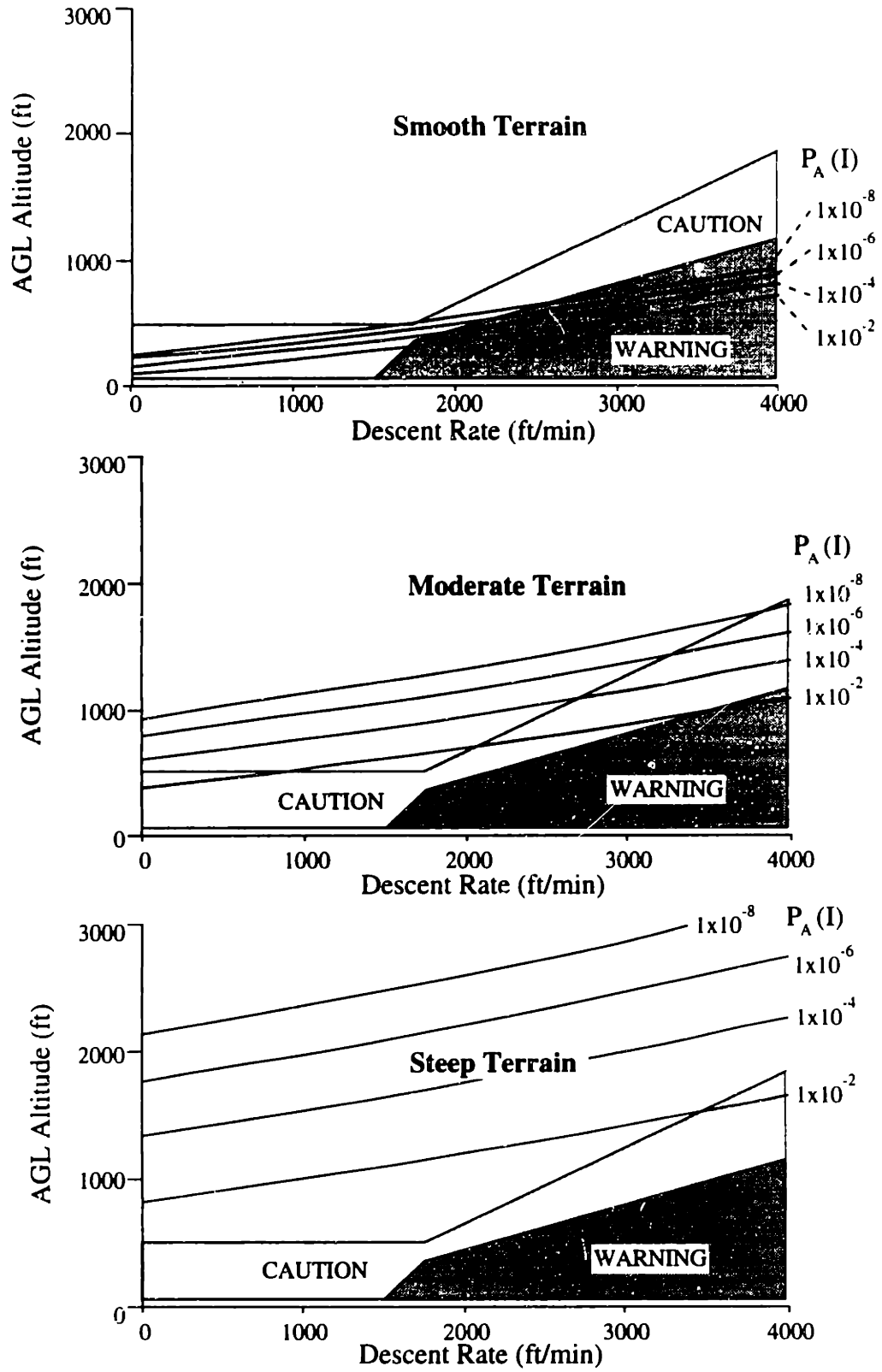


Figure 4.24: Comparison of $P_A(I)$ and GPWS Thresholds
170 knots, 5 sec response latency, 2°/sec flight path angle rate in response to alert. Not in landing configuration. Uses probabilistic terrain field developed in Appendix C.

From Figure 4.24, it can be seen that the GPWS thresholds provide for a very low ($< 1x10^{-8}$) probability of an incident for Smooth terrain when an alert is issued. In the Moderate and Steep terrain categories, however, GPWS does not provide as high a degree of protection. If Alert Space were expanded to compensate, an increased rate of false alarms would likely occur in Smooth terrain situations. The decreased performance of GPWS in Steep terrain is a symptom of the fact that the terrain ahead of the aircraft is unobservable to the alerting system. Note. however, that these results are based on the statistical terrain field that was generated as described in Appendix C, for an aircraft at 170 knots that flies a 2°/sec pull-up maneuver after a 5 sec latency when an alert is issued.

4.5 Summary

This chapter has demonstrated that the unified alerting system methodology can be applied to the evaluation of a terrain alerting system. The methodology can be used to examine the probability of false alarm or missed detection at a range of operating conditions and assumed pilot responses. In this manner, direct comparisons can be made between existing or proposed alerting thresholds and the probability of an incident or false alarm.

Because the pilot's response to an alert can be included probabilistically in the alerting system model, the effects of changes in pilot behavior can be examined. For example, it is possible to quantitatively compare the effect of a modification that decreases pilot response time to a modification that increases sensor accuracy. The costs of implementing these modifications can then be weighed against their benefits in terms of probabilities of missed detection or false alarm. In addition, the sensitivity of the probability of an incident to pilot blunders (e.g., not reacting in any manner to an alert) can be readily quantified.

In the performance evaluation of GPWS over flat terrain, a probabilistic model of the pilot's response to an alert is presented. Using this model, the GPWS alerting thresholds are shown to be consistent with a probability of an incident of approximately $1x10^{-7}$ when an alert is issued.

A successful method of generating a probabilistic model of terrain altitudes is also introduced. Using this terrain model, the probability of an incident is quantified at a range of descent rates and altitudes. For example, at a descent rate of 3,000 ft/min, the GPWS alerting threshold is shown to issue a false alarm with probability $1x10^{-4}$ in Smooth terrain and with probability 0.08 in Steep terrain. There is a corresponding increase in the

probability of a missed detection, changing from $<1x10^{-16}$ for Smooth terrain to 0.03 for Steep terrain.

A direct comparison of the GPWS alerting thresholds to the probability of an incident is also provided. The probability of an incident in Smooth terrain when a GPWS alert is issued is shown to be less than $1x10^{-8}$. In Steep terrain, this probability rises to greater than 0.01 for most descent rates.

These analyses show the limitations of GPWS that occur because information about terrain ahead of the aircraft is unavailable to the system. Were additional terrain information available from an onboard database or from sensors, the performance of GPWS could be improved when flying over rapidly varying terrain.

5. Example Application: Evaluation of the Traffic Alert and Collision Avoidance System

This chapter presents a different alerting system application for the unified methodology developed in Chapters 2 and 3. The application covered here involves small, rapidly moving hazards (other aircraft), in contrast to the large, static threat posed by terrain in the previous chapter.

The Traffic Alert and Collision Avoidance System (TCAS) is being used on transport aircraft in the U.S. to alert flight crews to potential mid-air collisions. Because of the challenging nature of this application, TCAS provides a good example with which to demonstrate the alerting system methodology.

5.1 Background

Following two mid-air collisions in the U.S. (a Boeing 727 and a Cessna near San Diego in 1978, and a DC-9 and a Piper at Cerritos, California in 1986), increased concerns were raised over the potential for future mid-air collisions. The Traffic Alert and Collision Avoidance System (TCAS) was implemented in the early 1990's on U.S. jet transport aircraft to alert flight crews to hazardous traffic and to provide instructions for avoiding a collision. Currently, TCAS is mandated on all U.S. passenger aircraft with more than 30 seats (FAA 1994: §121.356).

TCAS uses measurements of an intruding aircraft's range and altitude to estimate the collision danger. When a collision is predicted, vertical maneuver commands are given to the pilot to maximize the separation between the two aircraft. Two conflicting aircraft that are both equipped with TCAS coordinate their maneuver commands so that both aircraft do not follow the same course of action.

The alerting system has been designed to provide timely alerts to true threats and to keep false alarms to a minimum. However, because TCAS relies on only two measurements of the many parameters defining the dynamics of an encounter, its performance is inherently limited. The design tradeoffs in TCAS, therefore, provide a challenging example application for the alerting system methodology developed in the thesis.

5.2 Problem Statement

The examples used in this chapter concern a specific situation geometry between two aircraft that has been shown to cause false alarms during actual operations (Mellone & Frank 1993). The focus on this example allows for a detailed discussion of the design issues involved in the alerting system. The situation that is considered involves a "TCAS aircraft", which uses the alerting system under study, and an "intruder" aircraft that is assumed to not react to TCAS alerts. In this situation, the intruding aircraft is descending toward the TCAS aircraft. TCAS predicts a collision and issues an alert. However, the intruder levels off safely above the TCAS aircraft -- an alert was not truly needed.

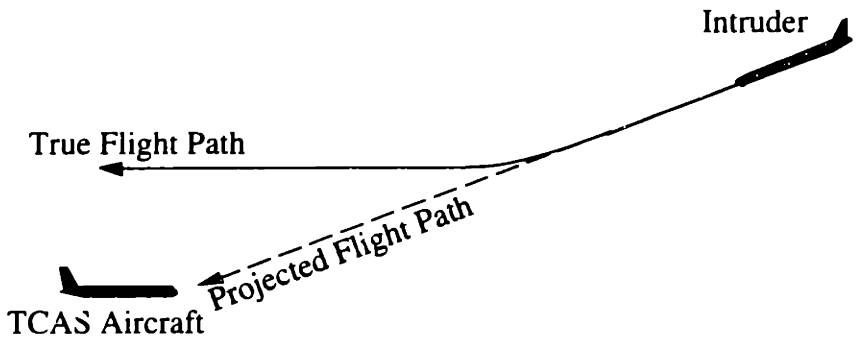


Figure 5.1: TCAS False Alarm Example

The situation shown in Figure 5.1 not only produces false alarms, but can also result in an incorrect vertical maneuvering command from TCAS. In some cases, TCAS may issue a command to the TCAS aircraft to climb to avoid a collision. A climb is the correct course of action if the intruder continues its descent, but if the intruder has leveled off, a climb may put the aircraft at additional risk.

Because of its tendency to produce false alarms and erroneous maneuvering commands, this hazard situation has resulted in the recommendation that aircraft slow their rates of descent when approaching their final altitude (Mellone & Frank 1993). In addition, the TCAS alerting thresholds have been modified several times to reduce false alarms in this type of encounter.

This chapter investigates the how these threshold modifications have affected the performance of the alerting system for the example situation shown in Figure 5.1. In particular, a direct comparison is made between the original version of TCAS (called here Version 1.0 (RTCA 1983)) and the latest implementation of TCAS called Version 6.04A,

which was mandated on U.S. aircraft in December, 1994 (Miller, et al. 1994, MITRE 1993).

In addition, a proposed modification to TCAS is examined. The fact that the intruder was intending to level off at a certain altitude could be transmitted to the TCAS aircraft. With this information, the false alarm in Figure 5.1 could be inhibited. The benefits of intent information are examined relative to the original TCAS Version 1.0. This exercise demonstrates the relative merits of changing the alerting thresholds on the existing system and of using additional information that more fully describes the hazard situation.

5.3 Architecture of TCAS

5.3.1 Dynamics of a Traffic Encounter

Figure 5.2 shows a generic traffic encounter situation. If the aircraft are assumed to move along straight trajectories, the situation can be described with a six-state vector that includes the relative position and velocity of the two aircraft. If the aircraft are turning or otherwise maneuvering, parameters such as turn rate and bank angle may need to be included.

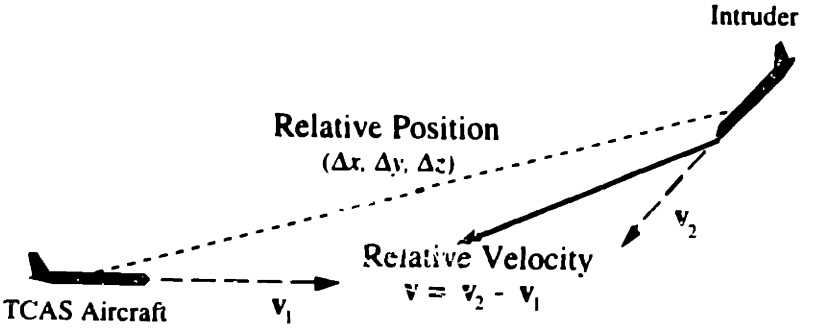


Figure 5.2: Generalized Traffic Encounter

The relative position of the two aircraft is defined as $(\Delta x, \Delta y, \Delta z)$ and the relative velocity is defined as $(\Delta v_x, \Delta v_y, \Delta v_z)$. The complete state vector, $\mathbf{x}(t)$, is then:

$$\mathbf{x}(t) = \left[\Delta x \; \Delta y \; \Delta z \; \Delta v_x \; \Delta v_y \; \Delta v_z\right]^T \tag{5.1}$$

As discussed in Appendix D, TCAS effectively has access to estimates of range (r), range rate ($\dot{r}$), relative altitude (h), and relative altitude rate ($\dot{h}$). Relative bearing information (ψ) is used for the alerting display but is not used by the alert criteria due to

large uncertainties. The observable state vector is then given by:

$$\mathbf{y}_a(t) = \left[r \; \dot{r} \; h \; \dot{h} \; \psi \right]^T \tag{5.2}$$

where $r = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$, $\dot{r} = \sqrt{\Delta v_x^2 + \Delta v_y^2 + \Delta v_z^2}$, $h = \Delta z$, $\dot{h} = \Delta v_z$, and $\psi = \tan^{-1}(\Delta y / \Delta x)$.

5.3.2 Alerting Thresholds

The current version of TCAS that is available on transport aircraft provides visual and aural indications of the traffic situation to the flight crew. The alerting criteria are based on the four states (r, $\dot{r}$, h, $\dot{h}$) that are inferred from measurements of range and altitude which are passed through an estimation system. An electronic traffic display is provided in the cockpit that uses three measurements (r, ψ, and h) to produce a plan-view depiction of the traffic situation.

TCAS uses two-stage alerting, with cautionary alerts called Traffic Advisories (TA) and warnings called Resolution Advisories (RA). TAs direct the crew's attention to a potential threat, but no avoidance information is provided. RAs also provide avoidance commands such as "Climb" or "Descend". Current procedures prohibit a maneuver in response to a TA unless the threat has been acquired visually, but mandate a maneuver in response to an RA (AA 1994). This chapter only considers Resolution Advisory (RA) alerts.

The alerting thresholds are rather complex and are described in more detail in Appendix D and in the TCAS operational specifications (RTCA 1976). However, for the simple traffic encounter situation considered here, the alerting thresholds can be reduced to a dependence on two parameters, called τ and DMOD. The parameter τ is a threshold on the estimated time to closest point of approach, and DMOD is a buffer distance around the TCAS aircraft. A range threshold requires that the intruder be projected to enter a sphere of radius DMOD around the TCAS aircraft within τ seconds. An altitude threshold requires that the intruder have a projected time to coaltitude less than τ seconds. The alerting thresholds can be summarized by Equation (5.3).

$$\text{if } r < (-\tau \dot{r} + \text{DMOD}) \text{ and } h < (-\tau \dot{h}), \text{ then issue an RA} \tag{5.3}$$

where the values of τ and DMOD are shown in Table 5.1.

Table 5.1
TCAS RA Alert Threshold Parameters
Aircraft at 15,000 ft MSL
(RTCA 1976, MITRE 1993)

Parameter	Version 1.0	Version 6.04A
τ	30 sec	22 sec
DMOD	1.0 nmi	0.8 nmi

Figure 5.3 shows a State-Space Hazard Diagram for the range vs. range-rate plane. The Figure shows a schematic of the different regions in state-space, assuming that the aircraft are on a collision course.

Hazard Space is defined by the extent of the cylinders that are used to model the aircraft (100 ft in radius and 100 ft high). Maneuver Limit Space is defined based on a 5°/sec flight path angle rate vertical pull-up maneuver as was used in Chapter 4.

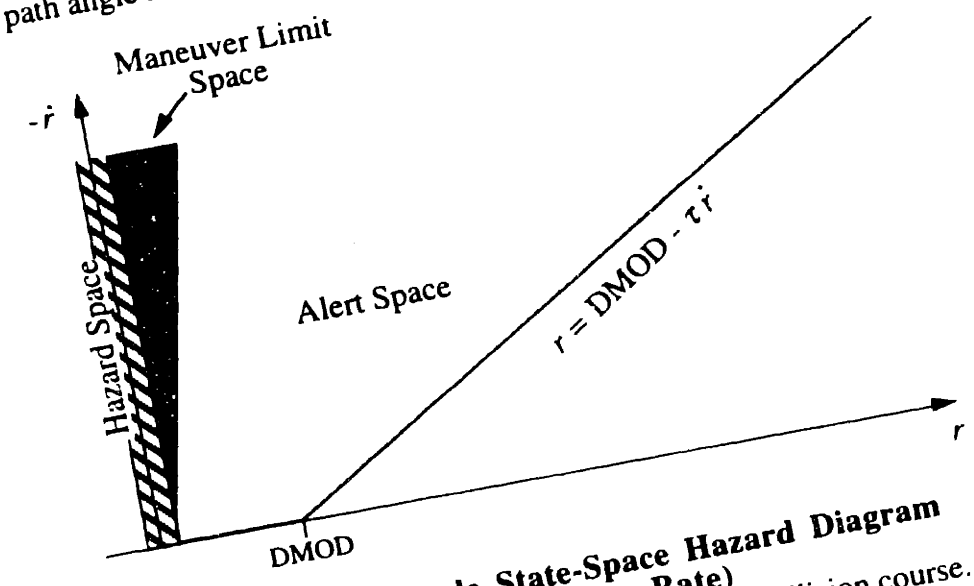


Figure 5.3: Example State-Space Hazard Diagram (Range vs. Range-Rate)
Assumes that aircraft are on a collision course.

Figure 5.4 shows a similar State-Space Hazard Diagram, but for the altitude vs. altitude-rate plane. Again, the aircraft are assumed to be on a collision course in this Figure.

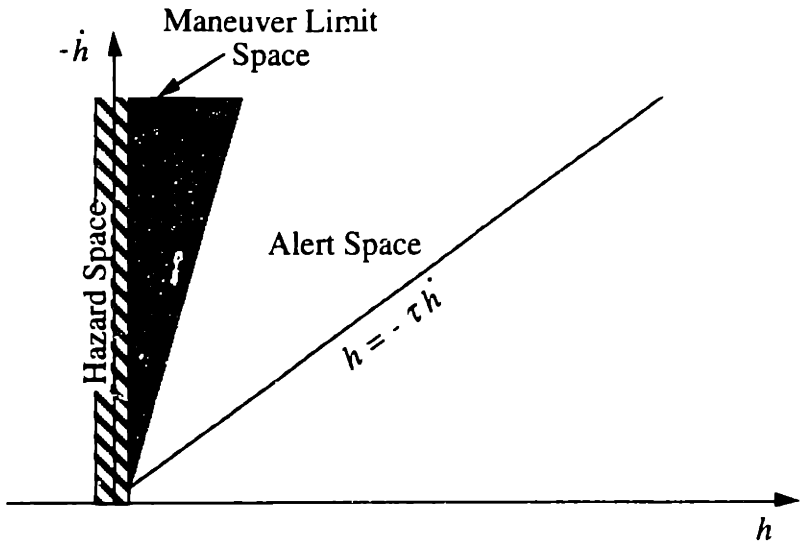


Figure 5.4: Example State-Space Hazard Diagram (Altitude vs. Altitude-Rate)
Assumes that aircraft are on collision course.

5.3.2.1 Resolution Advisory Selection

When an RA is issued, TCAS examines two potential avoidance maneuvers.[1] The TCAS aircraft's trajectory is extrapolated assuming a 5 second delay followed by a 0.25 g pitching maneuver to a vertical rate of 1,500 ft/min. Both a climbing and a descending maneuver are examined in relation to a straight-line extrapolation of the intruder's trajectory (Figure 5.5). The maneuvering advisory is chosen from the avoidance maneuver (either climb or descend) that provides the greatest vertical separation at the point of closest approach assuming that the intruder maintains a constant altitude rate. An aural alert of "Climb" or "Descend" is issued to the crew, as appropriate. In Figure 5.5, because the "Climb" projected vertical miss distance is larger than the "Descend" vertical miss distance, a "Climb" RA is chosen.

[1] The actual logic behind this decision is quite involved. A simplified discussion is given here, and more detail is available in Appendix D and RTCA (1983).

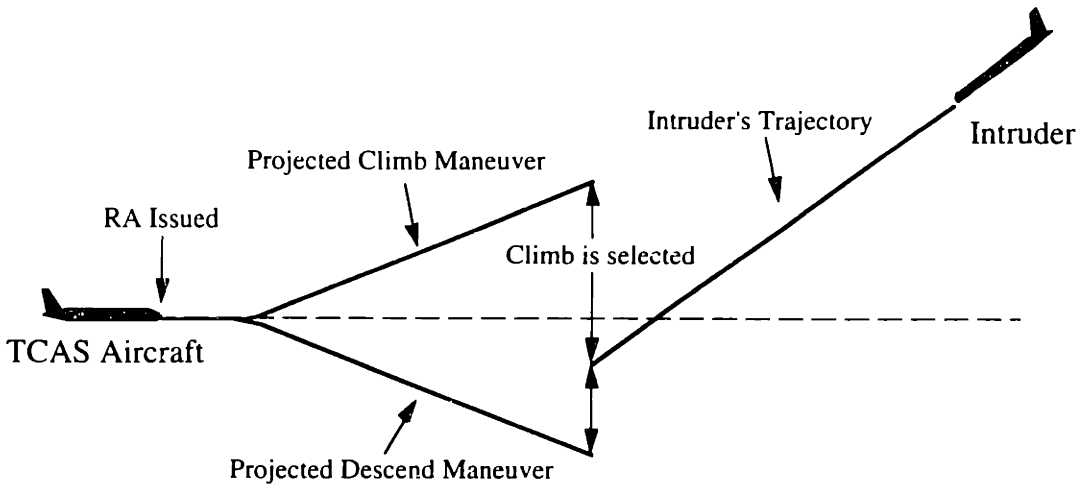


Figure 5.5: Example Selection of Climb or Descend RA

5.3.3 Alerting Displays

TCAS provides a visual display of the traffic situation, and augments this with visual and aural alerts when TAs or RAs are issued. The traffic display is normally provided in the cockpit either as part of the navigation display (Electronic Horizontal Situation Display) or as a separate display.

An example is shown in Figure 5.6. The display shows an aircraft track-up plan-view depiction of the traffic around the aircraft and can be changed in scale (AA 1994). Other aircraft are shown as icons on the display, and the pilot may select depictions showing the absolute altitude (MSL) of the intruder or the intruder's altitude relative to the TCAS aircraft. In addition, when TCAS estimates that the threat's altitude rate is at least 500 feet per minute, an arrow pointing up or down is shown next to the icon.

TCAS uses four stages of depiction on the display, distinguishing immediate threats from those that do not currently pose a hazard. Traffic beyond 6 nmi and 1,200 ft vertically is shown as a hollow white or light-blue (cyan) diamond icon ('Other Traffic'). When an aircraft closes within 6 nmi and 1,200 ft ('Proximate Traffic'), the diamond icon becomes filled. Aircraft within the TA thresholds are shown as filled amber circles on the display along with an aural "Traffic" alert. RAs are shown as red squares along with appropriate aural information such as "Climb".

Although the bearing of an intruder is shown on the display, pilots are instructed to use only vertical maneuvers to avoid traffic during an RA. The bearing information

contains large uncertainties and is intended to be used only to aid the crew in visually acquiring a threat.

