NASA Conference Publication 3131

Joint University Program for Air Transportation Research 1990–1991

Proceedings of a conference held in Athens, Ohio June 20–21, 1991

NASA

NASA Conference Publication 3131

Joint University Program for Air Transportation Research 1990–1991

Compiled by
Frederick R. Morrell
NASA Langley Research Center
Hampton, Virginia

Proceedings of a conference sponsored by the National Aeronautics and Space Administration, Washington, D.C., and the Federal Aviation Administration, Washington, D.C., and held in Athens, Ohio
June 20–21, 1991

NASA
National Aeronautics and Space Administration
Office of Management
Scientific and Technical Information Program
1991

PREFACE

The Joint University Program for Air Transportation Research is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). These research grants, which were instituted in 1971, build on the strengths of each institution. The goals of this program are consistent with the aeronautical interests of both NASA and the FAA in furthering the safety and efficiency of the National Airspace System. The continued development of the National Airspace System, however, requires advanced technology from a variety of disciplines, especially in the areas of computer science, guidance and control theory and practice, aircraft performance, flight dynamics, and applied experimental psychology. The Joint University Program was created to provide new methods for interdisciplinary education to develop research workers to solve these large scale problems. Each university submits a separate proposal yearly and is dealt with individually by NASA and FAA. At the completion of each research task, a comprehensive and detailed report is issued for distribution to the program participants. Typically, this is a thesis that fulfills the requirements for an advanced degree or a report describing an undergraduate research project. Also, papers are submitted to technical conferences and archival journals. These papers serve the Joint University Program as visibility to national and international audiences.

To promote technical interchange among the students, periodic reviews are held at the schools and at a NASA or FAA facility. The 1990-1991 year-end review was held at Ohio University, Athens, Ohio, June 20-21, 1991. At these reviews the program participants, both graduate and undergraduate, have an opportunity to present their research activities to their peers, to professors, and to invited guests from government and industry.

This conference publication represents the eleventh in a series of yearly summaries of the program. (The 1989-1990 summary appears in NASA CP-3095). Most of the material is the effort of students supported by the research grants. Four types of contributions are included in this publication: a summary of ongoing research relevant to the Joint University Program is presented by each principal investigator, completed works are represented by full technical papers, research previously in the open literature (e.g., theses or journal articles) is presented in an annotated bibliography, and status reports of ongoing research are represented by copies of presentations with accompanying text.

Use of trade names of manufacturers in this report does not constitute an official endorsement of such products or manufacturers, either expressed or implied, by the National Aeronautics and Space Administration or the Federal Aviation Administration.

Frederick R. Morrell
NASA Langley Research Center

CONTENTS

PREFACE iii

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AN INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 1990–91 3
Robert W. Simpson

EXPERIMENTAL INVESTIGATION OF PASSIVE INFRARED ICE DETECTION FOR HELICOPTER APPLICATIONS 11
Adam Dershowitz and R. John Hansman, Jr.

DETERMINISTIC MULTI-ZONE ICE ACCRETION MODELING 19
K. Yamaguchi, R. J. Hansman, Jr., and M. Kazmierczak

ALERT GENERATION AND COCKPIT PRESENTATION FOR AN INTEGRATED MICROBURST ALERTING SYSTEM 29
Craig Wanke and R. John Hansman, Jr.

COCKPIT DISPLAY OF HAZARDOUS WINDSHEAR INFORMATION 41
R. John Hansman, Jr., and Craig Wanke

OHIO UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY, 1990–1991 51
Robert W. Lilley

A HYBRID MODULATION FOR THE DISSEMINATION OF WEATHER DATA TO AIRCRAFT 55
Dennis M. Akos

FAULT DETECTION AND ISOLATION FOR MULTISENSOR NAVIGATION SYSTEMS 61
Paul A. Kline and Frank van Graas

ESTABLISHING AN EVOKED-POTENTIAL VISION-TRACKING SYSTEM 69
Trent A. Skidmore

GPS INTERFEROMETRIC ATTITUDE AND HEADING DETERMINATION: INITIAL FLIGHT TEST RESULTS 79
Frank van Graas and Michael Braasch

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1990–1991 103
Robert F. Stengel

TARGET PITCH ANGLE FOR THE MICROBURST ESCAPE MANEUVER 113
Sandeep S. Mulgund

STOCHASTIC ROBUSTNESS 127
C. Marrison

NEURAL NETWORKS FOR AIRCRAFT SYSTEM IDENTIFICATION 141
Dennis J. Linse

OPTIMAL AIRCRAFT PERFORMANCE DURING MICROBURST ENCOUNTER 155
Mark L. Psiaki and Robert F. Stengel

INTELLIGENT FAILURE-TOLERANT CONTROL 163
Robert F. Stengel

SYSTEMATIC METHODS FOR KNOWLEDGE ACQUISITION AND EXPERT SYSTEM DEVELOPMENT 173
Brenda L. Belkin and Robert F. Stengel

TECHNICAL NOTES AND CORRESPONDENCE: STOCHASTIC ROBUSTNESS OF LINEAR TIME-INVARIANT CONTROL SYSTEMS 181
Robert F. Stengel and Laura R. Ray

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AN INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1990 - 1991

Robert W. Simpson
Flight Transportation Laboratory
M.I.T., Cambridge, MA 02139

SUMMARY OF RESEARCH ACTIVITIES

1. INTRODUCTION

There are three completed projects and three continuing research activities under way under the sponsorship of the FAA/NASA Joint University Program as the 1990-91 period ends. There were a number of publications during the year and a set of Annotated References to them is provided in Section 3. As well, the complete conference paper is included for a few of these references. A brief summary of the continuing research activities is provided in the next section.