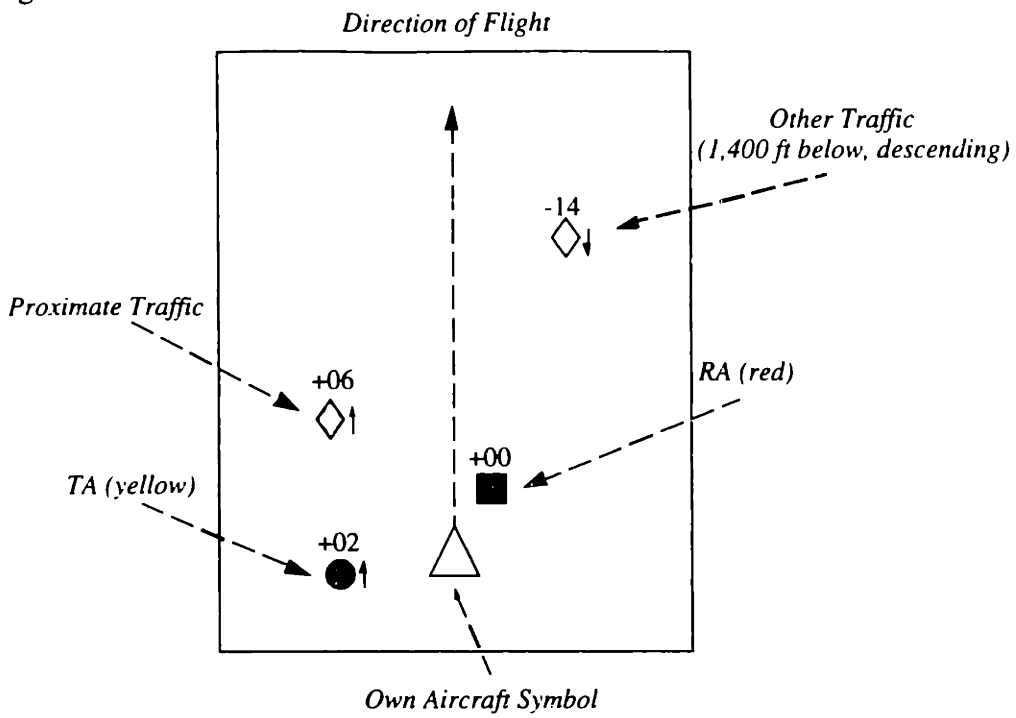


Figure 5.6: Diagram of TCAS Traffic Display

A graphical indication of the RA escape guidance is also provided on the vertical speed indicator. Green and red bands are used to indicate vertical speeds that should and should not be used, respectively.

5.4 TCAS Performance Evaluation

For the examples in this chapter, the hazard situation shown in Figure 5.1 is used to generate the state measurements that are available to TCAS. The TCAS aircraft is assumed to be flying at an altitude of 15,000 ft above Mean Sea Level (MSL). An intruder aircraft is flying directly toward the TCAS aircraft on a reciprocal course at a measured relative speed of 440 knots (733 ft/sec) and with a measured descent rate of 2,500 feet per minute. The two aircraft are modeled as cylinders 100 ft in radius and 100 ft high. If the cylinders intersect, a mid-air collision has occurred.

The intruder's trajectory is set so that the intruder is projected to pass 80 ft below the TCAS aircraft. This vertical miss distance is used so that nominally a collision will occur, but because the intruder is passing below the TCAS aircraft, TCAS will issue an erroneous "climb" command as is discussed in Section 5.4.1. In addition, the intruder

aircraft is assumed to not be using TCAS. Therefore, the intruder will not react during the situation.

As discussed in Appendix D, the uncertainties of the observable states generally decrease as the amount of time that the intruder has been tracked increases. For the purposes of the examples in this chapter, however, it is assumed that the TCAS estimators have converged to the steady-state errors given in Table 5.2. These values were obtained through a Monte Carlo simulation of state measurements using the TCAS estimation logic. The state estimates converge on these values before the intruder has been tracked for approximately 60 seconds, during which time both aircraft are assumed to follow straight lines.

Table 5.2
Estimate Accuracies for TCAS Examples
60 sec tracking time, 400 knot closure rate, 2,500 ft/min descent rate
(See Appendix D)

Parameter	Estimate Standard Deviation
r	18 ft
$\dot{r}$	6 ft/sec
h	82.6 ft
$\dot{h}$	0.56 ft/sec

5.4.1 Version 1.0: Intruder Descending Through Own Altitude

The examination of TCAS performance begins with a look at the situation where an intruder is flying directly toward the TCAS aircraft and is descending through the TCAS aircraft's altitude without leveling off (Figure 5.7). As mentioned in Section 5.2, this situation has been known to produce false alarms and is an important test case.

The Vertical Miss Distance (VMD) for this situation is a function of the range, range rate, relative altitude, and vertical rate:

$$\text{VMD} = \dot{h} - h\frac{\dot{r}}{r} \tag{5.4}$$

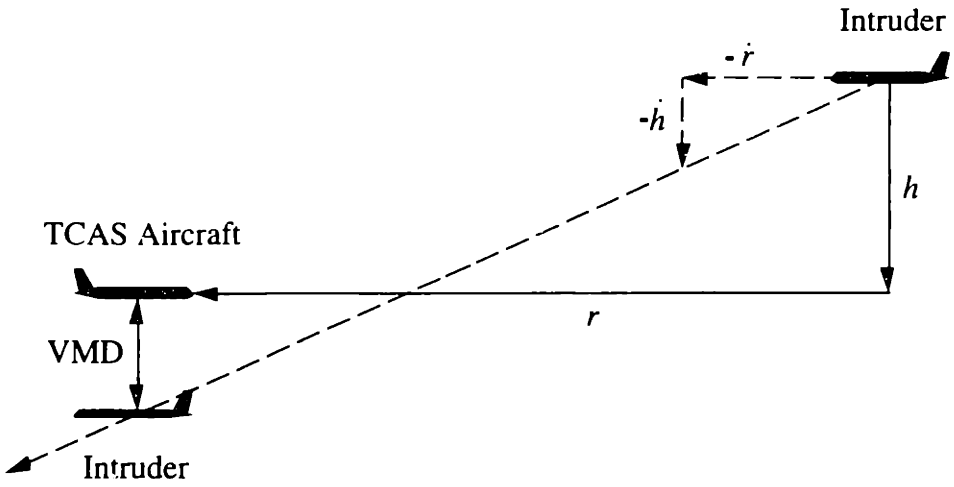


Figure 5.7: Intruder Descending Through Own Altitude

Following the discussion in Chapter 2, the measurements are assumed to be known deterministically and it is of interest to find the probability that a certain true situation has produced these estimates. Table 5.2 provides the standard deviations of all of the state estimates. Therefore, for each measured value, the probability that the state is truly some other value can be found.

The measurements described above were passed as inputs into a simulation of the TCAS range and altitude estimators to obtain estimated state values (Appendix D). Figure 5.8 shows a two-state diagram of the estimated altitude and range from the simulation. The intruder's estimated state is shown closing in range and altitude with the TCAS aircraft located at the origin in the Figure. Also shown are the Version 1.0 TCAS RA alerting thresholds for range and altitude from Table 5.1. Because the altitude alerting thresholds are sensitive to estimated altitude rate (which oscillates slightly in the simulation), the thresholds show some variability. As shown, an RA is issued at a range of approximately 24,000 ft, 32 seconds before the time of closest approach. In this example, the alert is issued more than $\tau = 30$ seconds from time of closest approach because the estimated descent rate at the time of alert is larger than the true value due to the dynamics of the altitude tracker. Note, however, that Figure 5.8 does not include the change in states that would occur due to an avoidance maneuver -- the Figure assumes that no action is taken after the RA is issued.

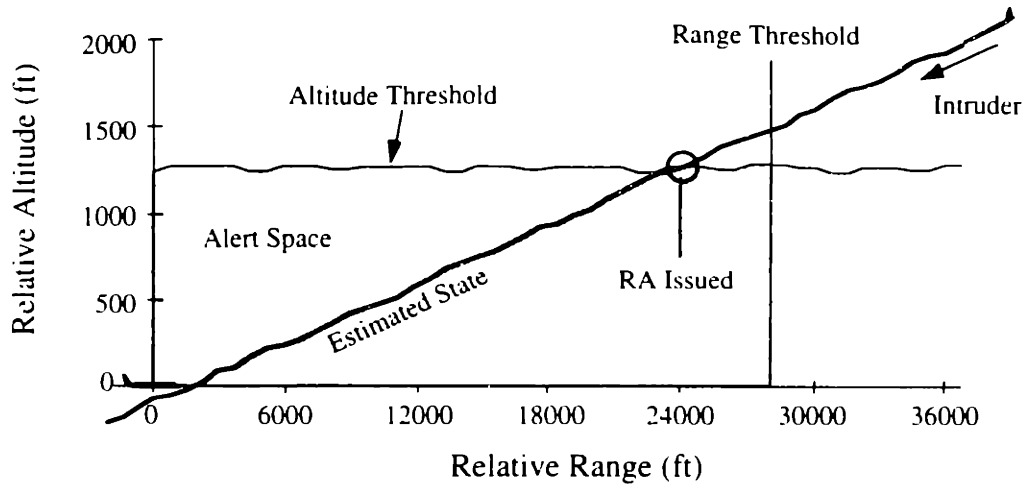


Figure 5.8: Simulated State and RA Alerting Thresholds, V 1.0 (Intruder Passing Through Own Altitude)

When the RA is issued, TCAS first determines whether it should issue a climb or a descend advisory. Figure 5.9 shows the projected trajectories of the intruder in relation to a "Climb" or "Descend" RA. The vertical miss distances for this situation are 732 ft for "Climb" and 540 ft for "Descend". Because the "Climb" response produces a larger vertical miss distance, it is selected as the RA.

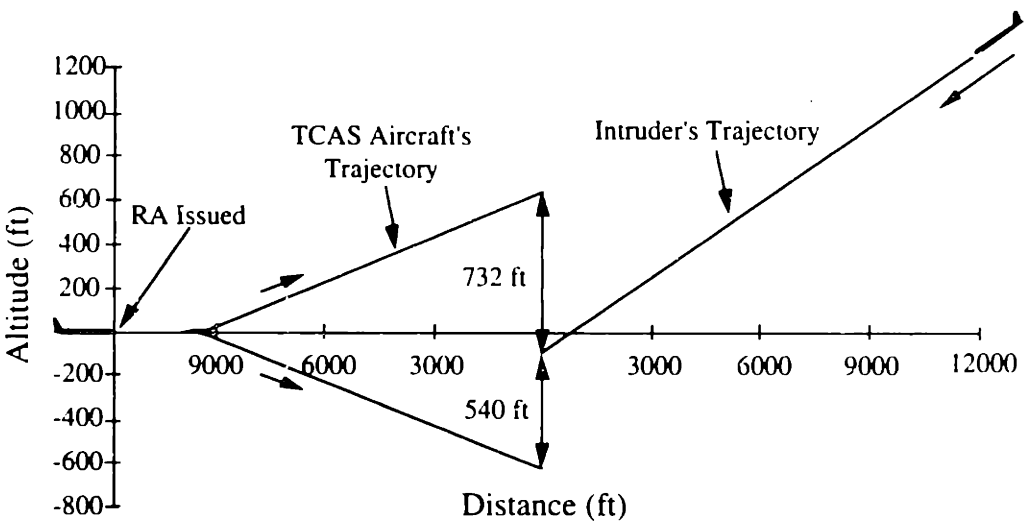


Figure 5.9: Climb and Descend RA Avoidance Maneuvers, V 1.0 (Intruder Passing Through Own Altitude)

5.4.2 Version 1.0: Intruder Stopping Descent Above Own Altitude

Next, consider a situation in which the intruding aircraft levels off 1,000 ft above the TCAS aircraft as discussed in Section 5.2. The intruder is assumed to perform a 0.1 g pitching maneuver as it levels off.

Figure 5.10 shows a range-altitude space diagram for the Version 1.0 TCAS RA thresholds. The altitude threshold is shown given the intruder's descent rate at each range value. Thus, as the altitude tracker begins to reflect the intruder's slackened descent rate, the altitude threshold drops in altitude.

Note that an RA is issued even though one is not truly needed. The Figure shows that the state vector leaves the alerting threshold region after several seconds. The RA is not ended, however, because once issued the TCAS logic requires an RA to continue as long as the intruder is closing in range with the TCAS aircraft. When the RA is issued, the state estimates are the same as that discussed in Section 5.4.1 -- TCAS has no way to predict that the intruding aircraft is not truly a threat.

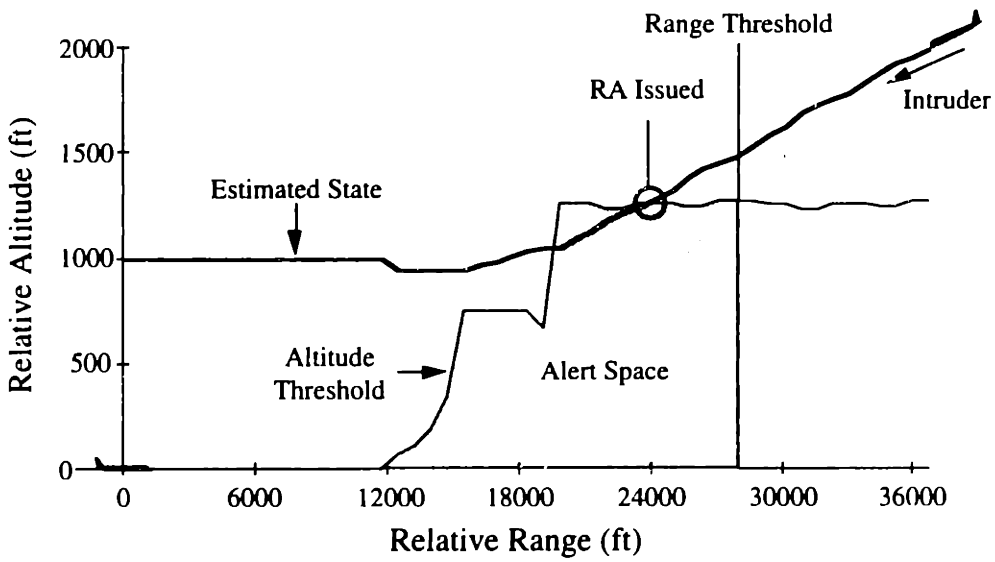


Figure 5.10: Simulated State and RA Alerting Thresholds, V 1.0 (Intruder Leveling Off 1,000 ft Above Own Aircraft)

Because the trajectory of the intruding aircraft (when the RA is issued) appears to be the same as that when no level off is made, TCAS again selects a "Climb" advisory. Note, though, that this maneuver actually results in a dangerous situation as the TCAS aircraft climbs toward the now-leveled intruder (Figure 5.11). The resulting vertical miss distance is only 364 ft.

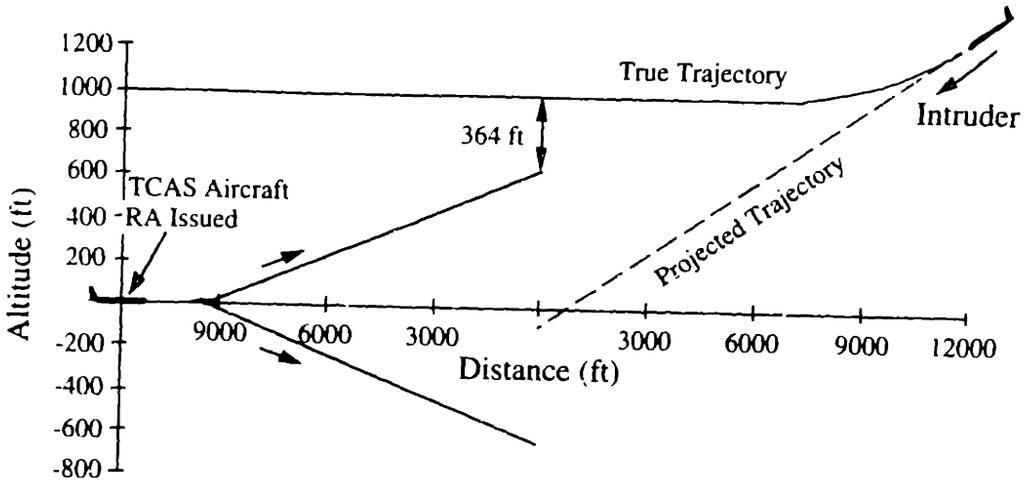


Figure 5.11: Climb and Descend RA Avoidance Maneuvers, V 1.0 (Intruder Leveling Off 1,000 ft Above Own Aircraft)

5.4.3 Version 6.04A: Intruder Descending Through Own Altitude

TCAS Version 6.04A uses the modified alerting thresholds shown in Table 5.1. The most significant change as far as this example is concerned involves the time to closest point of approach threshold, τ, which has been reduced from 30 sec to 22 sec. This change results in less available time to react to an RA, but eliminates the false alarm problem shown in Section 5.4.2. Figure 5.12 shows the same situation as was discussed in Section 5.4.1 in which the intruder passes through the TCAS aircraft's altitude, but with the TCAS Version 6.04A thresholds. An RA is issued approximately 23 seconds before the closest point of approach.

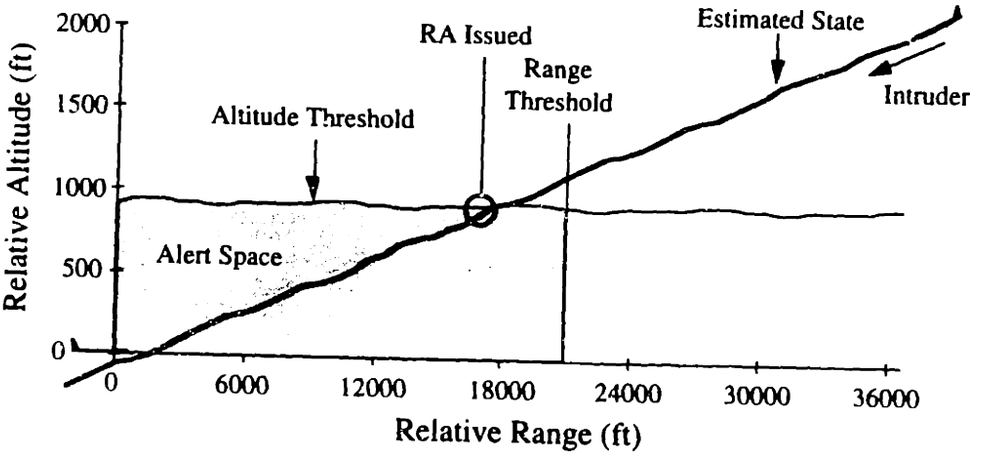


Figure 5.12: Simulated State and RA Thresholds, V 6.04A (Intruder Passing Through Own Altitude)

Because the RA is issued at a later time than in Version 1.0, the projected vertical miss distance between the TCAS and intruding aircraft is smaller than that in Section 5.4.1. A "Climb" RA is again issued. Figure 5.13 shows the vertical profile of the aircraft trajectories in this situation.

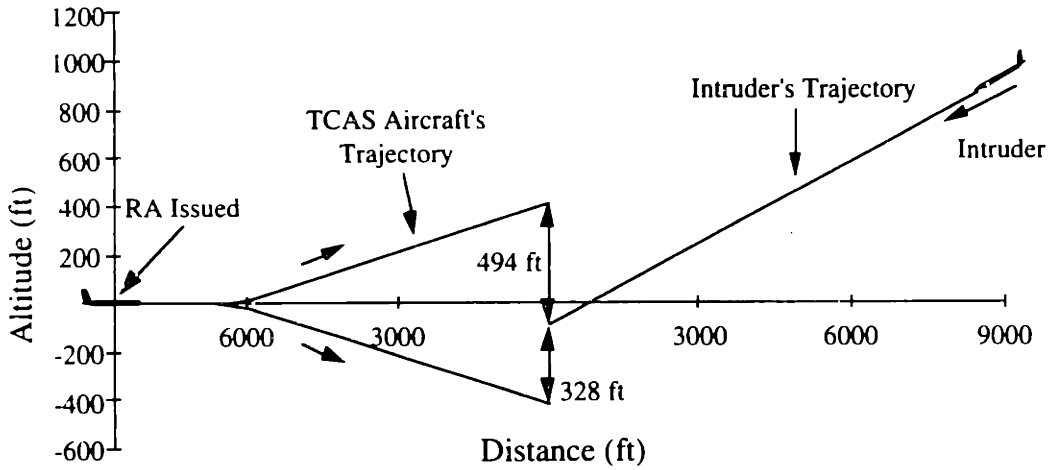


Figure 5.13: Climb and Descend RA Avoidance Maneuvers, V 6.04A (Intruder Passing Through Own Altitude)

5.4.4 Version 6.04A: Intruder Stopping Descent Above Own Altitude

Next, the intruder is assumed to level off 1,000 ft above the TCAS aircraft. Figure 5.14 shows that the state vector never enters the alerting region, so an RA is not issued.

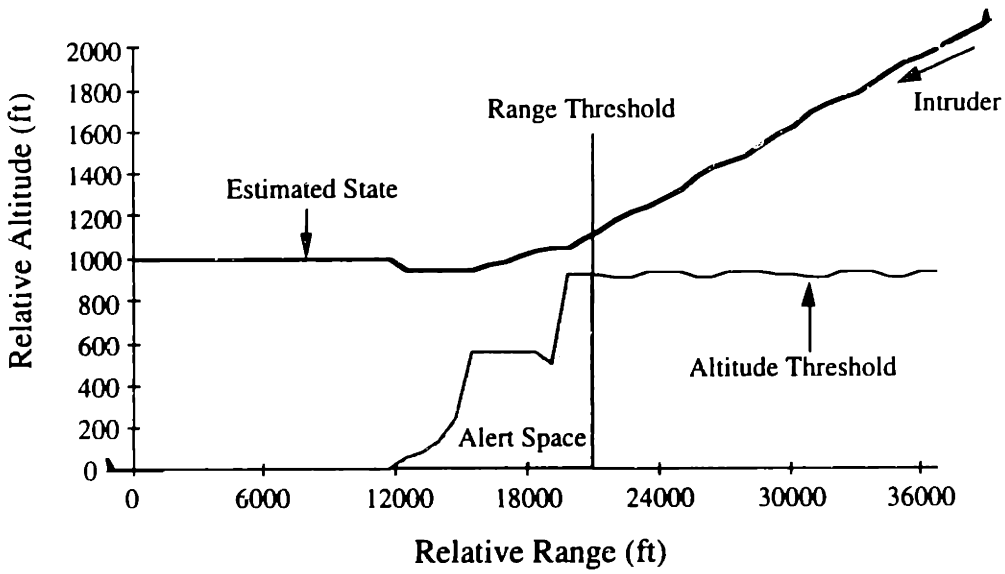


Figure 5.14: Simulated State and RA Thresholds, V 6.04A (Intruder Leveling Off 1,000 ft Above Own Aircraft)

5.5 Performance Comparison: Versions 1.0 and 6.04A

To demonstrate the tradeoffs between false alarms and missed detections for the Version 1.0 and 6.04A TCAS thresholds, the probabilistic analysis method developed in Appendix A is used. The estimated state values are assumed given, and the uncertainties in Table 5.2 are used to determine the probability that an alert is truly needed. For these examples, the aircraft are modeled as cylinders 100 ft in radius and 100 ft high. If the relative trajectory between the aircraft is such that the cylinders intersect somewhere along the trajectory, an incident is said to occur. Following the method in Appendix A, the probability that an incident occurs along a probabilistic trajectory T is given by the following equation:

$$P_T(I) = \int_{\mathbf{T}} \int_{\mathbf{Z}} f_\tau(\tau)\, f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x}\, d\tau \tag{5.5}$$

where $\tau \in \mathbf{T}$ is a possible trajectory between the two aircraft, and $\mathbf{Z}$ is the region in space where the intruder and the TCAS aircraft must be located simultaneously for an incident to occur.

As was shown in Figure 5.7, the direction of the trajectory of the threat relative to the TCAS aircraft is a function of the descent rate and the range rate. Accordingly, the integral over $\mathbf{T}$ in Equation (5.5) is a double integral over $\dot{h}$ and $\dot{r}$. Given a trajectory defined by $\dot{h}$ and $\dot{r}$, the probability that an incident will occur is then found by integrating the range and altitude error probability density functions over those values for which the two aircraft cylinders intersect (defined as $\mathbf{Z}$). The integral over $\mathbf{Z}$ begins by integrating the PDF for range over all possible values. Once $\dot{h}$, $\dot{r}$, and r are specified, the nominal vertical miss distance (VMD, Equation 5.4) can be found, and the limits of integration for the PDF of h can be determined. Recalling that the aircraft are modeled as 100 ft high cylinders, Figure 5.15 shows how the limits of integration for the altitude PDF are obtained for a given vertical miss distance.