2. REVIEW OF CONTINUING RESEARCH ACTIVITIES

2.1 Wanke, C., and Hansman, R. J., Hazard Assessment and Cockpit Presentation Issues For Microburst Alerting Systems.

Introduction

Low altitude wind shear events, and in particular microbursts, are the leading weather-related cause of fatal aviation accidents in the U.S. For this reason, wind shear hazard detection and alerting system development is a very active area of research, and has been a focus of the Joint University Program for several years. The variety of systems under development for wind shear detection and ground-to-air datalink, combined with the advent of electronic cockpit instrumentation, allow many options for implementation of an

integrated microburst warning system. Recent work at MIT has concentrated on two facets of this problem: (1) microburst hazard assessment, and (2) cockpit display of microburst alerts to flight crews.

Microburst Hazard Assessment

The first task for microburst alert generation is to combine data from several systems, which use different measurement techniques, into an estimate of the hazard posed to aircraft by a particular wind shear event. This must be done as accurately as possible to minimize both missed and unnecessary alarms. However, the danger to aircraft posed by a microburst is not a directly measurable quantity. Thus, a microburst "hazard criterion" is needed, a quantity which can be directly measured or inferred from the available measurements. One example of such a criterion, currently used by the Terminal Doppler Radar System (TDWR) is maximum wind change across the microburst. Alternatives to this are shear-based or energy-based criteria which are more difficult to measure but relate more directly to the effect of the windfield on the dynamics of an aircraft.

A technique based on batch flight simulation has been developed to evaluate some candidate hazard criteria.[1] Simulated approaches through numerically-modeled microbursts are flown, and a quantitative measure of approach degradation due to the windfield, a "microburst impact parameter," is computed. This measure includes both glideslope deviations and airspeed losses. The candidate microburst hazard criteria are then correlated with this microburst impact parameter. The value of the linear correlation coefficient is indicative of the usefulness of that criterion for evaluating the microburst hazard to aircraft.

The results, using 12 modeled microburst windfields in several geometrical configurations, indicate that shear-based (airspeed change per unit distance) and energy-based measures correlate very well with the microburst impact parameter. Also, maximum wind change was found to correlate very poorly. Further work is being done to extend these results to cover additional high-resolution microburst windfields, and to use these results to define "hazard thresholds" for accurate alert generation.

Cockpit Display of Microburst Alerts

Once alerts are generated, they need to be quickly and effectively disseminated to the flight crew. Recent work at MIT on this issue has focused on use of a digital datalink and electronic cockpit displays to graphically present microburst alerts to the pilot. A previous piloted simulator experiment performed earlier in this program demonstrated that graphical alerts are significantly more

---

[1] Wanke, C., and Hansman, R.J., "Alert Generation and Cockpit Presentation for an Integrated Microburst Alerting System," AIAA Paper 91-0260, January 1991.

effective than verbal or alphanumeric alerts.[2] A second experiment to further explore the use of graphical alerts was then designed and run in June of 1991. This experiment tested several different graphical alert formats using the MIT Advanced Cockpit Simulator, a part-task simulator of a modern transport aircraft with electronic cockpit displays. Alerts were presented on the Electronic Horizontal Situation Indicator; an example of a multi-level alert format is shown in Figure 1. Some of the issues being examined in this experiment are:

- Is presentation of microburst alerts on the EHSI display clear and effective?
- Should the available measurements be used with a hazard threshold to generate a single-level alert for "hazardous" microbursts only, or should multiple levels of intensity be displayed?
- Should measurements from all of the available sensors be combined to form a single "fused" alert, or should alerts from different sources be independently displayed?
- What are the procedural implications of displaying graphical alerts, i.e., does the positional information present in these alerts cause the pilot to significantly alter his avoidance strategy?

Analysis of the data from this experiment is in progress; nine active airline pilots have participated to date. The experiment involves flying a total of 12 approach scenarios with the simulator, using a variety of graphical alert formats. Preliminary results indicate that multi-level alerts are more desirable than single level alerts, and that display of alerts on the EHSI was found to be very clear and useful by the pilots. Further analysis is underway, and results will be presented at the next JUP quarterly meeting.

2.2 Hahn, E., and Hansman, R. J., The Situational Awareness Effect of Automated ATC Datalink Clearance Amendments.

Introduction

Among the capabilities envisioned for the future datalink (i.e. digital ground to air communications) system will be the use of these aircraft specific communications to supplement and possibly replace most routine Air Traffic Control (ATC) messages. By reducing the volume of ATC communications, including en route and terminal clearance amendments, datalink would relieve the already congested VHF voice frequencies. Additionally, having information being received in digital form makes it extremely easy for the

---

[2] Wanke, C., Chandra, D., Hansman, R. J., and Bussolari, S.R., "A Comparison of Voice and Datalink for ATC Amendments and Hazardous Wind Shear Alerts," 4th International Symposium on Aviation Space Safety, Toulouse, France, 20-22 November, 1990.

aircraft Flight Management Systems to read the information directly, increasing efficiency and reducing data transfer errors. However, the automation of datalinked ATC clearance amendments could adversely affect the flight crew's situational awareness by reducing their level of involvement in the amendment process. This project is investigating the effects of information presentation mode and amendment procedures on situational awareness of possible implementations of datalink.

Current Work

Work is proceeding on a part-time simulation study using the MIT Advanced Cockpit Simulator. The experiment centers on the information presentation mode and amendment procedures in possible ATC datalink implementations. Data is collected on the error detection capabilities of pilots when the procedure is altered to change the number of times the airline pilot subject must review the data during the process. In particular, the need for keying in of data and also the need for a readback to ATC is varied independently between scenarios. In addition, the mode of presentation is also independently varied between verbal, textual, and graphical formats to investigate whether one presentation mode is better than others. Data will also be collected on the pilot's time performance and subjective preferences.

The scenarios will simulate the Northeast Corridor during heavy traffic and thunderstorm conditions in order to increase the number of clearance amendments given to the pilot. Each is designed around two separate kinds of errors: the aircraft will be cleared into heavy thunderstorm cells and also will be routed incorrectly (i.e. to the wrong airport, the wrong approach fix, or on an illogical path). Each subject pilot will fly ten scenarios, each of which will have a different combination of presentation mode and amendment procedure. Eighteen subjects are currently being sought to participate in the experiment. The testing stage of the experiment is projected to terminate in September.

3. ANNOTATED REFERENCES OF 1990-91 PUBLICATIONS

3.1 Chi, Zhihang, A Graphic Simulation System for Adaptive, Automated Final Approach Spacing, Flight Transportation Laboratory Report 91-3, June 1991, Flight Transportation Laboratory, MIT, Cambridge, MA, 02139

As airline industry grows and air traffic increases drastically, terminal airspace around busy airports is becoming more and more crowded. To accommodate the soaring demand for use of airports, a plausible and profitable way is to improve the efficiency of existing airports. An automated final approach spacing system can improve the efficiency as well as alleviate the workload of air traffic controllers.

In this research we develop an automated adaptive and interactive Final Approach Spacing Advisory (FASA) system to be used in future at busy airports. Our system is able to generate and update final arrival paths for aircraft and guarantees that the aircraft land as scheduled and safely spaced. It prompts air traffic controllers for calls of turns and speed reductions until they merge into the approach center line. These automated prompts are adaptive to errors in the execution of the arrival paths due to winds or pilot response. Also our system grants the controllers the ability to manually alter the landing schedule without violating constraints and regulations.

It is assumed that a separate metering system has developed a plan for landing aircraft. This landing schedule is represented to the ATC controller in FASA as a sequence of "schedule boxes" or "slot markers" moving along the extended center line. The ATC controller has some degree of control over the planned landings by moving these slot markers. A range of feasible changes in the slot position ("feasible interval") can be displayed. If a slot is moved rearwards to delay its landing, or if it arrives late for its slot on the center line, all subsequent slot markers will automatically shift rearwards to maintain safe spacings, and the automated prompts are then based on the new plan. If an aircraft arrives early, slots are not shifted frontwards automatically, but if any aircraft arrives ahead of its slot and will violate safe separations before touchdown, an alert is given to the controller. The mathematical methods used to implement this system are capable of incorporating a large variety of constraints and other interactive features.

The FASA system was implemented on an Apollo Unix workstation in the C language and a graphical simulation using X-Windows was developed to test prototypical traffic situations for a single runway. The images of the elements of the FASA system, such as aircraft, slot markers and extended runway center line, are drawn in different colors on the color display of the Apollo workstation. The images of aircraft move along the runway center line. Prompts for calls of turns and speed reductions are implemented as the blinking of the images of the relevant aircraft. Error alerts are implemented as the change of colors of the relevant images. The user of our simulation program is given the ability to mouse any slot markers, drag them along the runway center line and put them anywhere within the feasible interval mentioned earlier.

The future research effort will extend the logic to incorporate planning for takeoffs, interactive insertion of missed approaches into the landing plan, and multiple runway operations. All of these features are easily implementable by extending the methods used in this thesis to develop FASA.

3.2 **Hazelton, Lyman R., An Expert System for Temporal Planning with an Application to Runway Configuration Management, Flight Transportation Laboratory Report 91-1, February 1991, Flight Transportation Laboratory, MIT, Cambridge, MA, 02139.**

The centralized Air Traffic Flow Control system depends critically on correct airport capacity estimates for its success. This requirement demands that area supervisors, responsible for the control facilities of the airports, construct a realistic *schedule* of the runway configurations to be used during the next several hours of operation. Constructing such a schedule is a difficult task, as it depends on several time dependent, strongly interacting factors. These variables include the airport ceiling and visibility, the wind speed and direction, the operational status of navigational aids, the expected traffic demand, noise abatement procedures and time of day, runway maintenance and repairs, and snow and ice removal. A computer program to weigh these factors and generate such a schedule would be very helpful and could have a major economic impact.

Significant problems stand in the way of creating a general scheduling program using standard techniques. Because each airport has difficult runway geometry, a different surrounding geography and population distribution, and different weather patterns, each scheduler program must be customized for the airport at which it is to operate. Using standard programming languages and techniques to create so many unique programs would be exorbitantly expensive. Further, the programs would require periodic updates to reflect changes in the economy and patterns of use at each airport, adding to the cost. Finally, some of the factors involved are not easily quantifiable, making the standard numerical methods of operations research difficult to apply.

Rule based "expert system" programs "reason", using logic, about a problem rather than just computing numerical results. They possess abilities which give them distinct advantages in the kind of problem domain typically encountered in Air Traffic Control: their structure consists of a standardized central "inference engine" and an easily customized set of rules and facts about a particular application; they can be built incrementally and are more easily maintained than traditional programs; and they can provide explanations of the logical processes used to make a decision, which can immensely simplify debugging.

At the time that this research was begun, there was a significant obstacle to using an expert systems approach, as well. While the program of runway configuration management clearly requires the ability to reason about time, there were no well defined mechanisms for carrying out this kind of reasoning in the logic of an expert system. In fact, there was almost no theoretical work in the area of temporal logic upon which one might base such a mechanism.

This research describes a number of powerful new ideas and techniques which extend the range of applicability of expert systems to temporally dependent problems. The most important of these is the description of a representation and a set of mechanisms which can be added to standard logic in order to reason about persistence. Also included are methods necessary to reason about irreversible actions. Implementation of these new ideas and mechanisms extends the temporal reasoning capabilities of an expert system inference engine so that it can successfully be applied to problems in Air Traffic Control such as runway configuration management. Indeed, they are of such a fundamental nature that they are applicable to a wide variety of planning and scheduling problems, perhaps including crew maintenance scheduling, space shuttle launch sequence scheduling, and autonomous robot task scheduling.