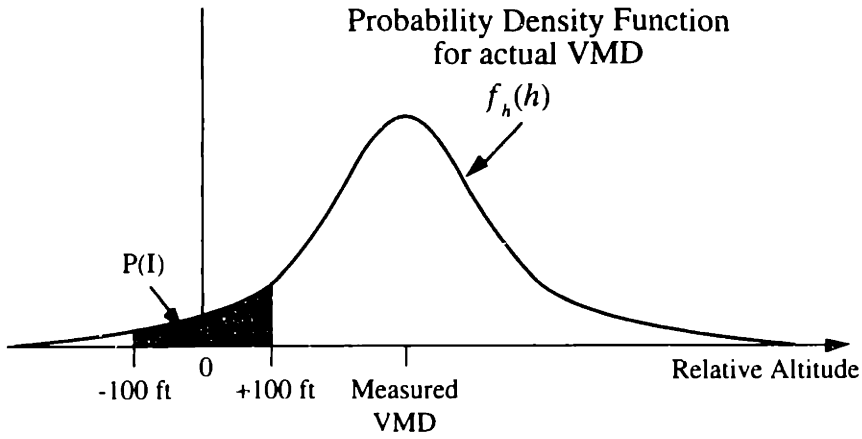


Figure 5.15: Limits of Integration for Altitude PDF

Expanding the integrals over **T** and **Z**:

$$P_T(I) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-100}^{+100} f_{\dot{h}}(\dot{h})\, f_{\dot{r}}(\dot{r})\, f_r(r)\, f_h(h)\, dh\, dr\, d\dot{r}\, d\dot{h} \qquad (5.6)$$

Equation (5.6) was solved through numerical integration for the vertical miss distances calculated for the two TCAS Versions in the situations covered in Sections 5.4.1 through 5.4.4. As Table 5.3 shows, the probability of an incident when the intruder levels off is reduced from $7x10^{-4}$ with Version 1.0 (VMD = 364 ft) to less than $1x10^{-16}$ with Version 6.04A (VMD = 1,000 ft). This improvement, however, comes at the cost of increasing $P_A(I)$ for the case in which the intruder does not level off (going from $1x10^{-14}$ to $9x10^{-7}$ between Versions 1.0 and 6.04A, respectively).

Table 5.3
Comparison of $P_A(I)$
TCAS Versions 1.0 and 6.04A

TCAS Version	Intruder Descends Through Own Altitude	Intruder Levels Off
1.0	$1x10^{-14}$	$7x10^{-4}$
6.04A	$9x10^{-7}$	$<1x10^{-16}$

Whether the tradeoffs involved with changing to Version 6.04A are acceptable depends on the relative likelihood between situations in which the intruder levels off above the TCAS aircraft and situations in which the intruder continues its descent. For example, assume that the intruder levels off 1,000 ft above the TCAS aircraft with a probability of 0.75, and that the intruder will continue to descend through the TCAS aircraft's altitude with a probability of 0.25. As shown in Figure 5.16, a window of potential altitudes at

which the level off maneuver begins is also defined. This window is centered on the nominal altitude at which a 0.1 g pull up maneuver would begin, and has a standard deviation of 2 seconds of flight time (approximately 80 ft at the descent rate of 2,500 ft/min).

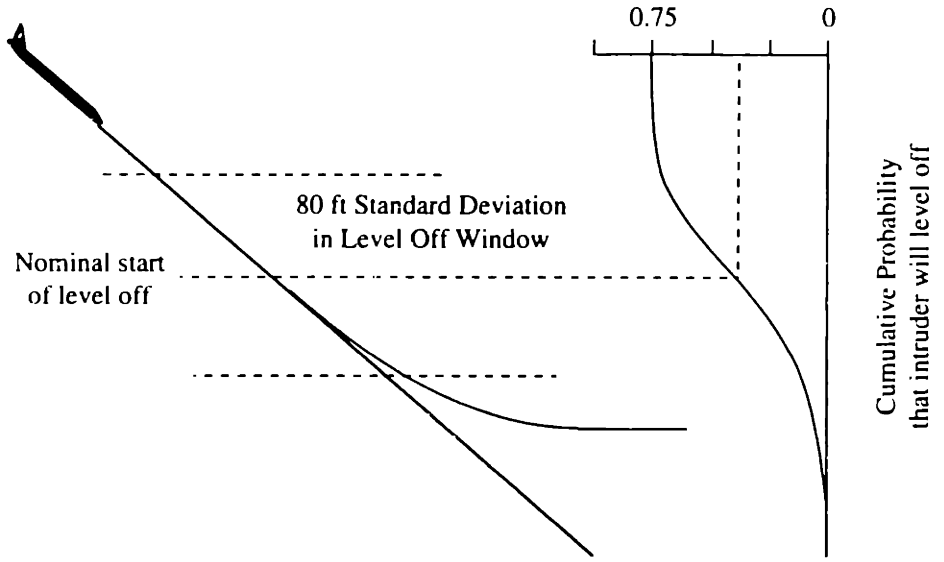


Figure 5.16: Example Probability that Intruder Will Level Off

Following the discussing in Section 2.7, two future trajectories for the TCAS aircraft are considered. In the Nominal Trajectory case (N), the TCAS aircraft is assumed to continue at its current altitude. In the Avoidance Trajectory case (A), the TCAS aircraft is assumed to fly a standard avoidance maneuver: after a 5 second delay, the aircraft performs a 0.25 g pull-up maneuver until a 1,500 feet per minute climb rate has been achieved (RTCA 1976).

Equation (5.6) was numerically integrated for varying relative altitudes between the intruder and the TCAS aircraft. Figure 5.17 shows curves of the probability that an incident will occur along N, $P_N(I)$, and along A, $P_A(I)$. Recall from Chapter 3 that $P(FA) = 1 - P_N(I)$ and $P(CD) = 1 - P_A(I)$ evaluated at the alerting threshold. Overlaid on these curves are the locations of the alerting thresholds for TCAS Versions 1.0 and 6.04A.

The curve of $P_N(I)$ shows the probability that an incident will occur if the TCAS aircraft does not maneuver, as a function of the relative altitude between the two aircraft. At large relative altitudes (location 1 in the Figure), $P_N(I)$ is approximately 0.14 -- this is the probability that an alert will eventually be needed when the two aircraft are still far from one another. As the relative altitude decreases and the two aircraft come closer together, the probability that the intruder will not level off increases, resulting in a corresponding

increase in $P_N(I)$. At low relative altitudes (location **2**), $P_N(I)$ is approximately 0.57. Note that even at a measured relative altitude of 0 ft, the probability that an incident will occur is less than 1 due to the relatively large uncertainties in the state estimates (especially altitude).

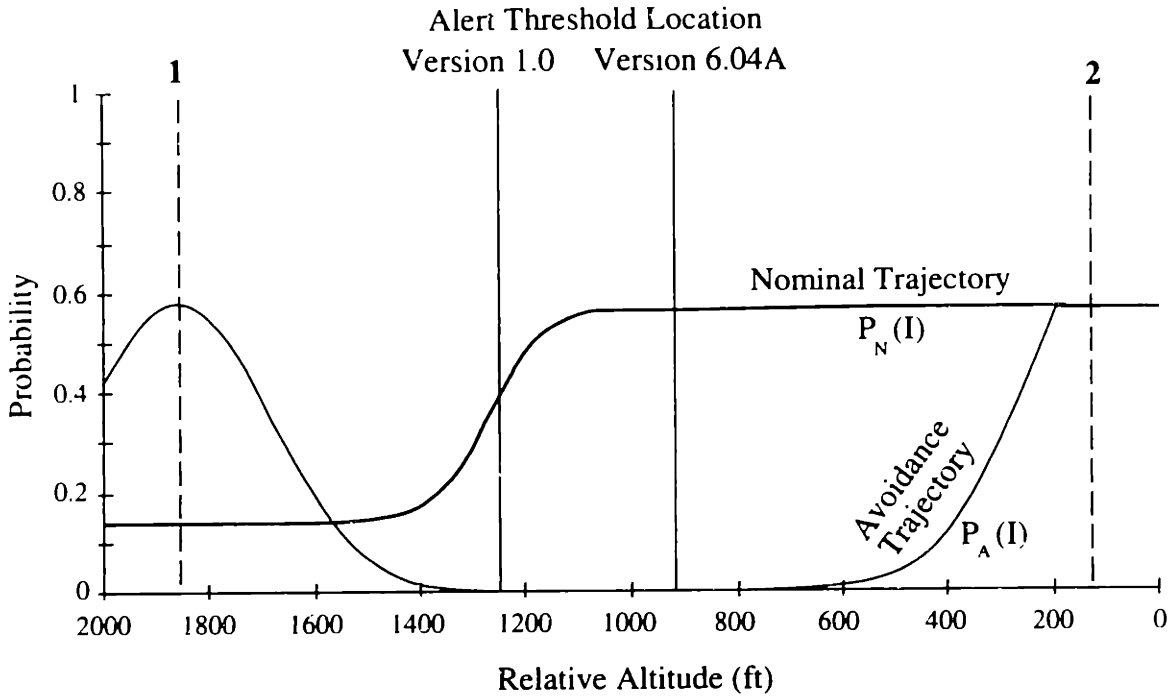


Figure 5.17: Probability of Incident vs. Relative Altitude TCAS Version 1.0 and 6.04A Threshold Comparison

The curve of $P_A(I)$ shows two regions of local maximum. Between relative altitudes of approximately 2,500 ft and 1,300 ft (location **1**), a "Climb" maneuver risks a collision with an intruder that has leveled off. Below a relative altitude of 700 ft (location **2**), $P_A(I)$ again rises, because there is not enough time available to avoid a threat that has not leveled off. At relative altitudes above 2,500 ft, a "Climb" RA will allow the TCAS aircraft sufficient time to climb well above the intruder's trajectory whether the intruder levels off or not.

Figure 5.18 shows an SOC curve for the data contained in Figure 5.17. Recall that the SOC curve is a plot of P(FA) and P(CD) at each possible alerting threshold location. The interesting shape of the SOC curve is due to the fact that both high and low thresholds (locations **1** and **2** in Figure 5.17) result in a low P(CD) in this situation. Because, as was shown in Figure 5.17, the probability of a false alarm does not drop below approximately 0.43, the SOC curve does not reach the y-axis in the plot.

As shown, the change in thresholds from Version 1.0 to Version 6.04A results in the reduction of the probability of a false alarm (summarized in Table 5.4). Furthermore, because the probability that the TCAS aircraft will climb into a leveled intruder is lower, P(MD) is also lower in Version 6.04A than Version 1.0. As can also be seen, Version 6.04A appears to have moved the threshold as far as possible before P(CD) drops sharply.

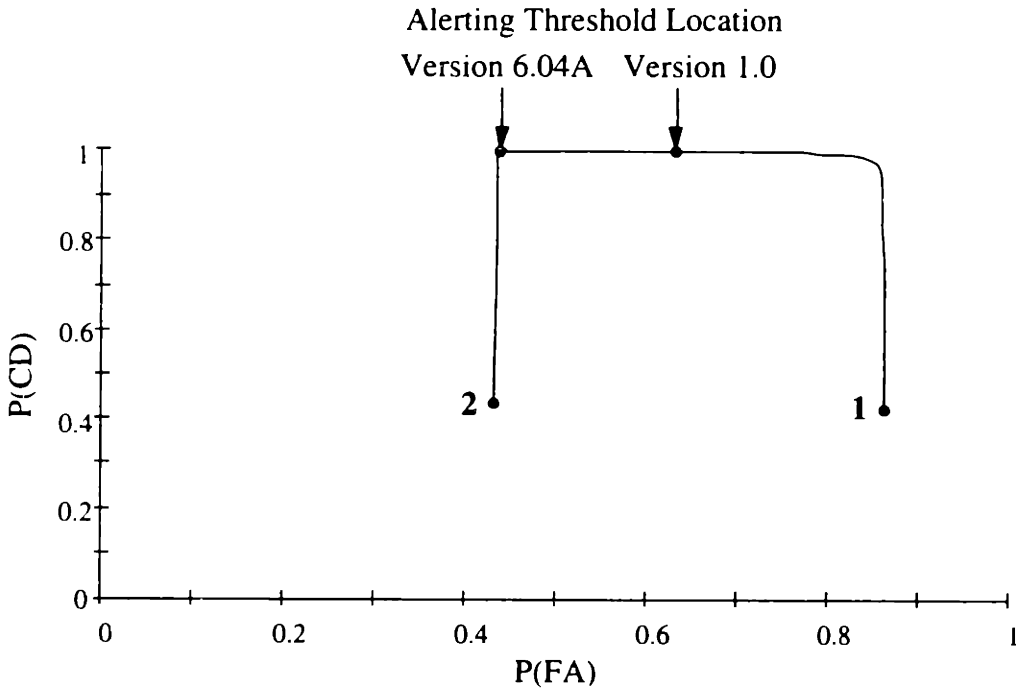


Figure 5.18: SOC Curve for TCAS Versions 1.0 and 6.04A

Table 5.4
Comparison of Late and Nuisance Alert Probabilities
TCAS Versions 1.0 and 6.04A
Intruder levels off with probability 0.75

TCAS Version	P(MD)	P(FA)
1.0	2×10^{-4}	0.63
6.04A	1×10^{-11}	0.44

This example is only one of many potential encounter geometries. A full evaluation of the changes from Version 1.0 to 6.04A would require an examination of this situation at other descent rates. At other descent rates, there may be less benefit from the TCAS threshold change than is evident in this example.

5.6 Analysis of the Benefit of Intended Trajectory Information

The modifications of TCAS resulting in Version 6.04A involved changes in the alerting thresholds. However, the total information available to TCAS regarding the intruder did not change. Thus, attempts to improve performance must follow the SOC curve in Figure 5.18. Another method that could be used to reduce false alarms is to provide TCAS with additional information from the intruding aircraft such as its intended level-off altitude. TCAS could then use this information to inhibit alerts for an intruder that will be leveling off, even when the standard TCAS alerting thresholds predict a potential collision. As an additional performance comparison, this section investigates the benefit of combining additional intent information with the Version 1.0 alerting thresholds.

A flowchart of the alerting logic used in this example is shown in Figure 5.19. Using the terminology from Chapter 2, the modification of TCAS considered here uses Intended Trajectory Extrapolation to determine whether an alert should be issued or inhibited. If no intent information is available, the system reverts to the alerting threshold logic covered in Section 5.3. However, if the level-off altitude of the intruder is available to TCAS, the system can inhibit alerting when the intruder is intending to arrest its descent safely above the TCAS aircraft. However, if it becomes clear that the intruder is deviating from its intended flight path, TCAS must disregard the intent information and issue an alert.

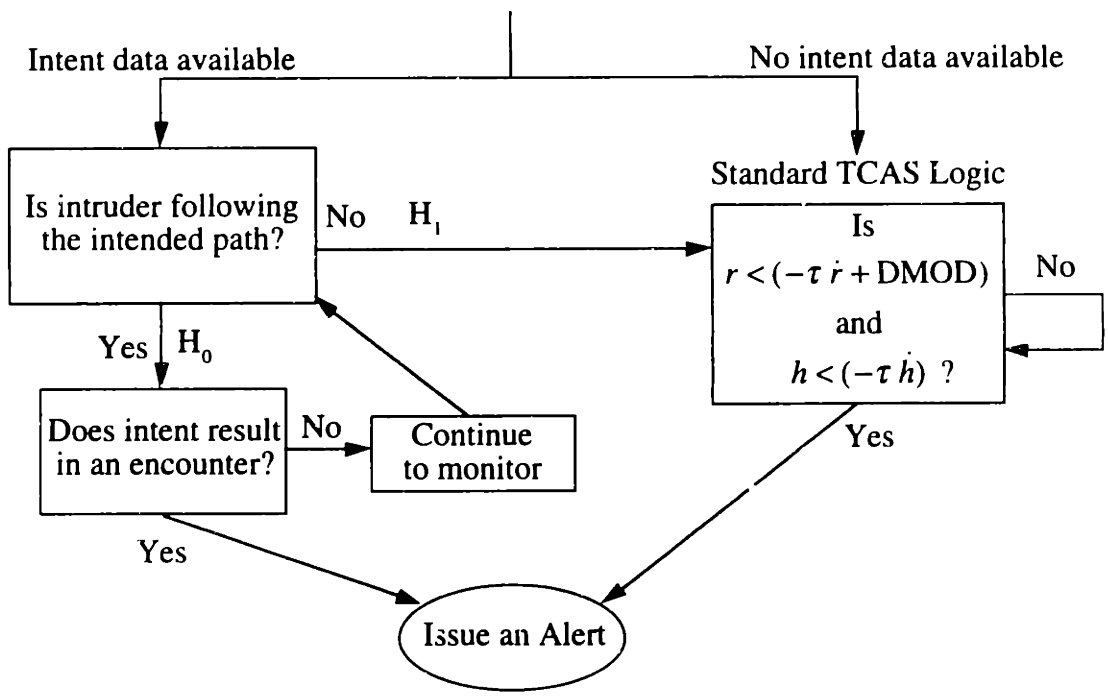


Figure 5.19: Logic Flowchart for TCAS with Intent Information

When provided with intent information, the TCAS logic continually decides between two hypotheses: H_0 is the hypothesis that the intended path is being followed, and H_1 is the hypothesis that the intruder is not following the intended path. To decide between the two hypotheses, the system must determine the confidence with which to believe that each hypothesis is being followed. For example, let $P(H_0)$ be an estimate of the probability that H_0 is being followed. The decision between the two hypotheses depends on a simple confidence test based on a threshold probability, p, as shown in Figure 5.20.

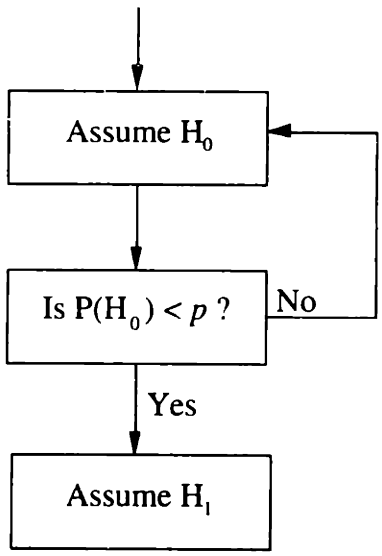


Figure 5.20: Intended-Path Hypothesis Test

The choice of p requires an examination of the effects that p has on false alarm and missed detection probabilities. If p is very large, then intent information is often disregarded and the benefits that might accrue are ignored. However, if p is very small, then the intent information may be believed even when it is not being followed, increasing the likelihood of a false alarm or missed detection.

Returning to the example situation studied in Sections 5.4.1 and 5.4.2, consider an aircraft that is descending toward the TCAS aircraft but is leveling off 1,000 ft above the TCAS aircraft. Also assume that the intruder has transmitted its leveling altitude to the TCAS aircraft.

As shown in Figure 5.21, the intended path of the intruder is such that no conflict is predicted; an alert that would otherwise be issued by the Version 1.0 thresholds is therefore inhibited. However, if measurements indicate that the intruder is straying from the intended path, the probability that the intended path is not being followed increases. If the

probability that the intruder will continue its descent results in the rejection of H_0, TCAS reverts to the standard thresholds and issues an alert. However, because in such a situation the alert has been delayed, the margin of safety built into the standard thresholds will decrease. The choice of p, therefore, will affect both the overall safety level of the system and the ability of the system to inhibit alerting in situations where a nuisance alert could be generated.

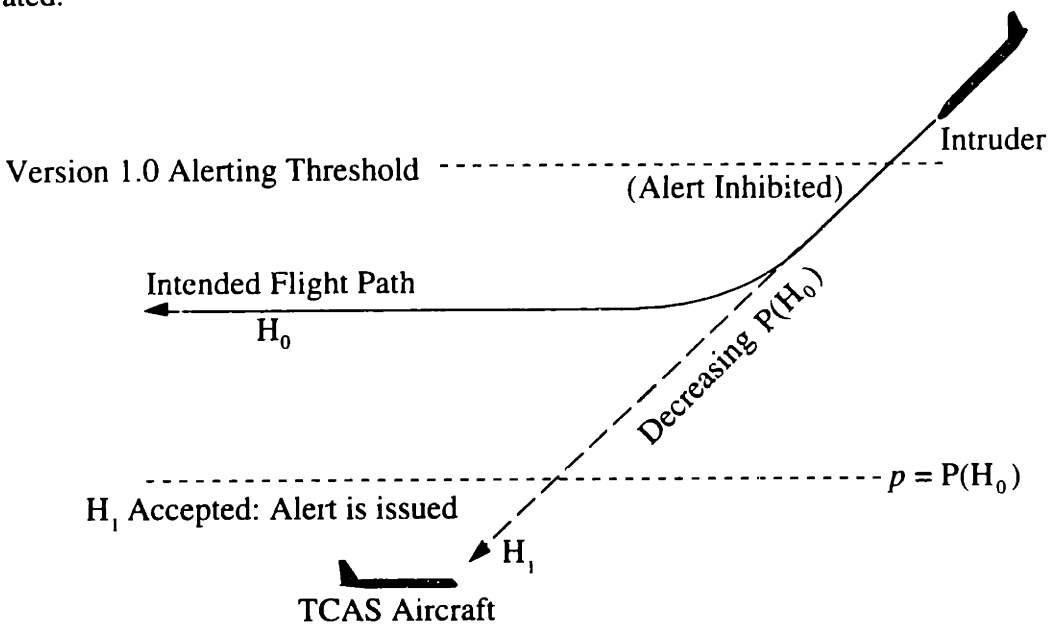


Figure 5.21: Example Situation: TCAS with Intent Information

The example TCAS system considered here assumes that nominally, intruding aircraft fly a 0.1 g maneuver such that they level off on the intended altitude. This maneuver can be shown in a state-space diagram (Figure 5.22). Statistically, it is assumed in this example that intruders begin this level off maneuver at an altitude that is centered on the nominal altitude at which a level-off maneuver would begin with a normal distribution that has a standard deviation of 2 seconds of flight time (approximately 80 ft at a descent rate of 2,500 ft/min).

The hypothesis test threshold, p, therefore creates an effective alerting threshold as shown in Figure 5.22. If the intruder crosses this threshold, H_1 is accepted and an alert is issued.

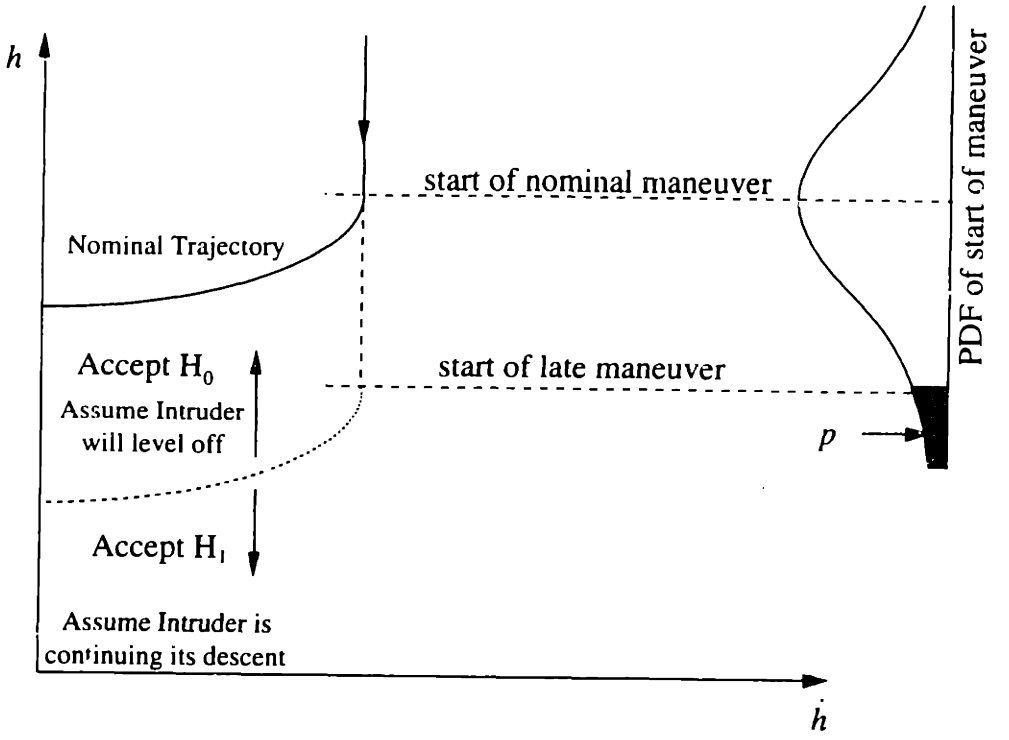


Figure 5.22: Alerting Threshold when Using Intent Information

A comparison of the Version 6.04A alerting threshold location and the probability density function in Figure 5.22, however, shows that when a Version 6.04A TCAS alert is issued, the probability that the intruder is intending to level off 1,000 ft above the TCAS aircraft is very low ($<1\text{x}10^{-5}$). Therefore, there is little benefit in inhibiting this alert based on intent information that is most likely not being followed. Thus, in this situation, the addition of intent information does not provide a significant performance benefit for TCAS over that provided by Version 6.04A.

5.7 Summary

This chapter provides an application of the alerting system methodology to an alerting system used on civil aircraft for the prevention of mid-air collisions (TCAS). The basic properties of TCAS are illustrated for an example situation that is known to produce false alarms in actual operations. This situation serves as a challenging example with which to demonstrate the usefulness of the methodology.

Using a probabilistic model of the state estimate uncertainties, the methodology is successfully used to determine the probability that a TCAS alert is a false alarm or correct detection. Two versions of TCAS are examined: the original Version 1.0, and the updated Version 6.04A.