3.3 **Yamaguchi, K., and Hansman, R. J., Deterministic Multi-Zone Ice Accretion Modeling, AIAA-91-0265, AIAA 28th Aerospace Sciences Meeting, January 1991.**

A multi-zone predictive analytical model was created to describe the growth of ice on aerofoil leading edges under glare icing conditions. Comparison with experimental ice shapes showed good agreement. (See copy of paper included.)

3.4 **Wanke, C., and Hansman, R. J., Alert Generation and Cockpit Presentation for an Integrated Microburst Alerting System, AIAA-91-0260, AIAA 28th Aerospace Sciences Meeting, January 1991.**

The issues of providing a cockpit alert for microbursts which combines ground and air-based detection systems are investigated. An experiment has been designed to evaluate candidate graphical cockpit displays which uses a part-task piloted simulator. (See included paper.)

3.5 **Dershowitz, Adam L., and Hansman, R. J., Passive Infrared Ice Detection for Helicopter Applications, paper presented at 46th Annual Forum of American Helicopter Society, Washington, DC, May 1990.**

A technique is proposed to detect remotely the icing on rotor leading edges by using passive infrared thermometry to detect the warning caused by the release of the latent heat of fusion as supercooled water freezes. (See included paper.)

Experimental Investigation of Passive Infrared Ice Detection for Helicopter Applications

Adam Dershowitz* and R. John Hansman, Jr.**
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts U.S.A

Abstract

A technique is proposed to remotely detect rotor icing on helicopters. Using passive infrared (IR) thermometry it is possible to detect the warming caused by latent heat released as supercooled water freezes. During icing, the ice accretion region on the blade leading edge will be warmer than the uniced trailing edge resulting in a chordwise temperature profile characteristic of icing. Preliminary tests were conducted on a static model in the NASA Icing Research Tunnel for a variety of wet (glaze) and dry (rime) ice conditions. The characteristic chordwise temperature profiles were observed with an IR thermal video system and confirmed with thermocouple measurements. A prototype detector system was built consisting of a single point IR pyrometer, and experiments were run on a small scale rotor model. Again the characteristic chordwise temperature profiles were observed during icing, and the IR system was able to remotely detect icing. Based on the static and subscale rotor tests the passive IR technique is promising for rotor ice detection.

1. Introduction

The reliable measurement of ice accretion is an important requirement for helicopter all-weather operation. Rotor icing can present a significant hazard due to increased torque requirements, aerodynamic performance degradation, control problems and severe vibrations resulting from uneven ice shedding. Helicopter exposure to icing conditions has increased over the past decade due to the increase in routine instrument flight operations.

The two primary helicopter applications for ice detection are the monitoring of critical components for Caution and Warning alerts, and the management of ice protection systems. For Caution and Warning applications, ice accretion on critical components should be detected before a hazardous condition exists.

Because rotating components often experience a significantly different icing condition than the fuselage, visual evidence of ice accretion on non-rotating structures is not adequate for Caution and Warning applications. Indirect indications of icing such as torque rise or vibration typically emerge only after the hazardous condition exists. Ideally, the direct measurement of ice accretion on the rotating components would provide the earliest and most accurate measurement of potential icing hazards.

For helicopters equipped with ice protection systems, the management of these systems can be critical. The most common type of rotor protection is electrothermal deicing. These systems require significant amounts of electrical power so the blades are typically deiced in segments. Because there are no effective rotor ice detection systems it is necessary to overheat and continually cycle the electrothermal deicing systems to assure adequate ice protection. If effective ice detection were employed the deicing systems could be operated more efficiently and effectively.

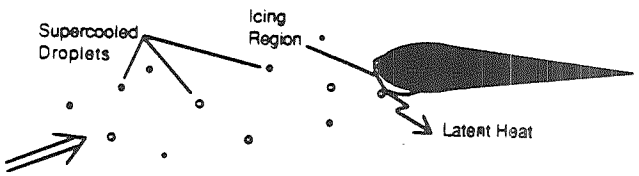


Fig. 1 Schematic representation of latent heat release during icing

* Research Assistant
** Associate Professor, Associate Fellow AIAA

2. Background

2.1 Existing Ice Detection Techniques

The most common ice detectors currently employed for aviation applications are direct contact type devices. In contact type detectors, either a probe or surface mounted transducer senses the ice presence through a variety of physical mechanisms. These include resonant frequency shifts, optical blocking, heat capacity, electrical capacity, and ultrasonic thickness gauging.[1]

Contact type ice detectors have not been successfully utilized for direct measurement of rotor ice accretion due to the difficulties inherent in instrumenting the rotors. These include mounting restrictions imposed by structural limitations of the blade, transmission of information between the rotating and non-rotating frames, acceleration limitations of the sensors, and erosion problems. As a result, contact type detectors are typically mounted on the fuselage and the rotor conditions are inferred.[2] These techniques have had limited success because of the complex dependence of the icing process on temperature, liquid water content and velocity, which vary significantly between the fuselage, and rotor as well as along the blade span itself.

2.2 Remote Monitoring of Rotating Blades

Remote sensing offers significant advantages for monitoring rotating blades. If ice accretion can be remotely detected by sensors mounted in the fuselage, then a direct measurement of rotor ice accretion can be accomplished without the difficulties identified above. Several potential techniques have been investigated for remote ice detection including; high speed video, and active microwave. The technique which has emerged as the most promising is passive infrared.