The methodology shows that Version 6.04A can reduce the probability of a false alarm (compared to Version 1.0) to the approximate minimum value that is possible given the information TCAS has available before the probability of missed detection increases rapidly. In addition, the probability of an incident is shown to decrease slightly with Version 6.04A. Thus, the methodology provides an analysis of TCAS that is consistent with the design changes that have occurred in response to the situation that is considered.

With Chapter 4, this chapter serves to show the wide applicability of the methodology. Both large, immobile threats and small, moving threats have been examined. In addition, the methodology can be used to examine small-scale changes in the location of the alerting threshold and large-scale changes in overall system design. Thus, tradeoffs between a simple modification of the alerting threshold and more complex, larger system changes such as providing datalinked intent information can be examined together using the methodology.

6. Summary and Conclusions

Hazard alerting systems can be effective components of human-operated systems by providing protection against unforeseen threats. However, alerting system design must balance a tradeoff between increasing safety and reducing unwanted false alarms. This tradeoff depends on a number of issues that are generic across many specific applications. In complex alerting systems, the tradeoffs between system components and a means of evaluating system performance are often unclear. A single methodology that can be used to model and evaluate alerting systems in a variety of applications is needed to provide insight into the issues that affect alerting system performance.

This thesis develops a unified methodology that has two primary elements. First, the methodology can be used to develop a model of the different components that affect the design and performance of an alerting system. The areas where effort can be directed toward enhancing an alerting system can therefore be identified. Second, the methodology provides a means by which the performance of alerting systems can be defined and evaluated by showing that the alerting decision is essentially a signal detection problem. To demonstrate the utility of the methodology, it is applied to two example alerting systems.

6.1 Summary

6.1.1 Major System Components

A control system view of alerting systems is taken to provide a generalized means of defining the major components of an alerting system in a specific application. Four major components of the system are identified as important to alerting system design and performance:

- Measurement Sources determine what states of the situation are observable to the operator and to the alerting system. Measurement uncertainties also affect the information available to the alerting system.
- The Alerting System includes definitions of the alerting thresholds and the methods of transmitting alert information to the operator.
- The Human Operator's response to an alert depends on the information coming from the Measurement Sources and the Alerting System, as well as external

factors such as training or procedures. This response can be probabilistic in nature.

- The Situation Dynamics define the effects of the operator's response on the evolution of the situation. Uncertainties in the dynamic model are also considered.

6.1.2 State-Space Hazard Diagram

The State-Space Hazard Diagram is developed to illustrate the relationship between the alerting thresholds and the overall hazard situation. Three regions are defined that describe the nature of the space in which the system could operate:

- Hazard Space is the space in which an undesirable incident can occur.
- Maneuver Limit Space is the space in which the system will be unable to prevent entry into Hazard Space at a later time due to maneuverability constraints.
- Alert Space denotes the region in which alerts are issued from the alerting system.

The State-Space Hazard Diagram is used to identify those areas where the alerting system coverage is excessive or deficient. In addition, it illustrates tradeoffs in system performance that depend on the type and quality of the information available to the alerting system. Because of the generalized nature of Hazard Space, the methodology can also be used to examine the potential negative effects of an alert, such as inducing incidents that occur because an alert is issued.

6.1.3 Identification of Factors Affecting System Performance

The major factors that affect the performance of alerting systems are identified:

- Constraints on the types of actions that are used to react to an alert affect the performance of the alerting system. Limitations on these actions define the extent of Maneuver Limit Space and influence the coverage that is necessary from the alerting thresholds.
- States defining the hazard situation that are not observable by the alerting system result in decreased system performance. When unobservable states exist, the alerting system must make an alerting decision based on incomplete data.

- The means by which the future state trajectory of the system can be extrapolated affects the operation of the alerting system. Detailed knowledge of the future trajectory enhances the ability of the system to determine whether an alert is warranted.

- A means for incorporating uncertainties in the measurements, the dynamic model, and the model of the human operator's response is presented. The effects of these uncertainties can be evaluated by the methodology.

- Issues regarding multiple levels of alerts are identified. Low-urgency Caution alerts and high-urgency Warnings can be used to mitigate poor system performance due to uncertainties in the true situation.

6.1.4 Evaluation of Alerting System Performance

Once a model of the alerting system has been created, the methodology can be used to evaluate that system in various hazard situations. The issues involved in such an evaluation are also discussed.

The concept that two trajectories can be examined to determine the outcome of an alerting decision is fielded. This concept recasts the alerting decision as a signal detection problem, allowing the use of classical Signal Detection Theory methods to examine system performance.

- A Nominal Trajectory is used to examine the outcome of a decision not to alert.

- An Avoidance Trajectory is used to examine the outcome of a decision to issue an alert.

These trajectories can be probabilistic, and are used to examine the relative likelihood of several outcomes that are identified as part of this work:

- Missed Detection: An alert is not issued or is issued too late to avoid an incident.

- Correct Detection: An alert is issued and the hazard is avoided.

- False Alarm: An alert is issued, but action is not needed to avoid an incident.

- Correct Rejection: An alert is not issued, and an alert is not needed given the situation.

- Deferred Action: An alert is not issued and one may eventually be needed, but an alert at a later time will still enable the system to avoid an incident.

A procedure for calculating the probability that an incident will occur along the Nominal and Avoidance trajectories is developed. This procedure requires the equations describing the dynamics of the situation and probability density functions that describe the probabilistic nature of the Nominal and Avoidance trajectories. This methodology allows for the quantitative evaluation of the effectiveness of an alerting system in a given situation in terms of the probability of a false alarm or missed detection. The application of System Operating Characteristic (SOC) curves (analogous to Signal Detection Theory Receiver Operating Characteristic (ROC) curves) is also discussed.

6.1.5 Application of the Framework

The methodology developed in the thesis is applied to two different types of alerting systems. These examples serve to demonstrate the flexibility of the methodology and to highlight the potential similarities that can exist between applications.

6.1.5.1 Ground Proximity Warning System

The Ground Proximity Warning System (GPWS) currently used on civil jet transports is summarized and evaluated in two example situations. In one, an aircraft is assumed to be descending into a flat terrain field. The ability of the methodology to account for a probabilistic response from the pilot is demonstrated using example probability density functions, and the probability that an incident will occur is calculated. A safety level of approximately $1x10^{-7}$ is demonstrated for this situation.

Second, the performance of GPWS is evaluated with regard to a probabilistic terrain field. A Markov model of terrain is created based on actual terrain data and is used to find the probability that a false alarm or missed detection occurs. At a descent rate of 3,000 ft/min, for example, GPWS is shown to produce false alarms with probability $1x10^{-4}$ in Smooth terrain and with probability 0.08 in Steep terrain. There is also an increase in the probability of a missed detection, changing from $<1x10^{-16}$ for Smooth terrain to 0.03 for Steep terrain.

A comparison of the GPWS alerting thresholds with the probability of an incident is also provided for a range of descent rates. The probability of an incident in Smooth terrain

when a GPWS alert is issued is shown to be less than $1x10^{-8}$. In Steep terrain, this probability rises to greater than 0.01 for most descent rates.

6.1.5.2 Traffic Alert and Collision Avoidance System

The methodology is also demonstrated for a different application involving the Traffic Alert and Collision Avoidance System (TCAS). An example situation is considered in which an intruding aircraft is projected to descend dangerously near an aircraft equipped with TCAS. TCAS alerting thresholds are evaluated for the original (Version 1.0) thresholds and for an improved TCAS (Version 6.04A). The ability for the situation to produce false alarms with Version 1.0 is demonstrated, as is the ability of Version 6.04A to inhibit an alert in such a situation.

A quantitative comparison of the two threshold versions is performed. If the intruder continues its descent toward the TCAS aircraft, the probability of an incident when using Version 1.0 is $1x10^{-14}$. With Version 6.04A, this probability rises to $9x10^{-7}$. If the intruder levels off above the TCAS aircraft, however, Version 1.0 issues a false alarm, and the probability of an incident rises to $7x10^{-4}$. Version 6.04A does not issue a false alarm in this situation.

In a situation in which the intruder has a 0.75 probability of leveling off above the TCAS aircraft, the probability of a false alarm is shown to decrease from 0.63 for Version 1.0 to 0.44 for Version 6.04A. There is a corresponding decrease in the probability of an incident, going from $2x10^{-4}$ to $1x10^{-11}$ between Version 1.0 and 6.04A, respectively. The Version 6.04A alerting threshold is shown to be located such that it effectively minimizes the probability of a false alarm while maintaining a relatively high level of safety.

A potential improvement to TCAS that involves the ability to have access to the intended path of an intruding aircraft is also discussed. An enhanced TCAS of this sort is shown to not increase performance over that possible with TCAS Version 6.04A.

6.2 Conclusions

The primary contributions of this thesis are discussed below.

1. A generalized model of the major alerting system components has been developed. This model has wide applicability over a range of specific alerting systems. Therefore, applications outside aerospace or to large-scale vehicle systems are possible. The need to consider factors such as maneuver limitations, unobservable states, and the

state extrapolation method is also presented. Taking a probabilistic approach, the methodology incorporates uncertainties in measurements, modeling, and the human response to an alert.

2. A probabilistic analysis methodology has been developed to estimate the probability that an incident will occur in a given situation. This methodology accounts for uncertainties in measurements, the vehicle trajectory, and the size, shape, and severity of the hazard.

3. Alerting system performance metrics such as the probabilities of false alarm and missed detection are defined. The alerting decision is shown to correspond to a signal detection problem, allowing the use of classical Signal Detection Theory methods. The concept that the outcomes of an alerting decision are dependent upon two potential future trajectories (Nominal and Avoidance) is presented. These definitions are needed to provide common metrics against which different system design options can be compared. System Operating Characteristic (SOC) curves are defined to illustrate the tradeoff between false alarms and missed detections.

4. In an evaluation of the Ground Proximity Warning System (GPWS), the methodology is shown to provide estimates of the probability of an incident that are consistent with existing GPWS alerting thresholds. Constraints on the performance of GPWS due to the lack of terrain information ahead of the aircraft are demonstrated. In Steep terrain regions, GPWS performance is shown to be more limited than in Smooth terrain regions. Thus, there would be more benefit to providing increased terrain information in Steep regions than in Smooth regions.

5. In an evaluation of the Traffic Alert and Collision Avoidance System (TCAS), the benefit to system performance of the Version 6.04A modification is shown using the methodology developed in the thesis. The methodology predicts a benefit in system performance between Version 1.0 and 6.04A that is consistent with the changes in TCAS alerting thresholds. Using the methodology, a simple hypothesis test approach is demonstrated in an application of TCAS with intent information. Little reduction in the probability of false alarms is predicted if intent information is available to TCAS in the simplified situation that is studied.

6.3 Recommendations

This thesis has used a few illustrative situations to demonstrate the unified methodology. The methodology developed here can be applied more rigorously for a more complete, detailed examination of GPWS or TCAS. Parametric studies of the effect of changing different system parameters can be performed. For example, the effect on system performance of changing pilot reaction time can be compared with the effect of increasing sensor accuracy. If the costs of performing these changes can be quantified as well, the areas in which effort should be expended can be identified.

In addition, because of its generalized nature the methodology can be applied to other types of alerting systems. Advanced alerting systems are currently under consideration in the aerospace industry that would benefit from this thesis. Examples include an enhanced GPWS that uses a terrain database, a TCAS system that uses datalinked intent information, and alerting systems for simultaneous approach to parallel runways. Because of its ability to account for probabilistic hazards, the methodology would be especially useful in wake vortex alerting system applications. The methodology would also prove valuable to efforts in intelligent highway vehicles and large-scale traffic control systems.

References

American Airlines (AA). 1994. MD-11 operating manual, Vol 2. American Airlines flight department flight academy. Dallas, TX.

Barkat, Mourad. 1991. Signal detection and estimation. Artech House. Boston.

Bateman, Don. 1993. How to terrain-proof corporate and regional aircraft. The annual FSF European corporate and regional operations conference. Amsterdam. March 3.

Bateman, Don. 1994a. Ground Proximity Warning System (GPWS) - success and further progress. The international civil and military avionics conference. London. April 7.

Bateman, Don. 1994b. Development of Ground Proximity Warning Systems (GPWS). Royal Aeronautical Society controlled flight into terrain one day conference. London. November 8.

The Boeing Commercial Airplane Company. 1983. Boeing 767 operations manual. Boeing document No. D632T001-200. July 25.

Boucek, George. 1994. Personal communication. The Boeing Commercial Airplane Company. February 4.

Brown, Robert G., and Patrick Y.C. Hwang. 1992. Introduction to Random Signals and Applied Kalman Filtering. John Wiley & Sons. New York.

D'Azzo, John J., and Constantine H. Houpis. 1988. Linear Control System Analysis and Design. McGraw-Hill. New York.

DeCelles, J. L. 1991. The delayed GPWS response syndrome. Aviation Research & Education Foundation. Herndon, VA. July.

Federal Aviation Administration (FAA). 1994. U.S. Federal Aviation Regulations (FAR). Parts 25, 121, 135. Washington, D.C.

Gelb, Arthur (ed.). 1974. Applied optimal estimation. The Analytic Sciences Corporation. MIT Press. Cambridge, MA.

Hogg, Robert V., and Elliot A. Tanis. 1988. Probability and statistical inference. Macmillan. New York.

Hughes, David. 1994. Safety group highlights CFIT risk for regionals. Aviation Week & Space Technology. May 9.

Kemeny, J.G. and J.L. Snell. 1967. Finite Markov chains. D. Van Nostrand Co. Princeton, NJ.

Kirk, Donald E. 1970. Optimal control theory. Prentice Hall. Englewood Cliffs, NJ.

Kuchar, James K., and R.J. Hansman, Jr. 1993. Part-task simulation study of candidate terrain alerting displays. MIT Aeronautical Systems Laboratory. Report No. ASL-93-6. Cambridge, MA.

Lewis, Curt. 1994. Whoop, whoop, Pull Up! Flight Deck. American Airlines flight department flight academy. Dallas, TX. November / December.

Mellone, V.J. and Stephanie M. Frank. 1993. Behavioral impact of TCAS II on the national air traffic control system. Seventh international symposium on aviation psychology. The Ohio State University. April 27.

Mellone, V.J., 1993. TCAS II: Genie out of the bottle? Aviation Safety Reporting System (ASRS) Directline. No. 4.

Miller, C.A., Williamson, T., Walsh, J.A., Nivert, L.J., Anderson, J.L., and Andrew D. Zeitlin. 1994. Initiatives to improve TCAS-ATC compatibility. Journal of ATC. July-September.

The MITRE Corporation (MITRE). 1993. Final V6.04A change pages. Courtesy of Andrew D. Zeitlin. September 17.

National Transportation Safety Board (NTSB). 1975. Aircraft accident report - Trans World Airlines, Inc. Boeing 727-231, N54328, Berryville, Virginia, December 1, 1974. Report No. NTSB-AAR-75-16. Washington, D.C. November 26.

Niehaus, Axel and Robert F. Stengel. 1991. Rule-based guidance for vehicle highway driving in the presence of uncertainty. American Control Conference. Boston June.

Nordwall, Bruce. 1993. GPWS to improve regional safety. Aviation Week & Space Technology. April, 26.

Poole, Roy A. 1992. Whoop whoop pull up. Flight Safety. June.

Radio Technical Committee on Aeronautics (RTCA). 1976. Minimum performance standards - airborne ground proximity warning equipment. Document No. RTCA/DO-161A. Washington, D.C. May 27.

Radio Technical Committee on Aeronautics (RTCA). 1983. Minimum performance specifications for TCAS airborne equipment. Document No. RTCA/DO-185. Washington, D.C. September.

Rocky Mountain Communications, Inc. (RMC). 1991. CD-ROM #4, 3 arc second elevation data, 102 to 111 degrees west longitude. Golden, CO.

Sheridan, Thomas B. and William R. Ferrell. 1974. Man-machine systems: Information, control, and decision models of human performance. MIT Press. Cambridge, MA.

Sheridan, Thomas B. 1992. Telerobotics, automation, and human supervisory control. MIT Press. Cambridge, MA.

Shreider, Yuri A. (ed.). 1966. The Monte Carlo method. Pergamon. New York.

Swets, John A. and Ronald M. Pickett. 1982. Evaluation of diagnostic systems. Academic Press. New York.

Tantaratana, S. 1986. Sequential detection of a positive signal. In Blake, Ian F. and H. Vincent Poor (eds.). Communications and networks. Springer-Verlag. New York.

Weinstock, Herbert. 1963. The description of stationary random rate processes. MIT Instrumentation Laboratory Report E-1377. Cambridge, MA. July 1.

Wald, Abraham. 1957. Sequential analysis. John Wiley & Sons, New York.

Wickens, Christopher D. 1992. Engineering psychology and human performance. Harper Collins.

Wood, Loren. 1993. Lincoln Laboratory, Bedford, MA. Personal communication. November 1.

Appendix A.
Probabilistic Analysis Methodology

This appendix develops a methodology that can be used to estimate the probability that an incident with a hazard will occur for a given situation. The methodology is developed in a generalized manner to enable it to be applied to a wide variety of hazard situations. The examples used in this appendix consider an aircraft flying toward external, spatial hazards. However, the methodology is equally applicable to any hazard defined in state-space and need not be a spatial object.

Three definitions from Chapter 2 are repeated here because they are used extensively in this appendix:

An *incident*, I, is an event that the alerting system is designed to prevent.

Hazard Space is that subset of state-space in which an incident could occur.

A *hazard encounter*, E, is an event in which the system state vector $\mathbf{x}(t)$ lies within Hazard Space.

Examples of hazard encounters include the collision of an aircraft with terrain or the flight of an aircraft through a severe weather region. Depending on the specific type of hazard, an encounter may or may not actually result in the aircraft being adversely affected by the hazard. An event in which the aircraft is affected by the hazard is an incident. In many cases, a hazard encounter and incident occur together: an encounter with another aircraft means that an incident (a mid-air collision) has occurred. An encounter with some hazards, however, may not lead directly to an incident. For example, given that a hazard encounter has occurred in which the aircraft traveled through severe weather, there is some probability that an incident will not occur -- the aircraft may exit Hazard Space.

P(E) is defined as the probability that a hazard encounter occurs in a certain situation, and P(I) is defined as the probability that an incident occurs in that same situation. From Bayes' Rule,

$$P(I, E) = P(I \mid E)\, P(E) = P(E \mid I)\, P(I) \qquad (A.1)$$

By definition, however, P(E | I) = 1; an incident can never occur without an encounter also occurring. Thus,

$$P(I) = P(I \mid E)\, P(E) \tag{A.2}$$

Hazards for which P(I | E) = 1 are termed *hard hazards*. An encounter with a hard hazard means that an incident has also occurred. Hazards for which P(I | E) < 1 are termed *soft hazards* - an encounter does not necessarily require that an incident also occur. Since the ultimate purpose of an alerting system is to protect against incidents, P(I) will be an important metric of system effectiveness.

The methodology developed in this appendix is intended to provide a means by which P(E) and ultimately P(I) may be calculated. There are several principal areas where uncertainty may need to be accounted for in these calculations: the location of the hazard, the size, shape and severity of the hazard, and the estimated trajectory of the aircraft relative to the hazard. After an overview of the procedure that is used, the methodology is developed for a generalized situation. Chapters 4 and 5 then apply the methodology to specific situations.

A.1 Overview of the Procedure

The general problem statement at this point is shown in Figure A.1. A hazard of a certain size, shape, and severity is located in an uncertain position in space relative to the aircraft, and it is desired to know P(E) and P(I). The probability of encountering the hazard, P(E), is calculated by first defining the region in space (termed an *exclusion zone*) in which the hazard must actually be for an encounter to occur. The exclusion zone is a function of the sizes, shapes, and trajectories of the aircraft and the hazard.

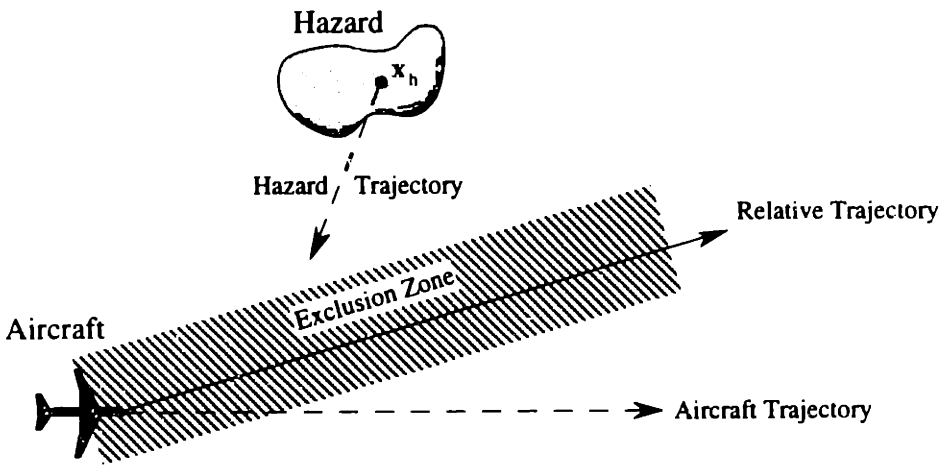


Figure A.1: Generalized Hazard Encounter

The probability density function describing the location of the hazard is then integrated over the exclusion zone to find P(E) for the given hazard. This procedure is performed over all possible trajectories and hazard shapes and sizes. Pr(I) may then be calculated using Equation A.2; P(I | E) depends on the type of hazard and may require additional calculations to determine the aircraft's exposure to the hazard.

A.2 Hazard Modeling

The first requirement for the methodology is a model of the hazard that is under study. While the size, shape, and severity of the hazard may be uncertain, a reference point in or near the location of the hazard, $\mathbf{x}_h = [x_h, y_h, z_h]^T$, must be defined (see Figure A.1).[1] This reference point can be located at any convenient position, but once chosen it must remain fixed relative to the hazard. Uncertainties in size and shape can then be described relative to $\mathbf{x}_h$. Following the method in Chapter 3, an *a priori* measurement of the location of $\mathbf{x}_h$ relative to the aircraft, denoted $\mathbf{x}_m$, is assumed. Given error distributions in the measurement of $\mathbf{x}_m$, the probability density function $f_{xyz}(\mathbf{x} - \mathbf{x}_m)$, is defined to describe the probability that the reference point, $\mathbf{x}_h$, is actually located at any arbitrary position $\mathbf{x} = [x, y, z]^T$.

When the location of the hazard is determined by an on-board sensor, the measured location of the hazard relative to the aircraft, $\mathbf{x}_m$, is obtained directly and $f_{xyz}(\mathbf{x} - \mathbf{x}_m)$ may generally be found from the errors in sensor estimates. In general, the aircraft will make an n-dimensional measurement of the hazard's location, $\mathbf{x}_m$. In many situations, $\mathbf{x}_m$ may be considered to be a normally distributed random variable with mean $\mathbf{x}$ and covariance matrix Σ, where:

$$\Sigma = E\left[(\mathbf{x} - \mathbf{x}_m)(\mathbf{x} - \mathbf{x}_m)^T\right] \tag{A.3}$$

and,

$$f_{xyz}(\mathbf{x} - \mathbf{x}_m) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_m)\Sigma^{-1}(\mathbf{x}-\mathbf{x}_m)^T\right]} \tag{A.4}$$

If, however, the location of the hazard and the aircraft are determined independently, $\mathbf{x}_m$ must be calculated as the vector difference of the locations of the hazard

---

[1] For simplicity, the time dependence of the vectors in this appendix is assumed implicitly, and $\mathbf{x}(t)$ is written as $\mathbf{x}$.

and the aircraft, as shown in Figure A.2. In the situation shown in the Figure, a ground station measures the location of the aircraft ($\mathbf{x}_{ga}$) and the hazard ($\mathbf{x}_{gh}$). The relative separation between the aircraft and the hazard is therefore given by $\mathbf{x}_m = \mathbf{x}_{gh} - \mathbf{x}_{ga}$. In such a case, $\mathbf{x}_m$ includes the errors in the estimates of both the aircraft and hazard location.