3. Passive Infrared Ice Detection

3.1 Theory of Operation

The passive IR technique detects ice accretion through remote IR thermometry of the icing surface. When ice accretes on an airfoil or rotor, the region where the accretion occurs becomes warmer than the surrounding surface due to release of the latent heat of fusion as the impinging supercooled water droplets freeze. This process is shown schematically in Fig.1. It should be noted that the ice accretion is typically limited to a small region near the leading edge, and that much of the rotor surface will remain clear of ice. These uniced areas will have surface temperatures at or near the ambient temperature which will always be below freezing in icing conditions.

As a result of the thermodynamic processes described above, the rotors will exhibit a characteristic chordwise surface temperature profile when in icing conditions. Schematic temperature profiles are shown in Fig. 2 for wet and dry icing conditions. Prior to icing, the rotors will equilibrate to some temperature slightly warmer than the static air temperature due to kinetic heating. During icing, the temperature will increase in the accretion region as indicated in Fig. 2. For wet (glaze) ice accretions, the wet stagnation region will remain at 0°C while for dry (rime) accretions the temperature will reach a peak in the stagnation region at some temperature below freezing.

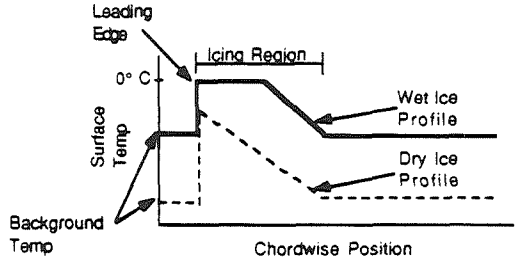


Fig. 2. Schematic chordwise temperature profiles for wet and dry icing conditions

In the proposed technique, shown in Fig. 3, the rotor surface temperature would be measured by an infrared pyrometer mounted on the fuselage which scans the blade in the chordwise direction as it passes overhead. The aim point of such a detector could be varied between spanwise positions to give full coverage of the blade. Infrared pyrometers are commonly used to remotely measure surface temperature through thermally emitted black body radiation in the IR spectrum.

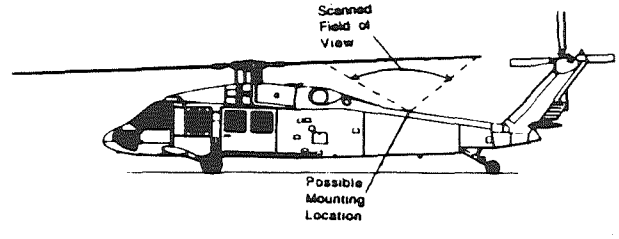


Fig. 3. Possible helicopter mounting scheme for a passive infrared ice detector

Passive IR video thermography has been used for several years in the U.S. and France as a diagnostic tool for validation of ice accretion modeling codes.[3,4] Synchronized IR video images have also been used to monitor rotating components in flight as part of the Advanced Turboprop development effort.[5] Due to uncertainty in surface emissivity, attenuation through the icing cloud and thermal drift of the detectors it is

sometimes difficult to accurately measure absolute temperature, however relative temperature resolution of 0.2°C is achievable. This is considered adequate for ice detection where only the relative temperature profile between the leading and trailing edge regions is required.

The passive IR system also has the potential to be used to monitor the operation of thermal deicing systems. Once ice sheds, the relatively hot blade surface will be exposed, resulting in an identifiable thermal signature.

3.2. Static Tests

3.2.1 Experimental Setup for Static Tests

In order to evaluate the feasibility of the passive IR ice detection concept, a series of static tests were conducted in the NASA Icing Research Tunnel (IRT). These preliminary tests were conducted on a non-rotating 6 ft. chord airfoil model with a 25° sweep which could be varied between ±20° angle of attack. The model was instrumented with thermocouples and electrothermal heaters in the leading edge region.

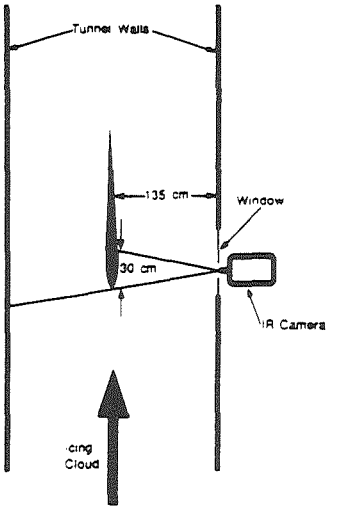
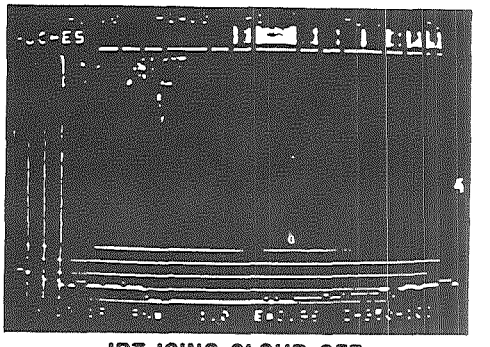


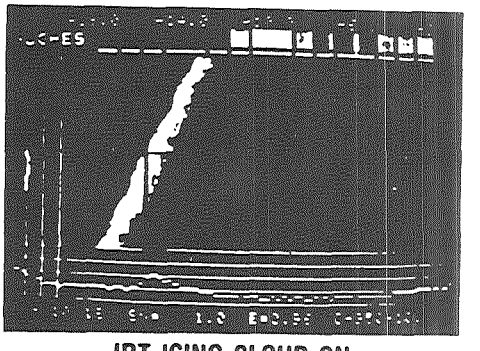
Fig. 4. Schematic of IRT setup

Infrared diagnostics were provided by a Hughes Probeye thermal imaging system which uses 6 Indium Antimonide (InSb) detectors and a spinning mirror to scan both horizontally and vertically. This imaging system, while somewhat more complex than required for an ice detection system, employs IR detectors very similar to those proposed for use in the passive system. These detectors, when cryogenically cooled, are sensitive in wavelengths from 2.0 to 5.6 microns. In the imaging system the output of the detectors was fed into a processor that developed thermal video images and chordwise temperature profiles. The data was stored on videotape. In the experimental set up, shown in Fig. 4, the IR camera was mounted outside of the tunnel and viewed the leading 30 cm of the airfoil through a hole in the tunnel window.

Experiments were run over a wide range of icing conditions, with airspeeds from 208 mph (93 m/s) to 129 mph (58 m/s) temperatures from -17°C (2.4 °F) to -3°C (26° F) and liquid water content (LWC) from 0.53 g/m^3 to 1.96 g/m^3. For each experimental run the airfoil was first cleared of any ice. The airspeed and temperature of the wind tunnel was stabilized and the spray was then turned on for a predetermined time span, typically 1 to 3 minutes. In some cases the spray was cycled on a second time, for 30 seconds, without deicing first. This second cycle allowed verification of the ability of the technique to detect ice accretion over an existing ice layer.



IRT ICING CLOUD OFF



IRT ICING CLOUD ON
(NOTE WARMING OF LEADING EDGE)

Fig. 5. IR video output of model lower surface (T=-11°C, V=92 m/s, LWC=0.17 g/m^3)

3.2.2. Static Test Results

A typical IR video image is presented in Fig. 5 for the swept airfoil at a tunnel temperature of -11°C, a LWC of 0.17 g/m^3, with a mean volumetric diameter (MVD) of approximately 15 microns, an angle of attack

of 4°, and a free stream velocity of 92 m/sec. Note that the chordwise temperature profile is shown across the bottom of the image. With the spray off, the airfoil surface temperature is uniform at approximately the tunnel temperature. When the spray is turned on, the increase in leading edge surface temperature resulting from the latent heat release is clearly visible. After the spray was turned off, the leading edge would slowly cool as the remaining liquid water on the surface froze.

The IR, and thermocouple measured chordwise temperature profiles for typical wet (glaze) ice conditions are presented in Fig. 6. In this case the tunnel temperature was -5°C, the LWC was 1.95 g/m^3, with a MVD of 20 microns, the angle of attack was 10°, and the free stream velocity was 59 m/s. With the spray off, the airfoil had a uniform indicated temperature. The warmer temperature observed upstream of the stagnation point is an image of the warm tunnel wall behind the airfoil. When the spray was turned on, the expected surface temperature profile for wet growth was observed. The leading 8 cm of the airfoil indicated a surface temperature of 0°C indicating wet ice growth. Aft of the wet region, the surface temperature dropped to the uniced value. The IR measurements were consistent with the thermocouple readings. The LEWICE ice accretion code was run to simulate ice growth at these same conditions. The LEWICE predicted surface temperatures, as shown in Fig 7, are consistent with those measured with the IR system. When the spray was turned off the surface temperature again became uniform as the surface water froze. When the spray was cycled on, the characteristic wet ice temperature profile was again observed.

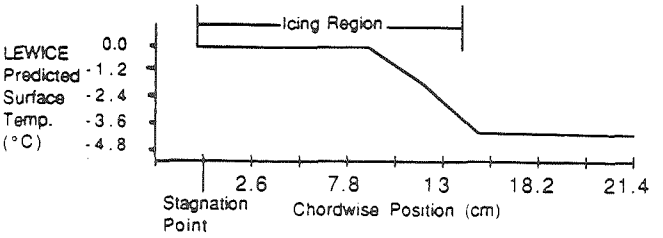


Fig. 7. LEWICE predicted temperature profile for wet (glaze) ice conditions (T=-5°C, V=59m/s, LWC=1.95 g/m^3).

The IR measured temperature profiles for a dry (rime) ice condition also show the expected temperature profile. An example is presented in Fig. 8. In this case the tunnel temperature was -17°C, the LWC was 0.17 g/m^3 , with a MVD of approximately 15 microns, the free stream velocity was 92 m/s and the angle of attack was 4.1°. With the spray off, the airfoil had a uniform indicated temperature. When the spray was turned on, the surface temperature was seen to peak at the stagnation point. This curve is characteristic of dry (rime) growth where the surface temperature reflects the local collection efficiency (i.e. increased impingement causes increased icing and increased heating). The stagnation point has the largest collection efficiency and consequently the greatest latent heat release. When the spray was turned off the temperature quickly dropped to the preicing value. The characteristic rime ice profile again appeared when the spray was cycled on.

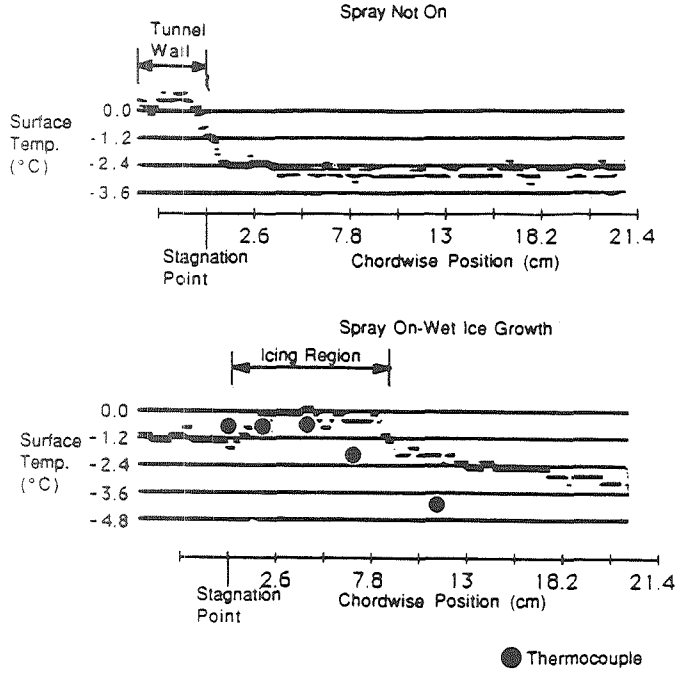


Fig. 6. Chordwise temperature profiles for wet (glaze) ice conditions (T=-5°C, V=59m/s, LWC=1.95 g/m^3)