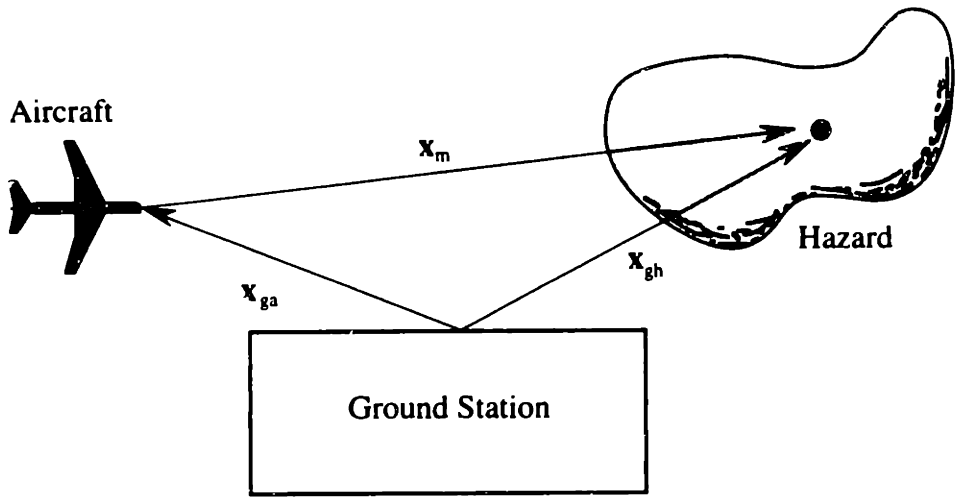


Figure A.2: Calculation of $\mathbf{x}_m$ From Two Measurements
$\mathbf{x}_m = \mathbf{x}_{gh} - \mathbf{x}_{ga}$

Once the hazard location is defined probabilistically with $f_{xyz}(\mathbf{x} - \mathbf{x}_m)$, the size and shape of the hazard need to be described. To begin, a set of possible hazard sizes and shapes is compiled, denoted as **S**. This set contains a finite or infinite number of elements, σ_i, each of which completely defines a particular hazard size and shape relative to $\mathbf{x}_h$ (see Figure A.3).

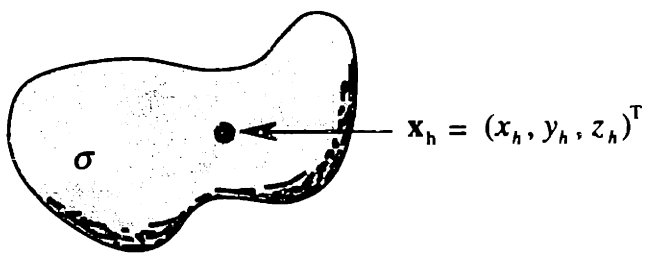


Figure A.3: Generic Hazard
Reference point: $\mathbf{x}_h$, Size and shape definition: σ

When **S** is composed of discrete σ_i, $P(\sigma_i)$ must be defined to describe the probability that each discrete event σ_i occurs. If **S** is a set of continuous values, $f_\sigma(\sigma)$ must instead be defined as the probability density function describing possible values of σ.

Note that each value of σ represents an independent, mutually exclusive event:

$$\begin{aligned} \sum_{\forall \sigma_i \in S} P(\sigma_i) &= 1 \qquad \text{(discrete } \sigma_i\text{)} \\ \int_S f_\sigma(\sigma)\, d\sigma &= 1 \qquad \text{(continuous } \sigma\text{)} \end{aligned} \tag{A.5}$$

As an example, consider a hazard that is modeled as a sphere with a probabilistic radius, r. Let $r = 0.4$ nmi with probability 0.25, 0.5 nmi with probability 0.50, or 0.6 nmi with probability 0.25. Each potential radius is considered to be a different discrete σ_i, with **S** defined as the set:

$$\mathbf{S} = \begin{Bmatrix} \sigma_1\text{: } r = 0.4 \text{ nmi, } P = 0.25 \\ \sigma_2\text{: } r = 0.5 \text{ nmi, } P = 0.50 \\ \sigma_3\text{: } r = 0.6 \text{ nmi, } P = 0.25 \end{Bmatrix} \tag{A.6}$$

Alternatively, the hazard's radius could be modeled as having a continuous distribution:

$$\mathbf{S} = \begin{Bmatrix} \sigma_0\text{: } r = 0.4 \text{ nmi, } f_\sigma(0.4) \\ \vdots \\ \sigma_\infty\text{: } r = 0.6 \text{ nmi, } f_\sigma(0.6) \end{Bmatrix} \tag{A.7}$$

where $f_\sigma(\sigma)$ is some probability density function (such as the one shown in Figure A.4).

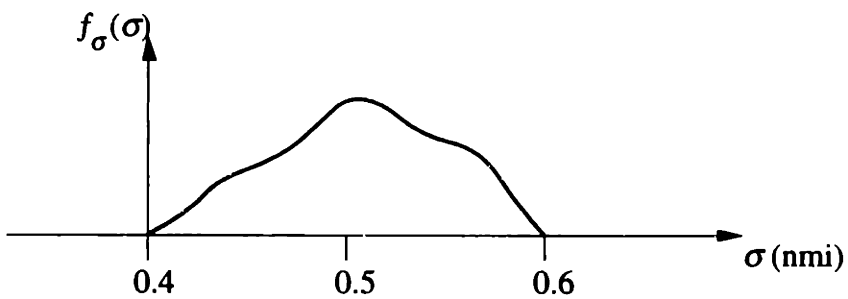


Figure A.4: Example Continuous Probability Density Function for Hazard Radius σ

If the hazard's size or shape changes with time, σ must include this time-dependent information also. In such cases, σ will, in general, describe a multi-dimensional volume in state-space and time.

A.3 Aircraft Modeling

The size and shape of the aircraft also must be described, using a parameter similar to the hazard definition σ. In the case of the aircraft, it can be assumed that the size and shape are known *a priori*, and do not change with time. The parameter ρ is the description of the extent of the aircraft: the space in which a hazard must be relative to the aircraft for an encounter to occur.

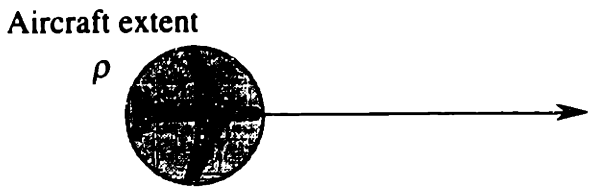


Figure A.5: Example Definition of Aircraft Extent, ρ

The parameter ρ can be defined as the specific problem dictates. For example, ρ might be defined as the smallest cylinder that contains the aircraft, in which case an encounter represents the event that the aircraft collides with a hazard. Alternatively, ρ might include a safety buffer around the aircraft, in which case an encounter will represent the event that a hazard intrudes into the safety buffer.

A.4 Trajectory Modeling

Next, a set of potential trajectories needs to be defined that describes the potential paths of the aircraft relative to the measured hazard location, $\mathbf{x}_m$ (shown in Figure A.6). In cases where the hazard is monitored from sensors on board the aircraft, the relative trajectory can be measured directly from the sensor outputs. When the aircraft and the hazard are tracked separately using different sensors, the random variables describing each trajectory must be combined (through a vector difference) to create a single random variable describing the relative trajectory.

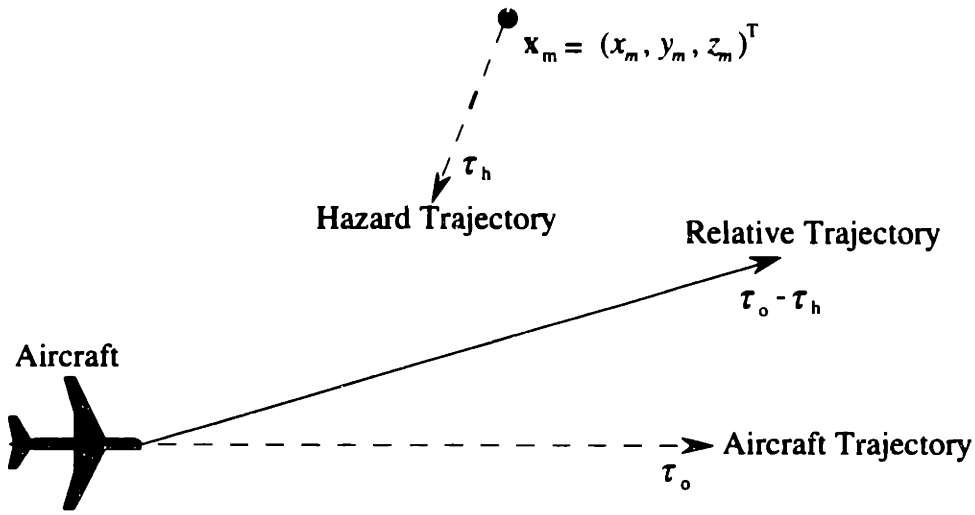


Figure A.6: Definition of Relative Trajectory

The entire set of possible relative trajectories will be denoted as **T**, with a particular trajectory defined as $\tau_i \in \mathbf{T}$. As with the hazard description set **S**, **T** can be a set of discrete trajectories or a continuous distribution of trajectories. Each trajectory is an independent, mutually exclusive event that has a defined probability of occurring, $P(\tau_i)$, for discrete τ_i, or is described with $f_\tau(\tau)$ for continuous τ:

$$\sum_{\forall \tau_i \in \mathbf{T}} P(\tau_i) = 1 \qquad \text{(discrete } \tau_i\text{)}$$

$$\int_{\mathbf{T}} f_\tau(\tau)\, d\tau = 1 \qquad \text{(continuous } \tau\text{)} \tag{A.8}$$

As with σ, τ is in general a multi-dimensional description of the trajectory in state-space and time. Aircraft flying the same path in space at different velocities will have different trajectories by this definition. As a simple example, consider a two-dimensional trajectory in space:

$$y(t) = x(t) \text{ where } x(t) = vt \tag{A.9}$$

Figure A.7 shows two trajectories (in three dimensions with time) for $v = 0.5$ and $v = 3.0$.

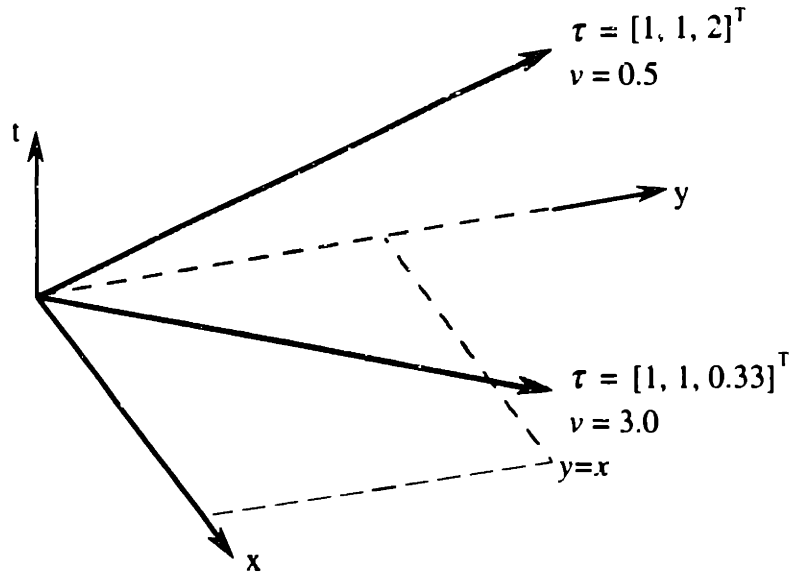


Figure A.7: Example Trajectories
$y(t) = x(t)$, $x(t) = vt$; $v = 0.5$ and $v = 3.0$.

A.5 Calculation of Probability of Encounter

A methodology is now derived that can be used to calculate the probability that the aircraft will encounter a hazard along a known trajectory τ in a situation with a given hazard size and shape σ and the aircraft extent ρ. This probability is denoted as $P(E \mid \rho, \sigma, \tau)$.[2] This methodology is subsequently expanded to account for uncertainties in σ or τ.

$\mathbf{R}(\rho, \tau)$ is defined as the space through which the aircraft will fly. $\mathbf{R}(\rho, \tau)$ is formed by extrapolating the aircraft extent definition ρ along the trajectory τ and is shown as the crosshatched area in Figure A.8.

---

[2] Note that this is actually $P(E \mid \mathbf{x}_m, \rho, \sigma, \tau)$, but in this and all future notation, $\mathbf{x}_m$ will be assumed known *a priori* and not explicitly mentioned.

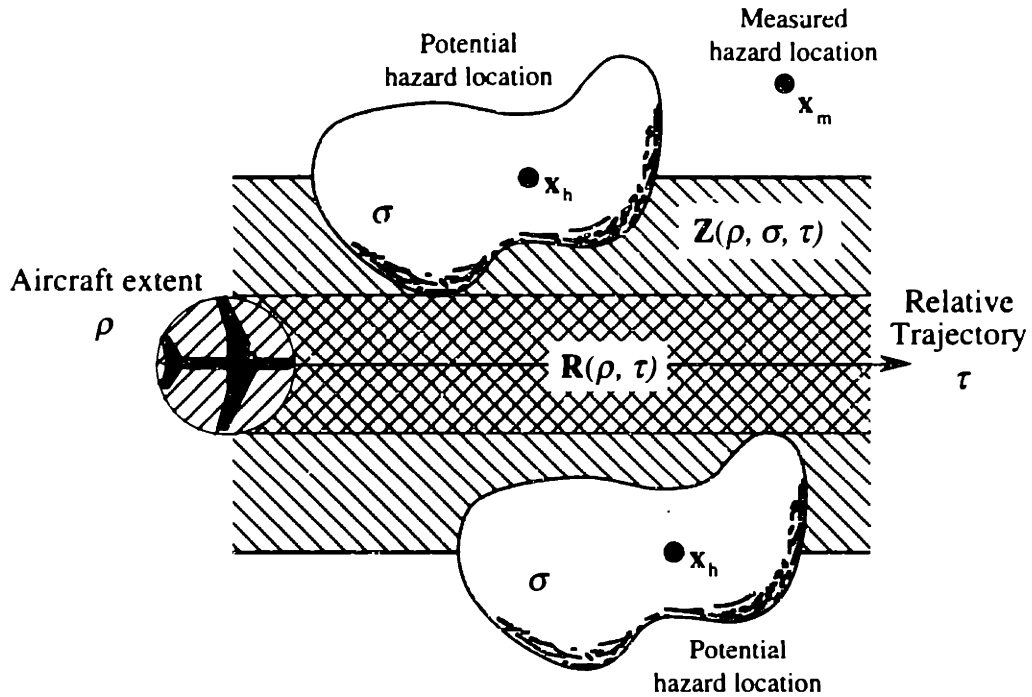


Figure A.8: Definition of Exclusion Zone, $\mathbf{Z}(\rho,\sigma,\tau)$, and Aircraft Zone, $\mathbf{R}(\rho,\tau)$

The exclusion zone, $\mathbf{Z}(\rho,\sigma,\tau)$, is defined as the space in which $\mathbf{x}_h$ must be to cause a conflict along τ. As shown in Figure A.8, $\mathbf{Z}(\rho,\sigma,\tau)$ is the union of all points $\mathbf{x}_h$ that cause the hazard to intrude into $\mathbf{R}(\rho,\tau)$. As defined, if $\mathbf{x}_h$ lies within $\mathbf{Z}(\rho,\sigma,\tau)$, the hazard will be encountered along τ.

The probability that $\mathbf{x}_h$ actually lies within $\mathbf{Z}(\rho,\sigma,\tau)$ (and therefore the probability that some part of the hazard intrudes into $\mathbf{R}(\rho,\tau)$) is then the integral of the hazard location density function over the space defined by $\mathbf{Z}(\rho,\sigma,\tau)$:

$$P(E \mid \rho,\sigma,\tau) = \int_{\mathbf{Z}(\rho,\sigma,\tau)} f_{xyz}(\mathbf{x}-\mathbf{x}_m)\, d\mathbf{x} \qquad (A.10)$$

A.5.1 Extensions to Uncertain Hazard State and Relative Trajectory

Equation (A.10) requires that ρ, σ, and τ are given *a priori*. As discussed earlier, ρ is considered to be given. In general, however, the size or shape of the hazard may not be known perfectly, and the aircraft's future trajectory relative to the hazard may likewise be uncertain. These uncertainties are now incorporated into the methodology.

With **S** and **T** defined as discussed in Sections A.2 and A.4, Equation (A.10) may be modified to take into account uncertainties in the hazard state or relative trajectory. Assume at first that σ is uncertain but that τ is given. Using the fact that the σ_i are mutually exclusive, exhaustive, and independent of τ (and dropping the known ρ for simplicity):

$$P(E \mid \tau) = P\big((E \mid \tau), \sigma_1\big) + P\big((E \mid \tau), \sigma_2\big) + \ldots + P\big((E \mid \tau), \sigma_n\big) \qquad (A.11)$$

From Bayes' Rule, this can be rewritten as:

$$P(E \mid \tau) = P(E \mid \sigma_1, \tau)P(\sigma_1) + P(E \mid \sigma_2, \tau)P(\sigma_2) + \ldots + P(E \mid \sigma_n, \tau)P(\sigma_n) \qquad (A.12)$$

Thus, for discrete σ_i:

$$P(E \mid \tau) = \sum_{\sigma_i \in \mathbf{S}} P(E \mid \sigma_i, \tau)P(\sigma_i) \qquad (A.13)$$

or, for continuous σ:

$$P(E \mid \tau) = \int_{\mathbf{S}} P(E \mid \sigma, \tau)\, f_\sigma(\sigma)\, d\sigma \qquad (A.14)$$

Thus, if given a trajectory τ, Equation (A.13) or (A.14) can be used to calculate P(E) along that trajectory given uncertainties in the size and shape of the hazard.

The same method is used to solve for P(E) when τ is also uncertain:

$$P(E) = P\big(E, \tau_1\big) + P\big(E, \tau_2\big) + \ldots + P\big(E, \tau_n\big) \qquad (A.15)$$

and again, by Bayes' Rule,

$$P(E) = P(E \mid \tau_1)P(\tau_1) + P(E \mid \tau_2)P(\tau_2) + \ldots + P(E \mid \tau_n)P(\tau_n) \qquad (A.16)$$

If the hazard size and shape are also uncertain, each $P(E \mid \tau_i)$ in Equation (A.16) is found by performing the calculation in Equations (A.13) or (A.14).

Thus, for discrete σ_j and τ_i:

$$P(E) = \sum_{\tau_i \in \mathbf{T}} P(E \mid \tau_i)P(\tau_i) = \sum_{\tau_i \in \mathbf{T}} \sum_{\sigma_j \in \mathbf{S}} P(E \mid \sigma_j, \tau_i)\, P(\sigma_j)\, P(\tau_i) \qquad (A.17)$$

Or, for continuous σ and τ:

$$P(E) = \int_T P(E \mid \tau)\, f_\tau(\tau)\, d\tau = \int_T \int_S P(E \mid \sigma, \tau)\, f_\sigma(\sigma)\, f_\tau(\tau)\, d\sigma\, d\tau \qquad (A.18)$$

It is also possible to have, for example, discrete σ_j but continuous τ, in which case the appropriate combinations of summation or integration are used. For the remainder of this appendix, it is assumed that all uncertain variables are continuous, though similar steps can be performed for discrete variables.

Recall that Equation (A.10) allowed the calculation of $P(E \mid \sigma, \tau)$ for a known σ and τ. Equation (A.10) can then be substituted into Equation (A.18) to obtain an expression for P(E) given uncertain trajectories and hazard size and shape:[3]

$$P(E) = \int_T \int_S \int_{Z(\rho,\sigma,\tau)} f_{x_m z}(\mathbf{x} - \mathbf{x}_m)\; f_\sigma(\sigma)\, f_\tau(\tau)\, d\mathbf{x}\; d\sigma\, d\tau \qquad (A.19)$$

A.6 Calculation of the Probability of an Incident

In some situations, such as flight near hazardous weather, an encounter with a hazard does not necessarily mean that an incident also occurs. In addition, the probability of an incident posed by an encounter with a hazard may not be known perfectly. The methodology developed so far is now extended to cases where even if an encounter were to occur, there is some probability that an incident will not occur. As discussed earlier, such a hazard is termed a *soft hazard* to distinguish it from *hard hazards* such as traffic or terrain where an encounter is equivalent to an incident.

There are two major types of soft hazard: those in which the probability of an incident depends on the time or distance over which an encounter occurs (termed an *exposure-dependent hazard*), and those in which the probability of an incident is independent of exposure (termed an *exposure-independent hazard*). An example of the former type is hazardous weather - the longer an aircraft remains inside a thunderstorm, the greater the probability that an incident has occurred. An example of an exposure-independent hazard is a missile site that has a certain probability of being active on a particular day, regardless of how long the aircraft flies near it.

---

[3] For completeness, note that P(E) here is actually $P(E \mid \rho, \mathbf{x}_m)$.

Equation (A.19) gives the probability that a hazard will be encountered inside the exclusion zone $\mathbf{Z}(\rho, \sigma, \tau)$. When soft hazard severity is added, an additional term, $P(I \mid E)$, must be included as a measure of the probability of an incident given that exposure to a hazard has occurred. For example, if $P(I \mid E) = 0.5$, then there is a 0.5 probability that an incident will occur if an encounter occurs. In cases where an encounter with a threat is the same as an incident (hard hazards), $P(I \mid E) = 1$ and equation (A.19) gives the probability of such an incident occurring:

P(I) for hard hazard

$$P(I) = P(I \mid E)P(E) = P(E) \tag{A.20}$$

When an encounter with a hazard does not necessarily mean that an incident will occur (a soft hazard), the methodology must be further modified to account for $P(I \mid E) < 1$.

A.6.1 Exposure-Independent Hazards

When a soft exposure-independent hazard is encountered, $P(I \mid E)$ is less than 1. Assuming that this probability is constant over the entire hazard, P(I) is given by:

P(I) for soft, exposure-independent hazard

$$P(I) = P(I \mid E)P(E) \tag{A.21}$$

A.6.2 Exposure-Dependent Hazards

When dealing with exposure-dependent hazards, $P(I \mid E)$ depends on the amount of time or distance that the trajectory remains in Hazard Space. The total exposure to the hazard along the flight path is needed to calculate P(I). To begin, some estimate of the severity per unit of exposure to the hazard is needed. For this purpose, the parameter Θ is defined to represent an estimate of the mean amount of exposure (in units of distance or time) until an incident occurs.

For example, a severe weather hazard that is modeled as having a mean distance to an incident of $\Theta = 10$ nmi implies that, on average, a hazard encounter will result in an incident after flying 10 nmi through the hazard. Alternatively, the weather can be modeled as having a mean time to an incident of $\Theta = 10$ minutes. As Θ tends toward zero, the hazard becomes more like a hard hazard -- very small exposures will result in incidents. A very large value of Θ indicates an insubstantial hazard to which a large exposure is required

before an incident will likely occur. Θ is assumed to be constant over the extent of the hazard, though in general this does not have to be so.

With Θ defined as above, P(I | E) is modeled using an exponential distribution in a manner similar to that used for component failure:

$$P(I \mid E) = \int_{\tau} \frac{1}{\Theta} e^{-s/\Theta} \, ds \tag{A.22}$$

Note that in Equation (A.22), the integration proceeds over the curvilinear path defined by τ. The location of the aircraft along this trajectory is given by $\mathbf{x}(s)$. However, as shown in Figure A.9, because the aircraft has some spatial extent, $\rho(s)$, exposure to the hazard may occur before $\mathbf{x}(s)$ itself enters hazard space.

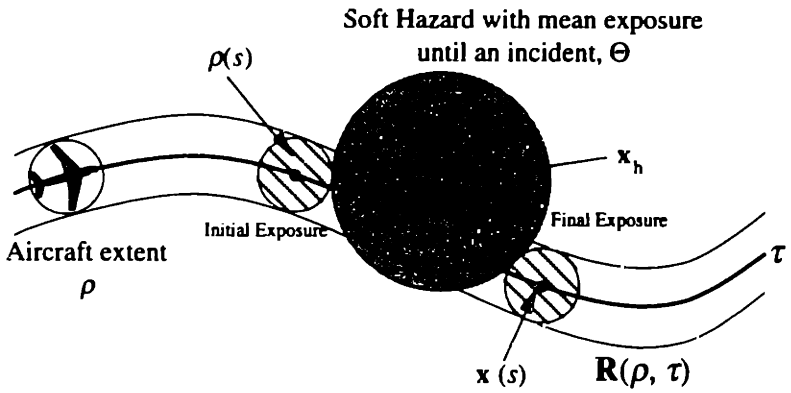


Figure A.9: Soft, Exposure-Dependent Hazard

The set of points at which exposure to the hazard occurs, **E**, is therefore defined as:

E is the set of $\mathbf{x}(s) \in \tau$ for which there exists an $\mathbf{x}_i \in \rho(s)$ such that $\mathbf{x} \in \mathbf{HS}$ (A.23)

where **HS** is the Hazard Space defined by the extent of the hazard.

The final equation for P(I) for exposure-dependent hazards is then written:

$$P(I) = \int_{T} \int_{S} \int_{Z(\rho,\sigma,\tau)} f_{xyz}(\mathbf{x} - \mathbf{x}_m) \, f_{\sigma}(\sigma) \; f_{\tau}(\tau) \int_{E} \frac{1}{\Theta(\mathbf{x}(s))} e^{-s/\Theta(\mathbf{x}(s))} \, ds \, d\mathbf{x} \, d\sigma \, d\tau \tag{A.24}$$

A.7 Alternative Solution Methods

The methodology presented above is by no means the only way to calculate P(E) or P(I). A few additional, alternative methods are briefly outlined as an indication of other possibilities that might be well suited to certain types of problems.