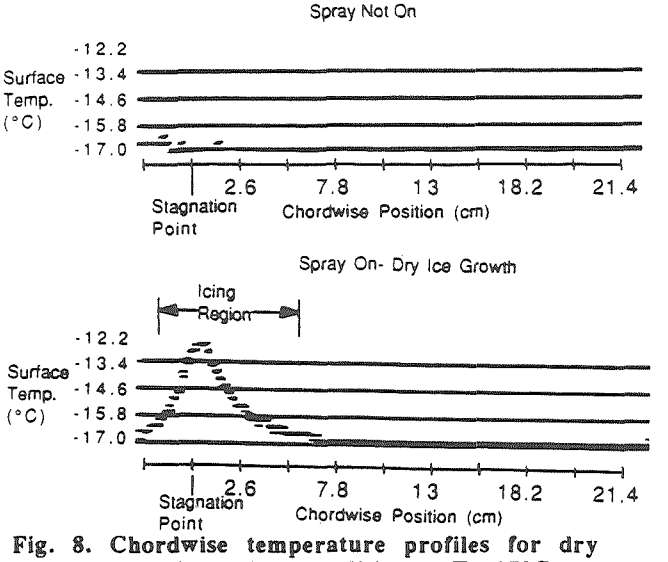


Fig. 8. Chordwise temperature profiles for dry (rime) ice conditions (T=-17°C, V=92m/s, LWC=0.17 g/m^3).

3.3 Rotor Tests

3.3.1 Infrared Ice Detector Design

In order to evaluate the IR ice detection technique on rotating systems a prototype detector, suitable for both flight and ground tests, was developed. This ice detector, shown in Fig. 9, is essentially a point infrared pyrometer. It was built around a HgCdTe (Mercury Cadmium Telluride) infrared detector. This detector was chosen as it offers high sensitivity in the range of 2-5 microns, using only thermoelectric cooling. Other detector types were considered, as they offer higher sensitivity, or operate at longer wavelengths. However they require cryogenic cooling, which would not be practical for flight operations. This detector uses a four stage thermoelectric cooler, with an automatic controller, to keep the temperature of the detector at -73° C. A gel cell battery was used to provide the bias for the detector at 2.6 V.

The HgCdTe detector is positioned at the focus of a single Calcium Fluoride (CaF_2) lens. The simple biconvex lens had a 10 cm focal length. Calcium Fluoride was selected as it has high transmissivity in the necessary IR band and is not overly sensitive to thermal shock, or water degradation. Concentric aluminum tubes were used, with a Teflon sleeve, to allow the length of the pyrometer to be adjusted giving a focus for a target distance from 15 cm to infinity. This range allows the prototype to be used on a small test rotor or in full scale flight tests. This lens gave the detector a target spot size of 0.18 cm diameter at 56 cm as used in the tests, and a spot size of 1 cm diameter at 3 m, typical of a helicopter application. In order to prevent icing or fogging of the lens small gas jets were set up to spray dry nitrogen across the lens.

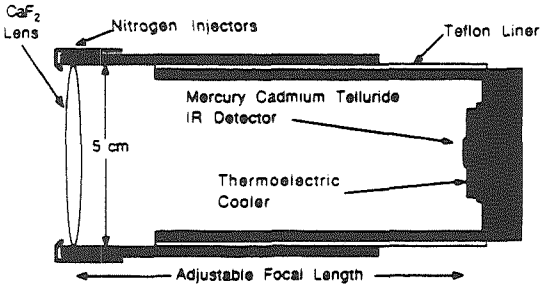


Fig. 9. Infrared ice detector schematic

The signal from the pyrometer itself was processed as shown in Fig. 10. It was first fed into a high gain, low noise amplifier. The amplifier was AC coupled to prevent saturation. This had the effect of amplifying only short term temperature changes. The signal was sent into a digital oscilloscope, and stored for post processing if required.

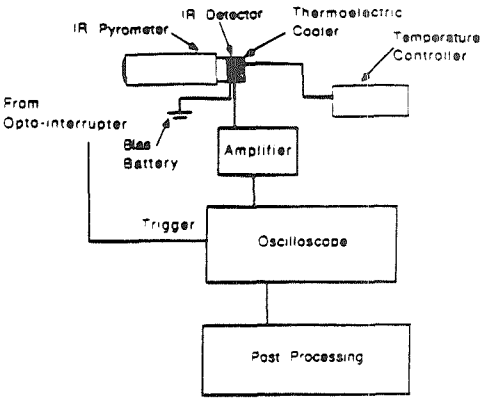


Fig. 10. Ice detector signal processing path

3.3.2. Test Rig For Rotating Tests

In order to test the prototype ice detector, a test rig was constructed. It consisted of a single blade counter-balanced rotor with a 22 cm span and a 2 cm chord, built of aluminum, schematically shown in Fig. 11. The small rotor was chosen to allowed the entire rig to be placed in a lab freezer. This system was able to run over a wide range of rotation speeds, from 0 Hz to 40 Hz. The tests were nominally run at 22 Hz which allowed good ice formation, at reasonable rotor Reynolds numbers. The blade passed through the field of view of the pyrometer at an angular speed typical of a tail rotor, which is several times faster than that of a main rotor. This test therefore was conservative in that significantly faster detector dynamics were required than would be necessary for main rotor use. The angle of attack was adjustable over a range of positive and negative values. An opto-interrupter was set up across the blade to measure rotation rate and to provide trigger pulses for tracking the blade.