A.7.1 Calculation by Conditional Probabilities

It may appear that a straightforward method to find $P(E \mid \tau)$ would be to break the trajectory into a number of points and find $P(E \mid \mathbf{x}_i)$ at each point $\mathbf{x}_i$ along the trajectory (see Figure A.10). By combining these probabilities in some manner, it is possible to calculate $P(E \mid \tau)$. However, this method becomes difficult in all but the most simple situations for two reasons. First, the event of encountering a hazard at two different points is in general not mutually exclusive given a finite-sized hazard. Therefore, the values of $P(E \mid \mathbf{x}_i)$ cannot be added together to find $P(E \mid \tau)$. Second, the event of encountering a hazard at one point is conditional on the event of encountering the hazard at a previous location - the events are not independent. This method therefore requires the use of unwieldy conditional probabilities to calculate $P(E \mid \tau)$ along the trajectory.

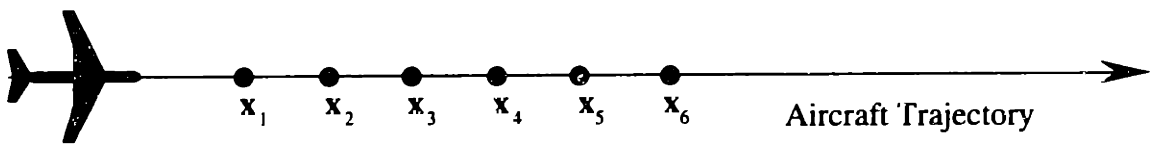


Figure A.10: Breaking Trajectory into Set of Discrete Points

With the shorthand $P(\bar{\mathbf{x}}_i)$ defined to be $P(\bar{E} \mid \mathbf{x}_i)$, i.e., the probability that a hazard does not exist at point $\mathbf{x}_i$:

$$P(E|\tau) = 1 - P(\bar{E}|\tau) = 1 - P(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \ldots)$$

$$= 1 - P(\bar{\mathbf{x}}_1)P(\bar{\mathbf{x}}_2|\bar{\mathbf{x}}_1)P(\bar{\mathbf{x}}_3|\bar{\mathbf{x}}_2, \bar{\mathbf{x}}_1)P(\bar{\mathbf{x}}_4|\bar{\mathbf{x}}_3, \bar{\mathbf{x}}_2, \bar{\mathbf{x}}_1)\ldots \tag{A.25}$$

This method may provide an accurate result, but the necessity of carrying along conditional probabilities requires too many computations to be of practical use in most situations.

A.7.2 Monte Carlo Methods

Monte Carlo methods may be applied by generating a large number of random events and recording that fraction of the events that lead to hazard encounters or incidents. The principal drawback of this method is the fact that very large sample sizes may be required in order to produce accurate results.

When a Monte Carlo simulation is used, the goal is to estimate a probability p (in this case, p would be P(E) or P(I)). At the end of the simulation, there will be S successes out of N trials. The fraction $\hat{p} = \frac{S}{N}$ is an estimator of p that follows a Bernoulli distribution with mean $\mu = p$ and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{N}}$ (Shreider 1966).

As p gets small, more evaluations are necessary to estimate p accurately. For example, for small p, $p(1 - p)$ approaches p. The standard deviation of the estimate is then approximately $\sqrt{\frac{p}{N}}$. If it is desired to obtain a certain accuracy in the estimate of p relative to the mean, $\varepsilon = \sigma/\mu$, the requirement on the number of trials needed is $N = 1/(\varepsilon^2 p)$.

As an example, assume it is desired to estimate p to an accuracy of $2\sigma = 0.01p$ (i.e., there is a 95% probability that $\hat{p}$ is within $0.01p$ of the true value). If $p = 10^{-5}$, at least 4×10^9 evaluations are needed.

Given that P(E) or P(I) may be small, the Monte Carlo method may require too many evaluations to be of practical use in some situations.

A.7.3 Estimation Using $P(E \mid x_i)_{max}$

An approximate method for estimating $P(E \mid \tau)$ has been used by Niehaus and Stengel (1991), for example. In their work on intelligent highway vehicle trajectory planning, Niehaus and Stengel needed an estimate of $P(E \mid \tau)$ along specific trajectories to choose a hazard-minimizing trajectory. Instead of calculating the actual value of $P(E \mid \tau)$ along an entire trajectory, they estimated its value by instead taking the maximum $P(E \mid \mathbf{x}_i)$ of all points $\mathbf{x}_i$ along the trajectory. This method works well for situations in which it is desired to compare the probability of an encounter along similar trajectories and for cases in which it is evident where the highest probability of an encounter is located along a trajectory. Some situations, however, may be complex and require a more detailed evaluation of P(E) along the trajectory.

Appendix B.
Signal Detection Theory Concepts

This appendix provides a brief summary of the principle concepts in Signal Detection Theory (SDT). More detailed descriptions of SDT can be found in the literature, such as Sheridan and Ferrell (1974), Swets and Pickett (1982), and Sheridan (1992).

The basic problem in SDT is one in which it is desired to decide whether a signal is present based on a measurement that contains background noise. As Figure B.1 shows, there is a certain probability density function (PDF) that describes the probability that a given measured value will be obtained when noise only is present. A second PDF describes the probability that a given measurement will occur when the signal is present. If the measurement is greater than some threshold value, the signal is said to be present.

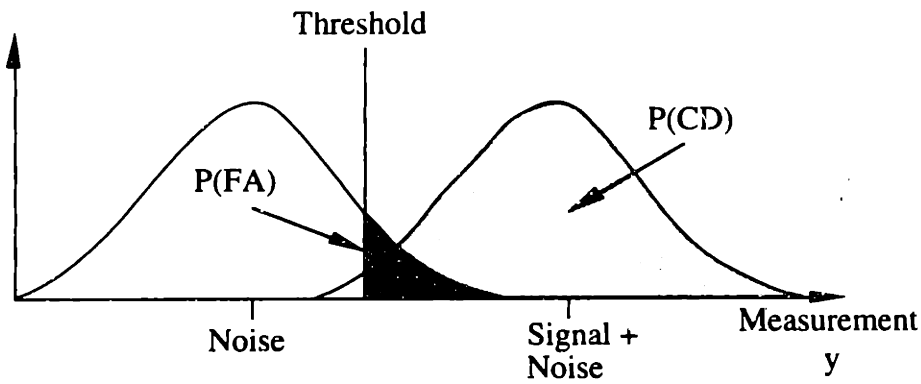


Figure B.1: PDFs of Noise and Signal + Noise

Due to the uncertainties in the problem, however, there are four possible outcomes of the signal detection problem, as shown in Table B.1.

Table B.1
Signal Detection Outcomes

Decision	Truly Noise	Truly Signal + Noise
Noise only	Correct Rejection (CR)	Missed Detection (MD)
Signal + Noise	False Alarm (FA)	Correct Detection (CD)

For example, a false alarm occurs when the signal is said to be present when in fact it is not. A missed detection occurs when the signal is present but a decision is made that the signal is not present. The probability that a false alarm or correct detection occurs (denoted P(FA) and P(CD), respectively) is given by the integral of the areas under the PDFs to the right of the threshold as shown in Figure B.1. In addition, because the integrals of each PDF over all measurements must sum to 1, the following two relations hold:

$$\begin{aligned} P(FA) + P(CR) &= 1 \\ P(MD) + P(CD) &= 1 \end{aligned} \qquad \text{(B.1)}$$

An alternative method for viewing the SDT problem is to plot the cumulative probability distributions for the two PDFs shown in Figure B.1. As shown in Figure B.2, the probability of a false alarm is then given by the value above the noise distribution (i.e., the integral of the noise PDF to the right of the threshold), and the probability of a missed detection is given by the value below the signal + noise distribution.

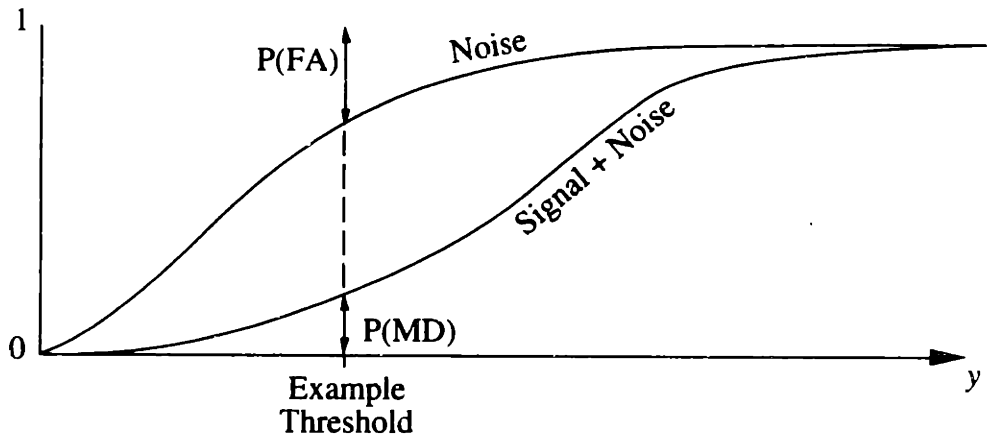


Figure B.2: Cumulative Probability Distributions of Noise and Signal + Noise

As the decision threshold is moved, false alarms are traded off against missed detections. For the example threshold location in Figure B.2, a higher probability of false alarm than a probability of missed detection is to be expected.

The tradeoff between false alarms and missed detections can be seen more directly through a Receiver Operating Characteristic (ROC) curve. ROC curves show a plot of the probability of a correct detection, P(CD), against the probability of false alarm, P(FA). An example is shown in Figure B.3.

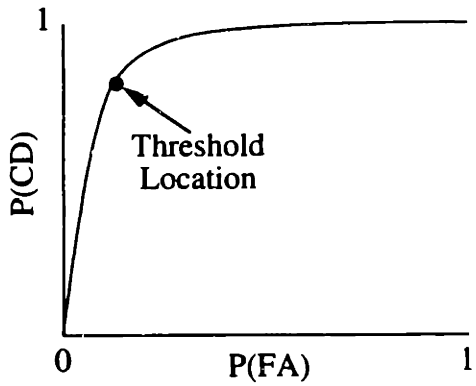


Figure B.3: Example ROC Curve

A particular set of PDFs for the noise and signal + noise cases results in a single ROC curve. The location of the decision threshold maps onto a single point along this ROC curve. Thus, the tradeoff between false alarms and missed detections can be shown. For example, in Figure B.3, any changes in the threshold to reduce the probability of missed detection below the value shown will result in a rapid increase in the probability of false alarm.

Appendix C.
GPWS Maneuver Model and Terrain Modeling

C.1 Derivation of Minimum Altitude During Pull-Up

For the calculations in Chapter 3, it is desired to know the minimum altitude attained by an aircraft that is performing a constant acceleration pull-up maneuver. Figure C.1 shows the parameters involved in this calculation.

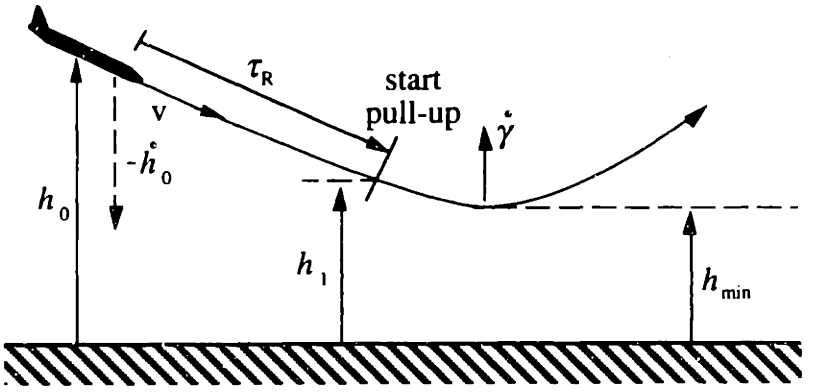


Figure C.1: Pull Up Maneuver

The basic defining parameters of the problem are: initial altitude (h_0), initial altitude rate ($\dot{h}_0$), the time delay until the pitching maneuver begins (τ_R), the flight path angle rate during the pull-up ($\dot{\gamma}$), and the aircraft's velocity, v.

The altitude at the start of the pull-up maneuver is:

$$h_1 = h_0 + \dot{h}_0 \tau_R \tag{C.1}$$

During the pull-up maneuver, it is assumed that $\dot{\gamma}$ and v are constant. The descent rate during the pull-up is then given by:

$$\dot{h}(t) = \mathrm{v}\sin(\gamma_0 + \dot{\gamma}\, t) \tag{C.2}$$

where the initial flight path angle $\gamma_0 = \sin^{-1}(\dot{h}_0 / \mathrm{v})$ and $t = 0$ is the time at the start of the pull-up maneuver.

The amount of time taken in the pull-up maneuver to arrest the descent rate is $\gamma_0 / \dot{\gamma}$. Therefore, the minimum altitude attained during the pull up maneuver is:

$$h_{min} = h_1 - \int_{t=0}^{t=\gamma_0/\dot{\gamma}} v\sin(\gamma_0 + \dot{\gamma}t)dt = h_1 - \frac{v}{\dot{\gamma}} + \frac{v}{\dot{\gamma}}\cos(\gamma_0) \tag{C.3}$$

noting that $\cos(\gamma_0) = \frac{\sqrt{v^2 - \dot{h}_0^2}}{v}$, and substituting Equation (C.1) for h_1,

$$h_{min} = h_0 + \dot{h}_0\tau_R - \frac{v}{\dot{\gamma}} + \frac{\sqrt{v^2 - \dot{h}_0^2}}{\dot{\gamma}} \tag{C.4}$$

which is the form used in Chapter 3.

C.2 Statistical Terrain Model

Several potential methods may be applied when creating a probabilistic model of terrain. In this analysis, terrain is treated as a stochastic first-order Markov process for several reasons. First, Markov models are often used to describe natural processes (Gelb 1974); they therefore are likely to also describe the statistics of terrain. Second, Markov processes are simple to define and manipulate using probability theory. The calculations that must be performed on the probabilistic model of terrain are extensive, so a simple model is desirable.

C.2.1 First-Order Gauss-Markov Process Properties

An important function describing the statistical properties of a stochastic signal is the autocorrelation function (Gelb 1974, Brown & Hwang 1992). This function describes the statistical correlation between two samples of a signal taken a lag time τ apart. The autocorrelation function, $\phi_{yy}(\tau)$, for a continuous-time, stationary, ergodic function $y(t)$ is defined by Equation (C.5):

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty} y(t)y(t+\tau)\,d\tau \tag{C.5}$$

For a first-order Gauss-Markov process, Equation (C.5) reduces to an even exponential function:

$$\phi_{yy}(\tau) = \sigma^2 e^{-|\tau|/\tau_0} \tag{C.6}$$

That is, the correlation between two samples of the process decreases exponentially as the lag time between samples increases. The parameters σ^2 and τ_0 reflect the variance and length scale, respectively, of the fluctuations in the process. The correlation time, τ_0, is defined as the lag time at which the autocorrelation function reaches $1/e$ of its peak value. Equation (C.6) is sometimes written instead in terms of $\beta = 1/\tau_0$.

Consider now a discrete-time Markov process that has taken values $y_0, y_1, y_2, \ldots, y_{n-1}, y_n$ up to the present time, n. A convenient property of first-order Markov processes is that the probability density function of the next value in the sequence, y_{n+1}, based on the previous values of the signal depends only upon the most recent value, y_n. In terms of probability density functions,

$$f(y_{n+1} \mid y_0, y_1, \ldots, y_{n-1}, y_n) = f(y_{n+1} \mid y_n) \tag{C.7}$$

Equation (C.7) implies, then, that a representative signal can be constructed without carrying along a complete history of all previous values in the signal. In fact, a discrete-time Markov process can be generated by the following system equation (Brown & Hwang 1992):

$$y_{n+1} = e^{-\beta} y_n + \xi_n \tag{C.8}$$

This process is termed a Gauss-Markov process when ξ_n is a zero-mean normally-distributed random variable with variance $\sigma^2(1-e^{-2\beta})$:

$$\xi_n \sim N\left(0, \sigma^2(1-e^{-2\beta})\right) \tag{C.9}$$

C.2.2 Empirical Generation of Autocorrelation Function

The first step in obtaining a Markov model involves presenting terrain altitude as a function of distance ahead of the own aircraft. A terrain altitude profile may then be thought of as a stochastic process. Although terrain is a spatial process, it can also be considered as a temporal process related by the aircraft's ground speed: $t = x / v$.

A large ensemble of actual terrain data can be used to build the statistics of this process. Accordingly, terrain profile data was obtained from a CD-ROM database covering the U.S. Great Plains and Rocky Mountain region, from 102° - 112° West longitude (RMC 1991). This database provided one meter altitude accuracy with 3 arc second

(approximately 300 ft) horizontal resolution. Examples of two true terrain profiles are shown in Figure C.2.

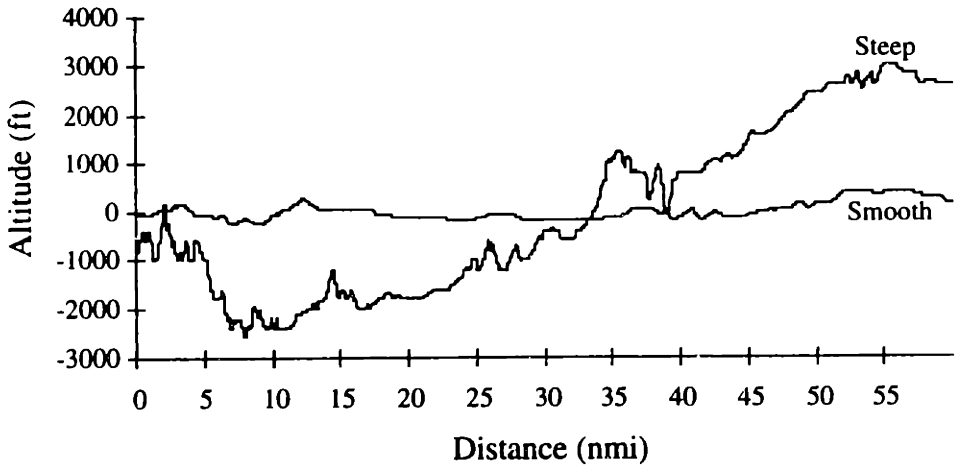


Figure C.2: Example True Terrain Samples
Samples shifted to zero-mean

Because the statistics of terrain in a flat region are not the same as those for a mountainous region, the terrain profiles were separated into five categories based upon the variance of terrain altitudes in the sample, shown in Table C.1. Given the altitude variances shown in the Table, 99% (3σ) confidence intervals for terrain altitude range in each sample may be approximated, and are also shown.

Table C.1
Terrain Categories

Terrain Category	Sample Altitude Variance (m^2)	99% Confidence Interval Altitude Span over 60 nmi	Number of Samples
Smooth	< 7,000	< 500 m	2,651
Moderately Smooth	7,000 - 27,778	500 - 1,000 m	1,996
Moderate	27,778 - 62,500	1,000 - 1,500 m	1,432
Moderately Steep	62,500 - 111,111	1,500 - 2,000 m	1,018
Steep	> 111,111	> 2,000 m	1,353

A total of 8,450 terrain samples were taken from the database, from latitudes between 32° - 49° North and longitudes between 102° - 112° West. Each sample was a North-South profile 1° long in latitude (60 nmi). Separate samples were taken every 72 arc seconds in longitude (approximately 1 nmi), as shown in Figure C.3. Because each sample was 1° (3,600 arc sec) in length with data point spacing every 3 arc seconds, 1,200 data points were available per terrain profile.

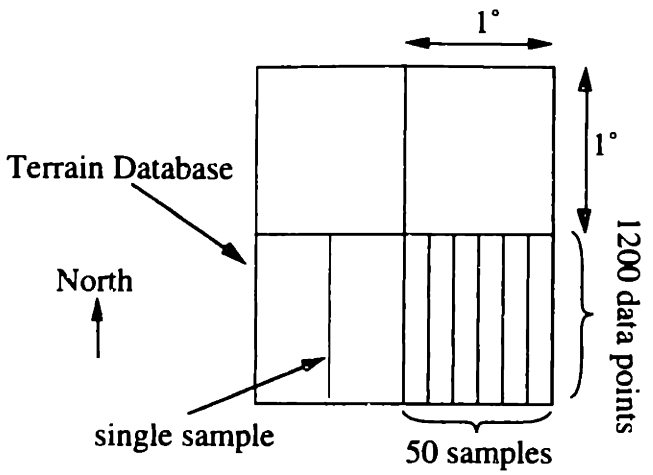


Figure C.3: Terrain Sample Selection
Fifty 1°-long samples taken per 1°x1° block

These samples were classified into one of the five terrain categories and the autocorrelation function, $\phi_{yy}(\tau)$, of the signal was calculated using Equation (C.10), as shown in Figure C.4. The values of $\phi_{yy}(\tau)$ were then averaged across the set of samples in the altitude category to arrive at an estimated autocorrelation function.

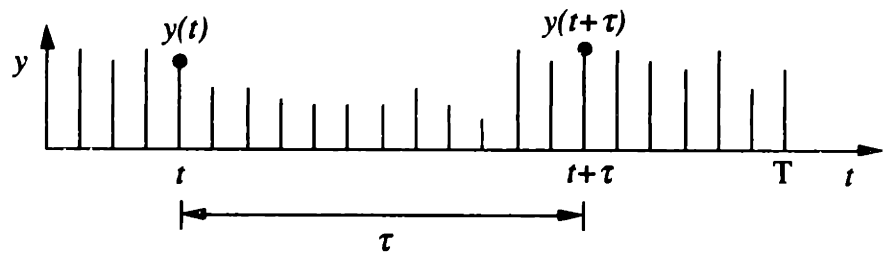


Figure C.4: Calculation of Autocorrelation Function
Lag time τ, record length T

$$\phi_{yy}(\tau) = \frac{1}{T-\tau}\sum_{t=0}^{T-\tau} y(t)y(t+\tau) \qquad \text{(C.10)}$$

The autocorrelation function of an infinite-length single-order Markov process is exponential, as was shown in Equation (C.6). However, when a finite-length sample is taken from a Markov process, the resulting empirically-determined autocorrelation function may be quite different (Weinstock 1963, Gelb 1974). The expected value of $\phi_{yy}(\tau)$ for a finite-length record is given by Equation (C.11). As shown in Figure C.5, as the record length increases, the empirical autocorrelation function more closely approaches the true exponential function.

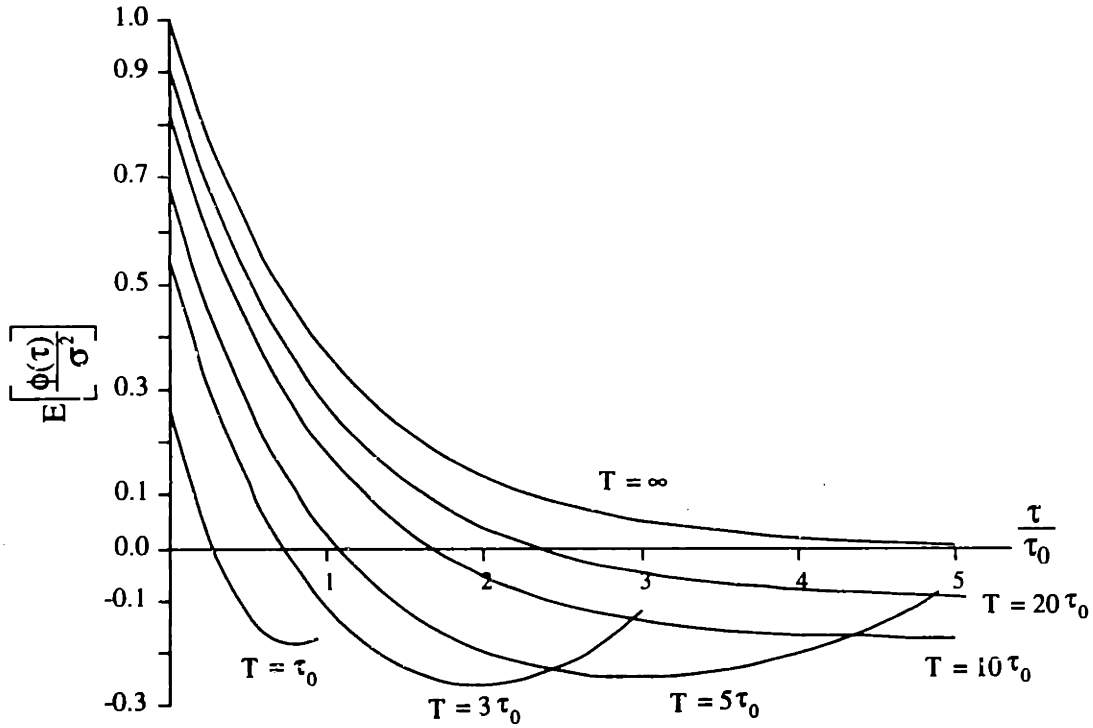


Figure C.5: Expected Value of Autocorrelation Function
T: Record length, τ_0: Correlation time

$$E\left[\phi_{yy}(\tau)\right] = \sigma^2 e^{-\tau/\tau_0} + \frac{2\sigma^2\tau_0^2}{T^2}\left(\frac{T}{\tau_0} - (1 - e^{-T/\tau_0})\right) - \frac{4\sigma^2\tau_0^2}{T(T-\tau)}\left(\frac{T-\tau}{\tau_0} - (1 - e^{-\frac{(T-\tau)}{\tau_0}})\right) - \frac{2\sigma^2\tau_0^2}{T(T-\tau)}\left(1 - e^{-\tau/\tau_0} + e^{-T/\tau_0} - e^{-\frac{(T-\tau)}{\tau_0}}\right) \quad \text{(C.11)}$$

(From Weinstock 1963)

The empirical data can be corrected, however, to obtain an estimate of the actual autocorrelation function. An example empirical autocorrelation function and a fitted Markov function is shown for the Moderate Terrain category in Figure C.6. Equation (C.11) was solved for τ_0 such that the zero-crossing point ($\phi_{yy}(\tau)$=0) occurred at the estimated value of τ given the record length used (T = 1,200). Then, σ was chosen such that the expected value of $\phi_{yy}(0)$ for the theoretical Markov process equaled that obtained empirically.

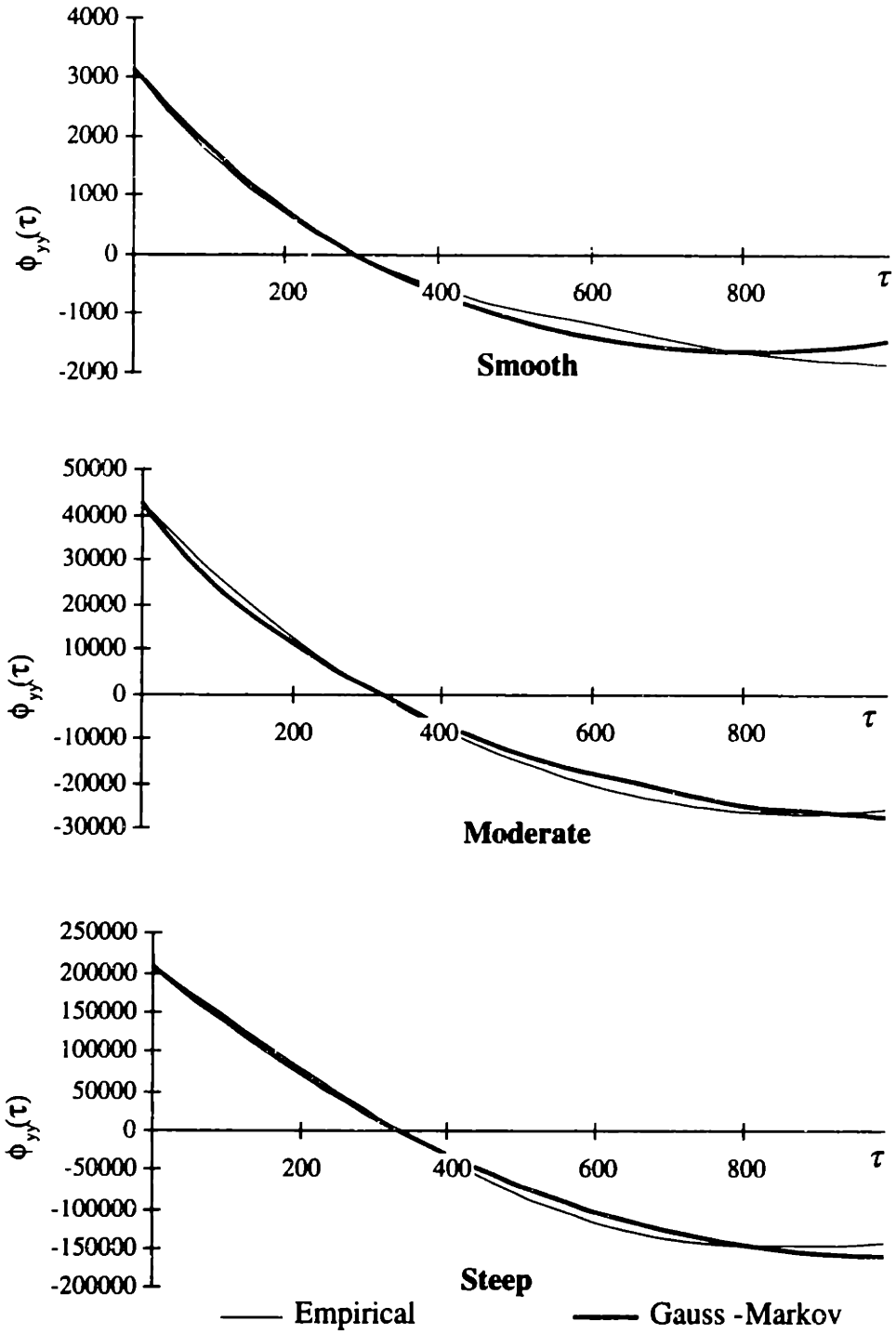


Figure C.6: Empirical and Theoretical Autocorrelation Functions
Record Length: T = 1,200

Table C.2 shows the function parameters for the fitted autocorrelation function for each terrain category. Note that the values for σ^2 increase as the terrain categories change from Smooth to Steep. However, τ_0 does not show much correlation with the terrain category. Because the categories were defined based on variance, it is expected that σ^2

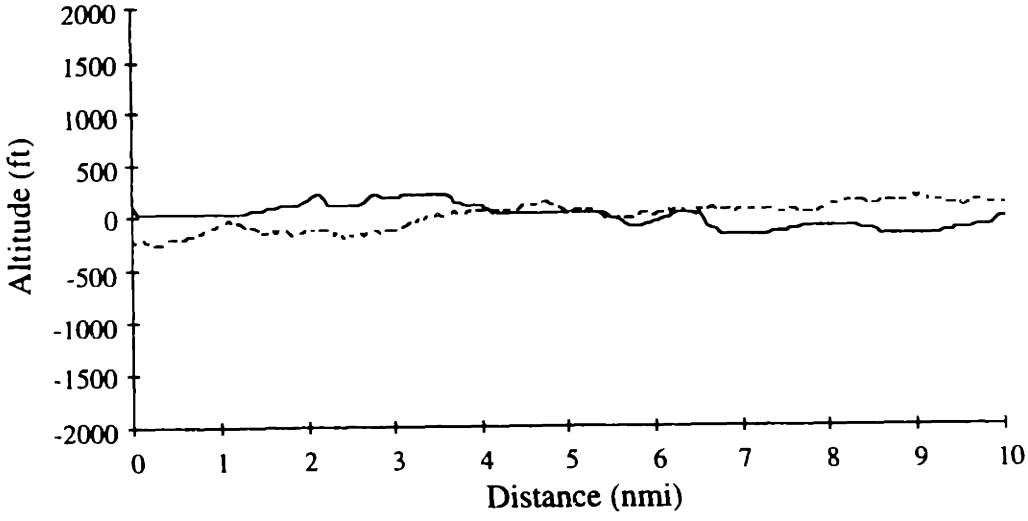
should be different in each category. It appears, however, that the correlation distance for terrain is not strongly coupled with variance; smooth regions may have similar correlation distances to steep regions. A more detailed categorization of terrain may then require a two-dimensional matrix based on variance and correlation distance. For the purposes of this analysis, however, categories based on variance alone are believed to be sufficient to enable a study of the effects of different types of terrain on system performance.

Table C.2
Corrected Autocorrelation Function Parameters

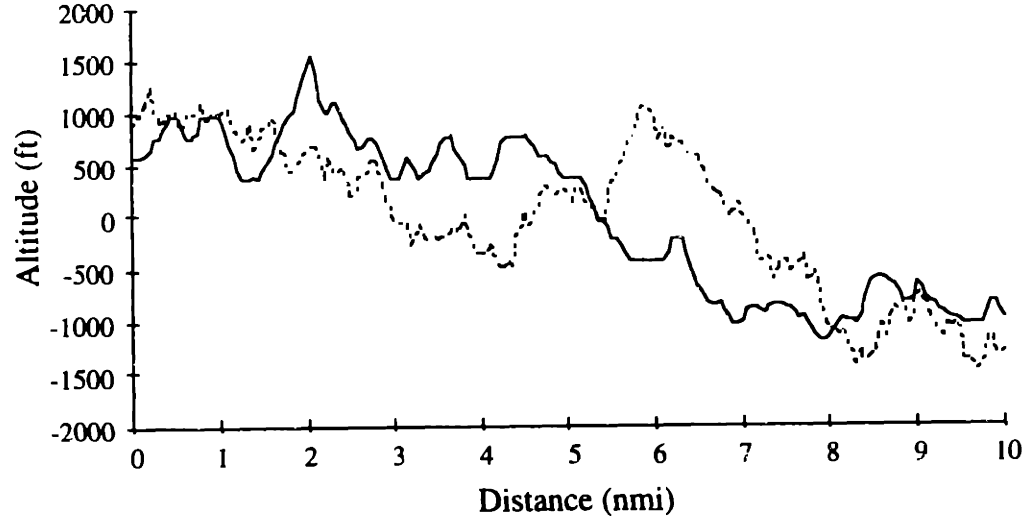
$$\phi_{yy}(\tau) = \sigma^2 e^{-|\tau|/\tau_0}$$

Terrain Category	σ^2 (m^2)	τ_0
Smooth	6,233	458
Moderately Smooth	72,212	1,551
Moderate	116,793	773
Moderately Steep	172,506	492
Steep	1,013,555	1,633

As a confirmation that the Markov modeling process is reasonable, comparisons of true terrain profiles with terrain generated using a Gauss-Markov process are shown in Figure C.7. The top plot shows a randomly-selected Smooth category terrain profile from the CD-ROM database and a randomly generated Gauss-Markov profile using Equations (C.8) and (C.9). The bottom plot shows randomly-selected profiles from the Steep terrain category. Note that the Gauss-Markov model appears to represent the large-scale changes in terrain altitude well, and appears to be a reasonable method by which terrain may be handled probabilistically.



Smooth (48°-49° N, 104° 30' W)



Steep (33°-34° N, 109° 30' W)

Figure C.7: Example Terrain Profiles
Solid: true terrain profile.
Dashed: random Gauss-Markov sample.
Samples shifted to zero mean.

C.3 Markov Chain Construction

C.3.1 The Markov Chain

One application of Markov processes that is of great value to a probabilistic analysis of systems is the Markov Chain (Kemeny & Snell 1967). Given a discrete-time Markov process, a set of discrete states can be constructed that describe the possible values that the process may take. In any time step, then, there is a certain probability (given by

$f(y_{n+1} \mid y_n)$) that the value of the process will change from its current value, y_n, to some other value, y_{n+1}. In Figure C.8, a five-state Markov Chain is shown, with the transition probability p_{ji} defined as the probability that the process will change from state i to state j in one time step.

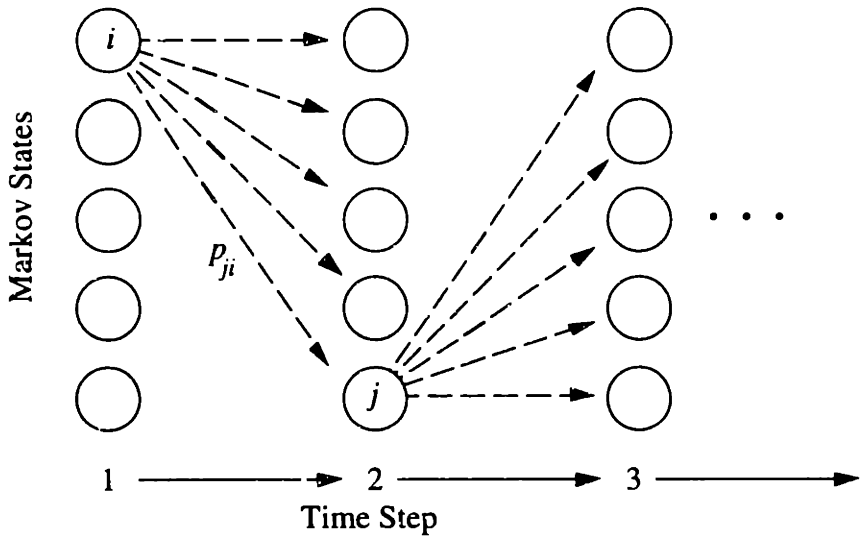


Figure C.8: Markov Chain

Next, a state vector $\mathbf{y}(t)$ is defined that holds the probabilities that the process takes on the value of each state in the chain at time t.

$$\mathbf{y}(t) = \begin{bmatrix} \text{probability that value is state 0 at time } t \\ \vdots \\ \text{probability that value is state } n \text{ at time } t \end{bmatrix} \tag{C.12}$$

Thus, the i^{th} element of $\mathbf{y}(t)$ is the probability that the process is in the i^{th} state at time t. The transition probabilities can also be placed into a state transformation matrix, shown in Equation (C.13).

$$\mathbf{T}(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1n}(t) \\ p_{21}(t) & p_{22}(t) & \cdots & p_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}(t) & p_{n2}(t) & \cdots & p_{nn}(t) \end{bmatrix} \tag{C.13}$$

Where $p_{ji}(t)$ is the probability that the state changes from i to j at time step t. Note that, in general, $p_{ji}(t)$ may vary with time.

Given an initial state probability vector, $\mathbf{y}(0)$, the probabilities of the process being in each state after one time step are given by $\mathbf{y}(1)$:

$$\mathbf{y}(1) = \mathbf{T}(1)\,\mathbf{y}(0) \tag{C.14}$$

Similarly, at time 2,

$$\mathbf{y}(2) = \mathbf{T}(2)\,\mathbf{T}(1)\,\mathbf{y}(0) \tag{C.15}$$

By extension, the probability of the process taking on a certain value at time t is then given by $\mathbf{y}(t)$ in Equation (C.16).

$$\mathbf{y}(t) = \mathbf{T}(t)\,\mathbf{T}(t-1)\ldots\mathbf{T}(1)\,\mathbf{y}(0) \tag{C.16}$$

Thus, by concatenating the transformation matrices together, the probability that the process takes on a certain value at a future time step may be calculated. This method is used to estimate the probability that terrain rises above certain altitudes.

Having found the parameters needed to describe and generate terrain profiles using a Markov process, it is now possible to construct a Markov Chain. The purpose of the chain is to determine the probability that the altitude of terrain will exceed the aircraft's altitude along a given flight path. The states in the Markov Chain represent different terrain altitudes, and the time steps described above are distance steps in this application.

The procedure for using the Markov Chain involves tracking the probability that the terrain reaches different altitudes as the aircraft flies along. Starting with an initial Above Ground Level (AGL) altitude, the probability that the terrain is at each state's altitude as the aircraft flies along is calculated. This involves using the state transition matrix described above. Those terrain altitudes above the aircraft's flight path represent cases in which an impact occurs. Accordingly, each state which is above the aircraft's altitude is tagged and no longer is treated using the Markov process model. Instead, these impact states will immediately transition to an absorbing Terrain Collision State. For example, in Figure C.9, the aircraft's flight path is shown relative to the Markov Chain. Consider the case in which terrain altitude is in state A. The state transitions shown with the dashed arrows represent potential future terrain altitudes below the aircraft. However, any transition from A to a state above the aircraft's flight path will result in a collision; the probabilities of these transitions are combined into a single transition into the Terrain Collision State (shown by the solid arrow). Similarly, terrain above the aircraft (B) or in a Terrain Collision State (C) will transition directly to the next Terrain Collision State. Thus, the probability that terrain

is in the Terrain Collision State at each step in the process represents the probability that a collision with terrain has occurred.

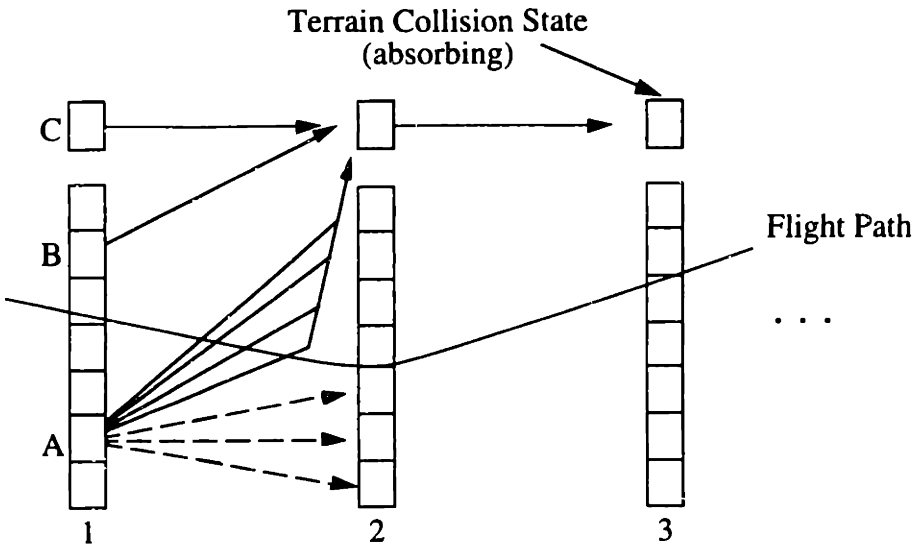


Figure C.9: Markov Chain Propagation Method

Each distance step (1, 2, 3 in Figure C.9) corresponds to one data point interval from the terrain database: 3 arc seconds (approximately 300 ft). A total of 50 states were used above the initial terrain altitude, each with a size of 10 to 20 meters (32 to 65 ft) depending on the type of terrain under study. Because terrain altitude can drop and then rise, an additional 50 states were also used below the initial terrain altitude. Additionally, the Terrain Collision State was created and used to store those cases in which terrain altitude had risen above the aircraft's altitude. This collision state was absorbing: once entered, the process would remain in that state. A schematic of the Markov Chain is shown in Figure C.10.

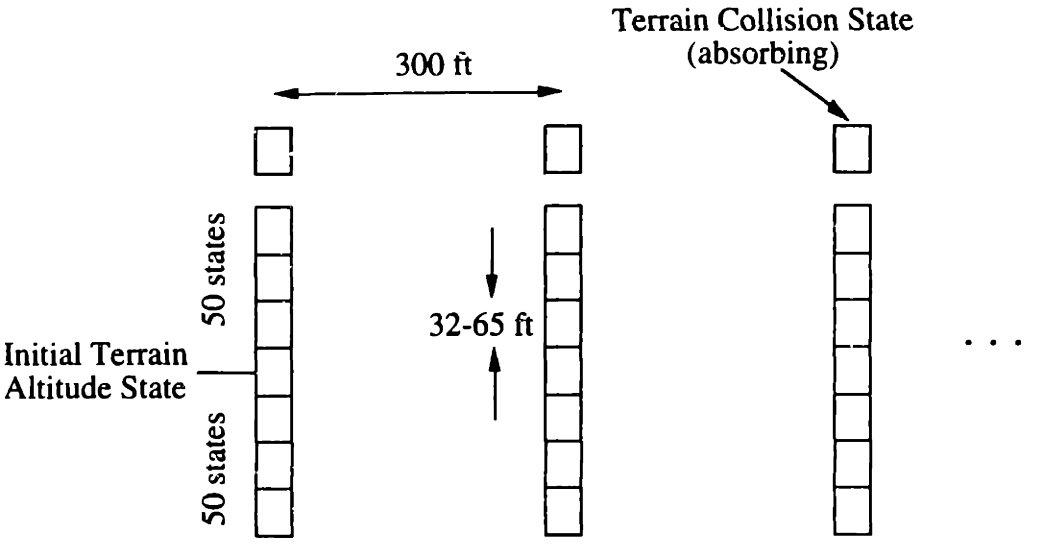


Figure C.10: Markov Chain Schematic
101 states, 300' apart horizontally

C.3.2 Calculation of Transition Probabilities

The probability that the process will transition from a starting value, y_n, to an ending value, y_{n+1}, can be found from the process defined by Equations (C.8) and (C.9). The probability density function for this process is:

$$f(y_{n+1} \mid y_n) = \frac{1}{\sqrt{2\pi\sigma^2(1-e^{-2\beta})}}\, e^{-\frac{1}{2}\frac{(y_{n+1}-e^{-\beta}y_n)^2}{\sigma^2(1-e^{-2\beta})}}. \qquad \text{(C.17)}$$

Next, the probability of transitioning from a given starting value to a given ending state (consisting of a range of values), $\Pr(y_n \Rightarrow y_{n+1})$, may be found. This situation is shown in Figure C.11, and Equation (C.18) is the required calculation.

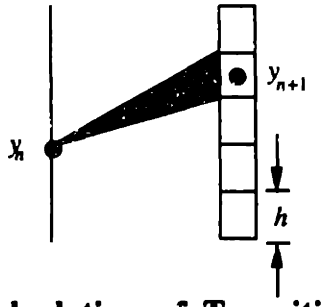


Figure C.11: Calculation of Transition Probability
Transition from y_n to state centered on y_{n+1}

$$p_{y_{n+1},y_n} = \int_{y_{n+1}-\frac{1}{2}}^{y_{n+1}+\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2(1-e^{-2\beta})}} e^{-\frac{1}{2}\frac{(y-e^{-\beta}y_n)^2}{\sigma^2(1-e^{-2\beta})}} dy \qquad \text{(C.18)}$$

where σ and β are the autocorrelation function parameters and h is the height of the altitude state block.

Equation (C.18) was used to construct a complete state transition matrix for a given terrain category. Next, the aircraft's flight altitude was superimposed over the Markov Chain. Those states above the aircraft's altitude resulted in a modification to the transition matrix such that there would be a transition to the absorbing state with probability 1. In addition, the probabilities corresponding to transitions to states above the aircraft's altitude were added and became the probability corresponding to a transition to the absorbing state.

C.3.3 Example State Transition Matrix Construction

Figure C.12 shows an example 5-state system with an aircraft flight path overlaid.

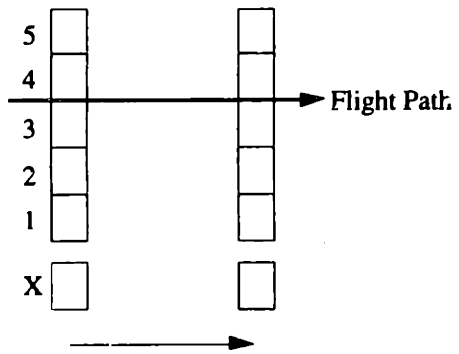


Figure C.12: Example 5-State Markov Chain
Flight path passes between states 3 and 4.
State X is the absorbing Terrain Collision State.

First, a complete state transition matrix is constructed:

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & 0 \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & 0 \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & 0 \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & 0 \\ p_{X1} & p_{X2} & p_{X3} & p_{X4} & p_{X5} & 1 \end{bmatrix} \qquad \text{(C.19)}$$

Recall that p_{ji} is the probability that the process transitions from state i to state j. For example, p_{13} is the transition probability for the change from state 3 to state 1. The sixth row and column of **T** represent the absorbing Terrain Collision State (state X in Figure C.12). Once entered, the process will remain in this state ($p_{XX} = 1$ and $p_{\bullet X} = 0$). Next, the possible transitions into state X need to be defined.

In this example, if terrain transitions to states 4 or 5 an impact will occur. Accordingly, starting with the complete **T** matrix in Equation (C.19), states 4 and 5 are modified such that they immediately transition into state X:

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ p_{21} & p_{22} & p_{23} & 0 & 0 & 0 \\ p_{31} & p_{32} & p_{33} & 0 & 0 & 0 \\ p_{41} & p_{42} & p_{43} & 0 & 0 & 0 \\ p_{51} & p_{52} & p_{53} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{C.20}$$

Finally, any transition from states 1, 2, or 3 into states 4 or 5 should be redirected into state X. This is done, for example, by setting p_{X1} to the sum of p_{41} and p_{51}, and setting p_{41} and p_{51} to zero. The final transition matrix is shown in Equation (C.21):

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 & 0 \\ p_{21} & p_{22} & p_{23} & 0 & 0 & 0 \\ p_{31} & p_{32} & p_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ p_{41} + p_{51} & p_{42} + p_{52} & p_{43} + p_{53} & 1 & 1 & 1 \end{bmatrix} \tag{C.21}$$

If the aircraft changes altitude, a new **T** matrix needs to be defined at each time step, taking into account the new impact states.

Appendix D.
TCAS Architecture and Alerting Logic

D.1 System Inputs

TCAS provides visual and aural indications of the traffic situation to the flight crew. The alerting criteria are based on four states (r, $\dot{r}$, h, $\dot{h}$) that are inferred from measurements of range and altitude. An electronic traffic display uses three measurements (r, ψ, and h) to produce an electronic depiction of the traffic situation.

As shown in Figure D.1, radar pulses are periodically sent from the own aircraft and are received by the transponder on another aircraft in the vicinity. The intruder's transponder then sends a reply pulse to the own aircraft, which calculates range using the round-trip time for the pulse.

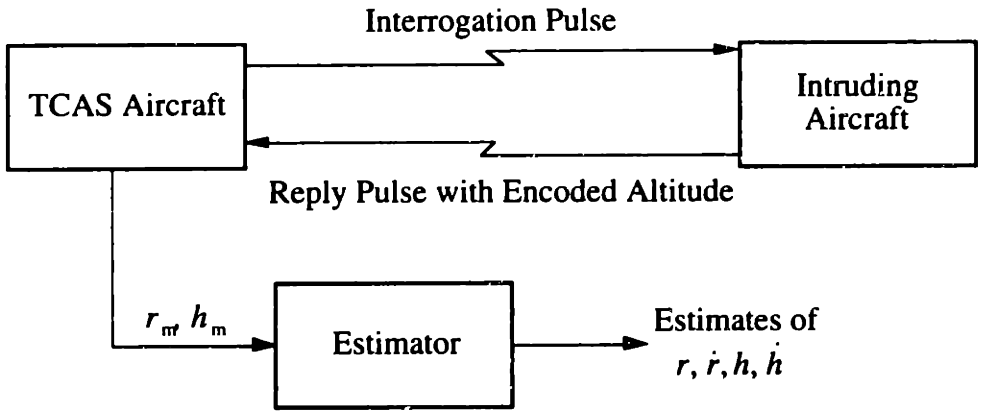


Figure D.1: TCAS Estimates of Range, Range Rate, Altitude, and Altitude Rate

To obtain altitude measurements, the intruding aircraft must have a transponder that sends its altitude encoded in the reply pulse. Currently, TCAS makes use of the same altitude encoding system that is used for transmitting altitude data to air traffic control radars. The altitude data from the transponders is discretized into 100 ft altitude bins.

D.1.1 Range and Range Rate Estimates

Because the measurements of range and altitude contain noise, an estimator, called an alpha-beta tracker, is used by TCAS to produce a filtered estimate of range and range rate (RTCA 1983). A more complex nonlinear tracker is used to estimate altitude, and

altitude rate. The alpha-beta tracker is a recursive estimator similar to a Kalman Filter but with constant filter gains of α and β.

The estimator for range and range rate uses the following discrete-time state vector:

$$\hat{\mathbf{x}}_k = \begin{bmatrix} r & \dot{r} \end{bmatrix}^T \tag{D.1}$$

Where the subscript k indicates the time step at which the estimate is made. The estimator works in one-second discrete time steps. The propagation of the state estimate is given by:

$$\hat{\mathbf{x}}_k^- = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \hat{\mathbf{x}}_{k-1}^+ \tag{D.2}$$

where a superscript (-) indicates the estimate prior to a measurement update, and (+) indicates the estimate after a measurement update. When a measurement of range ($\mathbf{z}_k$) is made, the estimate is updated using the following relation:

$$\hat{\mathbf{x}}_k^+ = (\mathbf{I} - \mathbf{K})\hat{\mathbf{x}}_k^- + \mathbf{K}\mathbf{z}_k \tag{D.3}$$

where the filter gain $\mathbf{K} = \begin{bmatrix} \alpha & 0 \\ \beta & 0 \end{bmatrix}$.

Range measurements are typically obtained with a standard deviation of approximately 30 ft, and can be assumed to be normally distributed (Wood 1993). TCAS uses filter gains of $\alpha = 0.4$ and $\beta = 0.15$ for range estimates. The filter equations were solved with a 30 ft standard deviation in range to determine the covariance matrix for the estimator, assuming a target moving with a constant range rate of 400 knots. The filter converges within approximately 17 seconds to a steady-state error standard deviation of approximately 18 ft in range and 6 feet per second in range rate. Range and range-rate estimate errors are correlated with a correlation coefficient of 0.69.

D.1.2 Altitude and Altitude Rate Estimates

Altitude measurements will, in general, contain a constant bias (due to calibration error or barometric pressure changes) and random noise. The minimum operational performance specifications for TCAS require altimetry accuracies at least as good as those given in Table D.1 (RTCA 1983). The examples in Chapter 5 assume that the aircraft is flying at 15,000 ft. Therefore, the total altimetry error for each aircraft has a standard deviation of $3\sigma = 174$ ft, or $\sigma = 58$ ft.

Table D.1
Required Altitude Accuracies for TCAS (RTCA 1983)
(99.7% Confidence Intervals)

MSL Altitude	3σ error (ft)
0	135
5,000	144
10,000	156
15,000	174
20,000	195
25,000	213
30,000	234
35,000	258
40,000	285

The altitudes encoded by the transponder are quantized into 100 ft increments. The nonlinear tracker used by TCAS is designed to produce smooth estimates of altitude despite the coarse, discrete resolution of the altitude data.

Monte Carlo simulations were performed to determine the accuracy with which the TCAS nonlinear tracker estimates altitude. In the simulation, the total altitude error was assumed to have a standard deviation of 58 ft, which was composed of a random noise component of $\sigma = 6$ ft and a random bias with $\sigma = 57.7$ ft (RTCA 1983). The flight of an intruder aircraft was simulated at a constant descent rate of 2,500 ft/min. Altitude reports from the intruder were quantized into 100 ft increments and passed into the altitude tracker algorithm. The standard deviation of the estimate converged within approximately 20 seconds in this simplified situation. The steady-state standard deviation of the estimated altitude was approximately 58.9 ft, with an altitude rate standard deviation of 0.56 feet per second.

The TCAS aircraft's altitude is assumed to be estimated to the tolerances given in Table D.1 ($\sigma = 58$ ft). The total relative altitude error is then given by:

$$\begin{aligned} \sigma_{tot} &= \sqrt{\sigma_{own}^2 + \sigma_{intruder}^2} \\ &= \sqrt{58^2 + 58.9^2} = 82.6 \text{ ft} \end{aligned} \tag{D.4}$$

D.2 Alerting Thresholds

D.2.1 Version 6.04A Thresholds

This section describes the alerting thresholds for TCAS Version 6.04A (RTCA 1983, MITRE 1993). TCAS uses two-stage alerting, with cautionary alerts called Traffic

Advisories (TA) and warnings called Resolution Advisories (RA). TAs direct the crew's attention to a potential threat, but no avoidance information is provided. RAs provide avoidance commands such as "Climb" or "Descend".

RAs and TAs are generated when the current state vector of an intruding aircraft passes two tests: range and altitude. These tests make up a four-state alerting threshold, which is more easily understood when broken up into two two-state thresholds. The range test is based on range (r) and range rate ($\dot{r}$), as shown in Figure D.2 for RAs and TAs. The values of the parameters shown in Figure D.2 are provided in Figure D.6 and Table D.2.

For threats that are closing with the aircraft ($\dot{r} < 0$), the range criteria are met if the threat is projected to enter a sphere of radius DMOD around the aircraft within τ seconds. An aircraft already within the sphere of radius DMOD is considered to be a threat if the product of range and range rate is small (the aircraft are not diverging quickly).

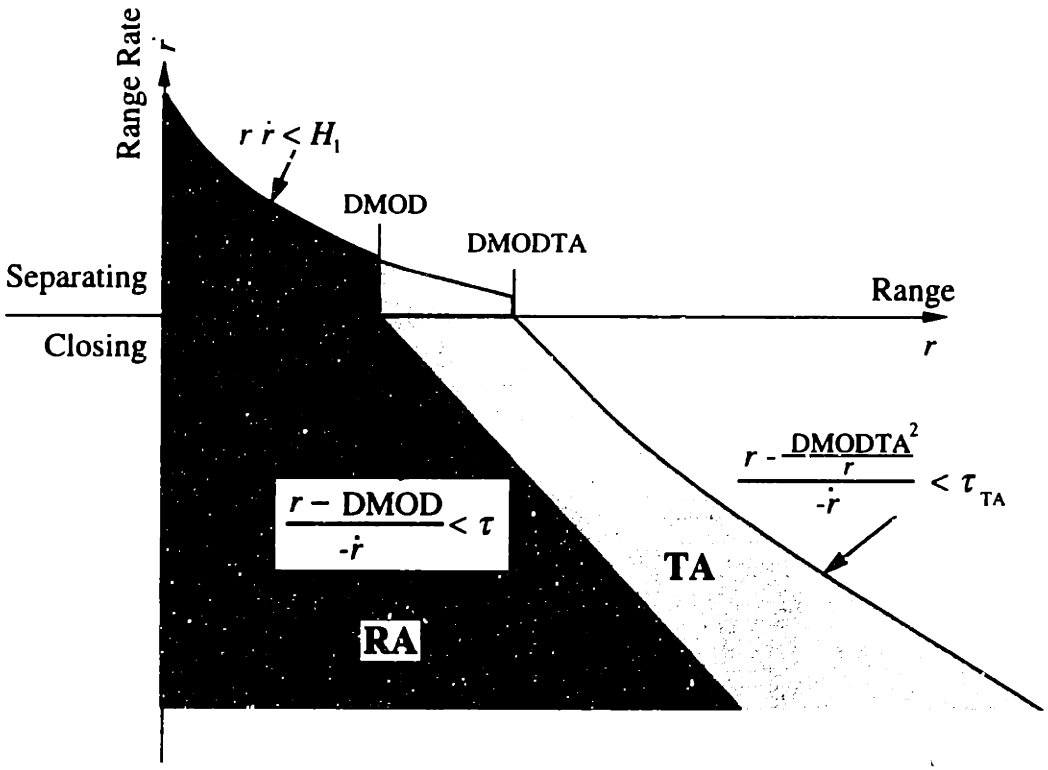


Figure D.2: Example TA and RA Range Test Thresholds
Parameter values are given in Figure D.6 & Table D.2.

The altitude test uses the relative altitude and relative altitude rate of the two aircraft, as shown in Figure D.3 for RAs and TAs.

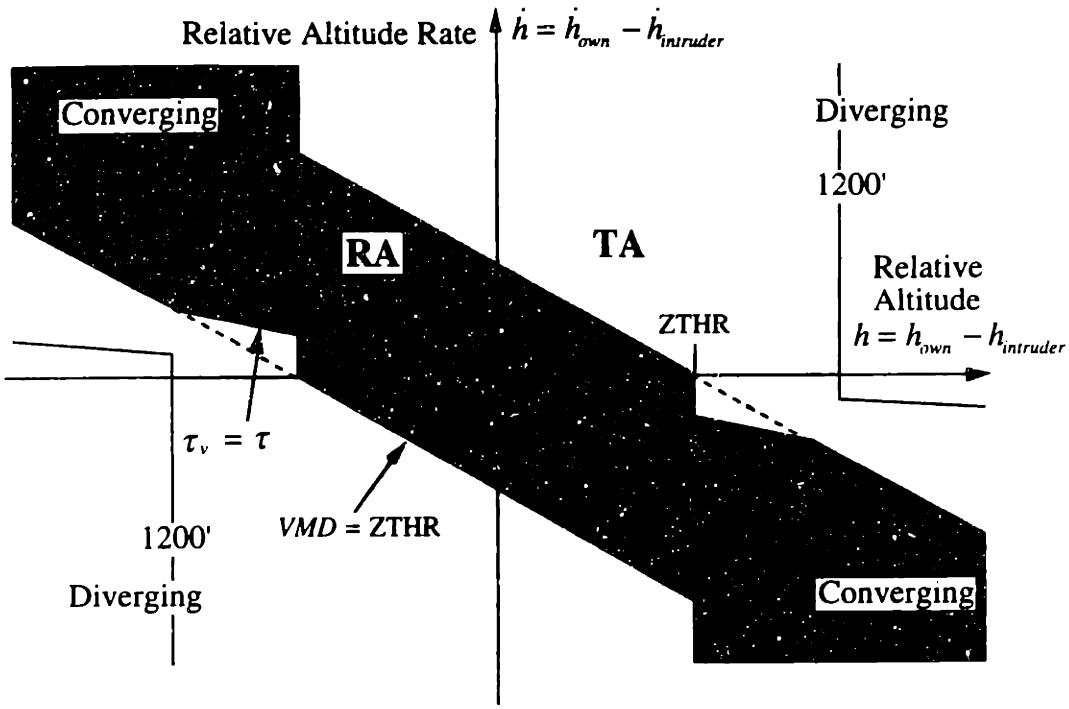


Figure D.3: Example TA and RA Altitude Test Thresholds
VMD = Vertical Miss Distance (see Figure D.4)
τ_v = time to coaltitude = $-h / \dot{h}$
Parameter values are given in Figure D.6 & Table D.2.

The altitude test uses two sets of thresholds, depending on whether the threat aircraft is currently within a certain altitude buffer around the aircraft, denoted ZTHR. If the threat is within the buffer (Figure D.4), it is considered to be a hazard if its projected altitude at the closest point of approach (called the Vertical Miss Distance, VMD) is also within the altitude buffer.

If the threat is currently outside the buffer (Figure D.5), it is considered to be a hazard when the time to coaltitude (τ_v) is small and either the VMD is projected to be within the altitude buffer or the intruder will cross the own aircraft's altitude before the closest point of approach.

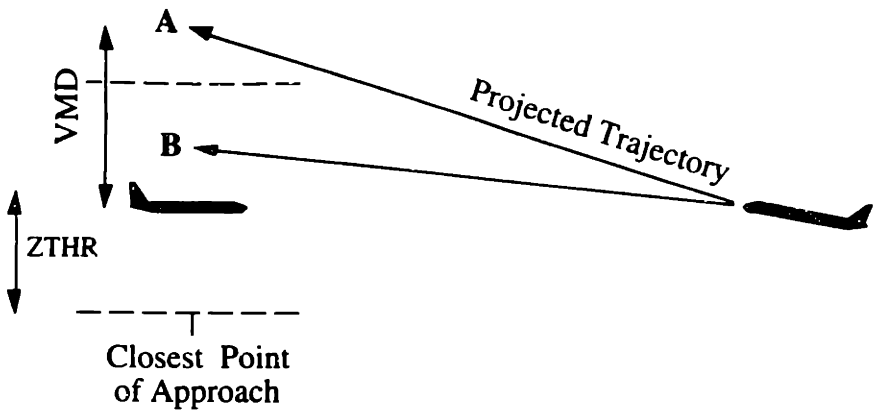


Figure D.4: Altitude Test When Intruder is Inside ZTHR
A: Test fails: VMD > ZTHR. No alert is issued.
B: Test passes: VMD < ZTHR. An alert is issued.

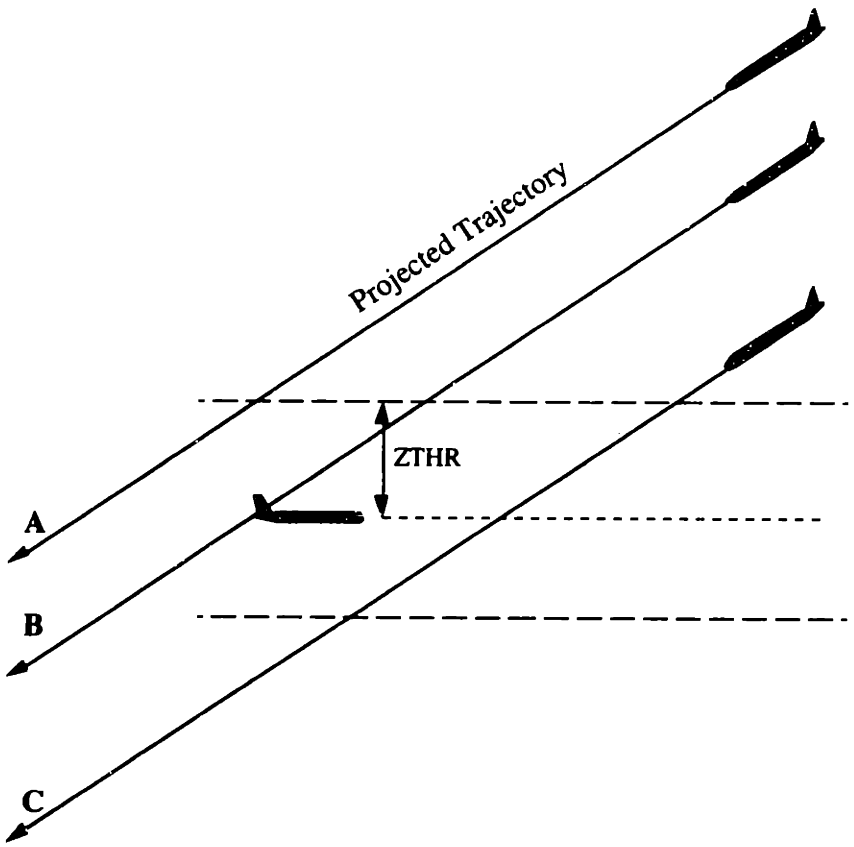


Figure D.5: Altitude Test When Intruder is Outside ZTHR
A: VMD > ZTHR: No alert is issued.
B: VMD < ZTHR: An alert is issued.
C: Time to coaltitude < time to closest point of approach: An alert is issued.

The values of the parameters defining the thresholds depend on the altitude of the aircraft, as shown in Figure D.6 and Table D.2. Only TAs can be generated when the

aircraft is below 1,000 ft AGL. The parameter τ is the threshold for predicted time to closest approach. In the notation used in the Figure, τ 30/35/48 indicates a threshold of 30 sec if the TCAS aircraft is level or climbs or descends in the same direction as the threat but at a lower rate, otherwise 35 sec for an RA, and 48 sec for a TA. DMOD is a range buffer used in the range alerting threshold (Figure D.2). ALIM is the minimum desired miss distance and ZTHR is an altitude buffer used in the altitude threshold (Figure D.3).

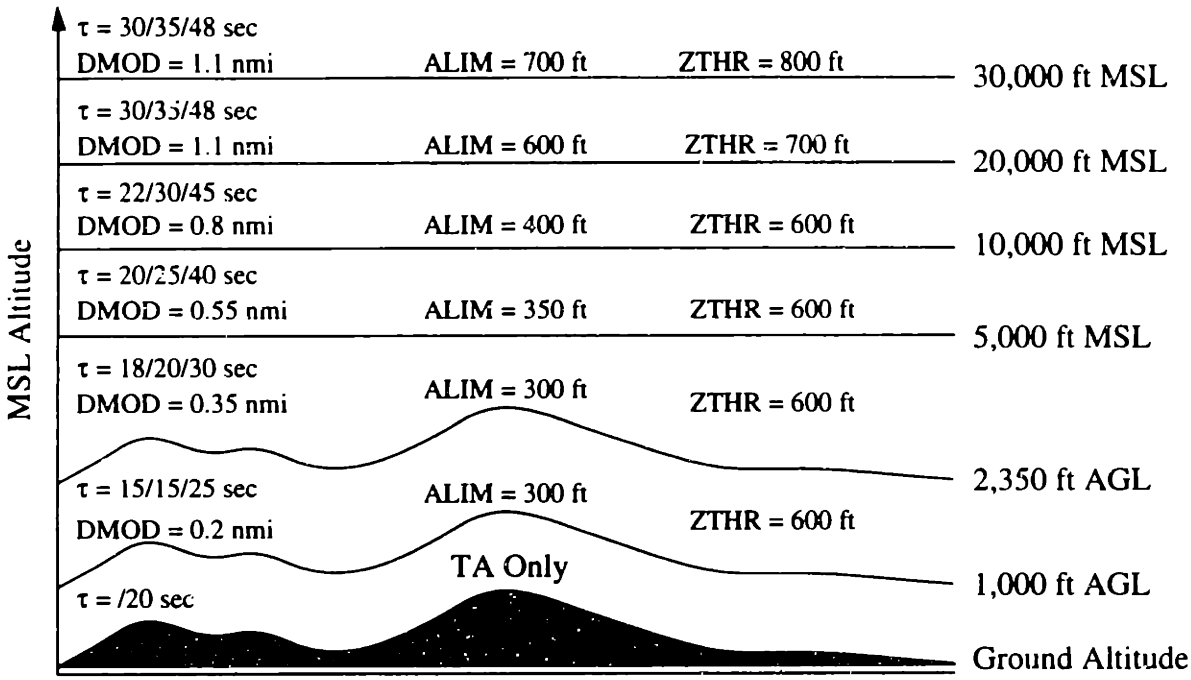


Figure D.6: TCAS Threshold Parameters
Figure adapted from (RTCA 1983) and (Miller et al 1994).

In Table D.2, H_1 is a parameter used in the range threshold, and DMODTA is the value of DMOD used for the TA alerting threshold (Figure D.2).

Table D.2
Additional TCAS Alert Threshold Parameters
(Aircraft at 15,000 ft MSL)

Parameter	Value
H_1 (nmi^2/sec)	0.00278
DMODTA (nmi)	1.50

D.2.2 Version 1.0 Thresholds

Several changes in the alerting logic were made between Version 1.0 and Version 6.04A (MITRE 1993). However, the major effects of these changes for the situation considered in Chapter 5 are the parameters τ and DMOD, summarized in Table D.3.

Table D.3
Comparison of TCAS RA Threshold Parameters
Aircraft at 15,000 ft AGL

Parameter	Version 1.0	Version 6.04A
τ	30 sec	22 sec
DMOD	1.0 nmi	0.8 nmi

D.3 Resolution Advisory Logic

When an RA is issued, TCAS examines the situation to determine an acceptable avoidance maneuver for the TCAS aircraft. All avoidance maneuvers are built on the assumption that the threat aircraft continues on its present trajectory and that the TCAS aircraft begins to maneuver after a 5 second delay with a vertical acceleration of 0.25 g. RAs are issued in one of two direction 'senses'. A climb sense is selected when the aircraft must not exceed some descent rate or must climb to avoid a collision. Descend sense is selected when the aircraft should not climb greater than some rate or must descend. The possible advisories are placed in three categories: positive -- Climb (Descend), negative -- Don't Descend (Don't Climb), and vertical speed limits -- Limit Descent (Climb) to 500, 1,000, or 2,000 feet per minute. A "Climb" RA assumes a 1,500 feet per minute climb rate. "Limit Descent" RAs assume the pilot descends no faster than the given rate. Similar rate assumptions are made for descend sense advisories.

When an RA is called for, TCAS searches through the possible advisories, from weakest to strongest, and selects the advisory that provides at least ALIM separation between the aircraft and the intruder (the value of ALIM is defined in Figure D.6). For example, in Figure D.7, the own aircraft selects climb sense to avoid an intruder. TCAS checks each advisory in turn, from "Limit Descent to 2,000 fpm" to "Climb". "Limit Descent to 500 fpm" is the first advisory that provides ALIM vertical separation (see Figure D.7), so it is the advisory that is issued to the crew.

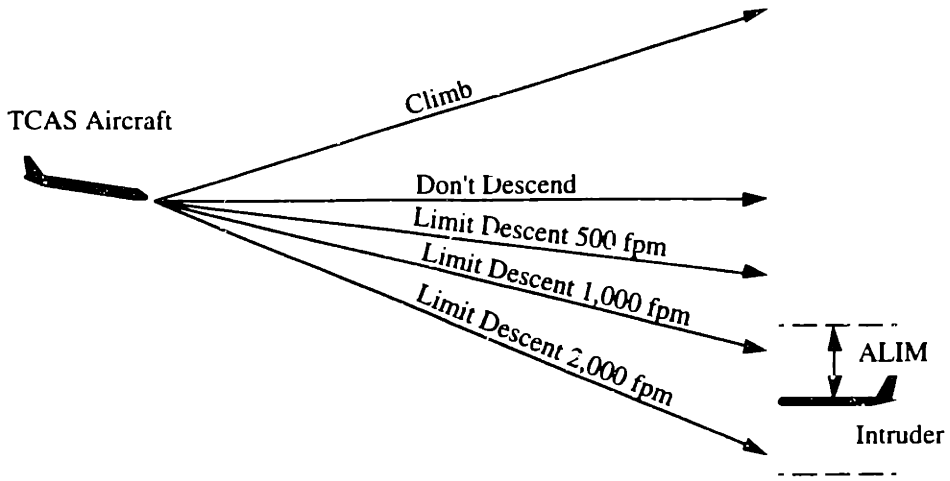


Figure D.7: Potential Climb Sense Resolution Advisories

Once TCAS has issued an RA, it will not change the climb or descend sense of the alert. For example, if a "Climb" RA is given, TCAS will not later change the alert to a "Descend" RA. The motivation behind this logic is to prevent TCAS from issuing seemingly contradictory commands to the flight crew.