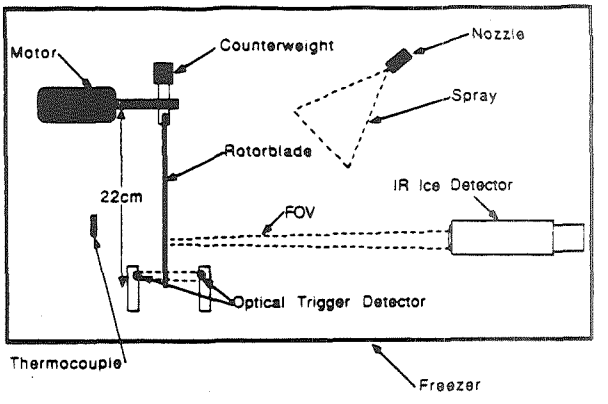


Fig. 11. Rotor test rig

The rotor and ice detector were both mounted in a freezer which had a minimum temperature of -23° C. The lens was 56 cm from the rotor plane. The ice detector was aimed at the 13.5 cm span position of the rotor, and focused to provided the sharpest signal possible. Because of the exploratory nature of the tests, an uncalibrated water spray was considered adequate. A simple spray system with an estimated LWC of 0.9 g/m^3 at the rotor was used. The freezer temperature was measured using a thermocouple.

3.3.3. Experimental Procedure

For a typical experimental run the blade rotation rate and freezer temperature were allowed to stabilize. The temperature and rotation rate were recorded. An uniced temperature profile was then recorded. Water at 0° C was then sprayed for 10 sec and a temperature profile during icing was recorded. Several profiles were recorded after the icing event while the blade temperature profile was returning to equilibrium. Again the temperature and rotation rate were recorded. Finally the blade was deiced and the cycle was repeated. These runs were done over a range of temperatures from -20° C to 60° C, and at several different angles of attack. During icing the rotation rate would drop slightly due to the increased drag from the ice, and the temperature would rise from opening the freezer, which was necessary to initiate the spray.

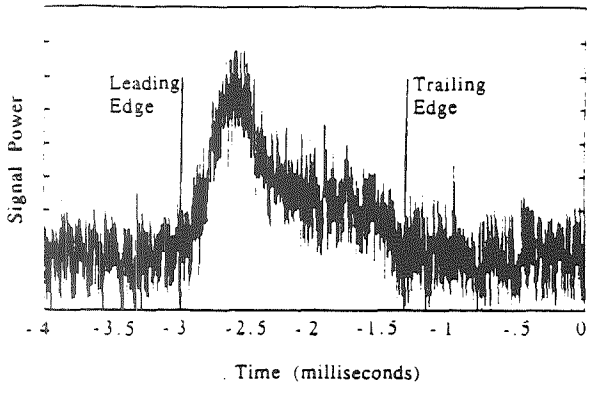


Fig. 12a. Unfiltered IR temperature profile

3.3.4. Rotor Results

A typical unfiltered signal from the pyrometer is shown in Fig. 12a at -6.8° C, 22 Hz and 0° angle of attack. The signal is shown in arbitrary units because only the temperature profile is of interest and the signal scales with amplifier gain, DC offset, and detector bias voltage. Because the signals contained a component of high frequency noise, they were digitally low pass filtered at 10 kHz resulting in signals as shown in Fig. 12b. As expected there was significant heating in the leading edge region, where there was icing occurring, while the rest of the blade remained at a relatively constant temperature. This is the same characteristic profile that was measured in the static tests. Before the spray began, the blade and the back wall of the freezer were at thermal equilibrium as indicated by the relatively flat line.

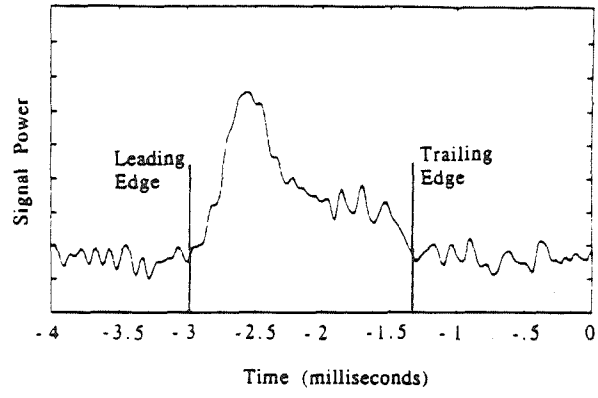


Fig. 12b. IR profile filtered at 10 kHz

Figs. 13 through 16 show the effect of ambient temperature on the signal. For all of these cases the rotor started at 22 Hz with 0° angle of attack, and the spray LWC is 0.9 g/m^3. During very cold conditions, such as Fig. 13 (-15° C) where there was a large temperature difference between ambient and freezing, there was a correspondingly large temperature change in the icing (leading edge) region of the blade. This resulted in the strong temperature profile shown in Fig. 13. Fig. 14 (-6.8° C) show a signal typical of glaze ice, where the icing profile is clearly visible. Fig. 15 (-3° C) shows a case just below freezing, where there was minimal glaze ice. Because the ambient temperature was close to 0° C the signal is small but clearly detectable. When the blade was stopped to be deiced there was only a small amount of ice present. A temperature profile taken at a temperature just above icing conditions is shown in Fig. 16. There was no signal indicative of icing, which is consistent with the observation that there was no ice accretion on the blade after the run.

In order to evaluate any changes in detectability with angle of attack, the angle of attack was varied over several runs. There was very little change in the temperature profile signal, although the higher angle of attack runs did include a slightly wider heating region, due to the increased size of the icing region.

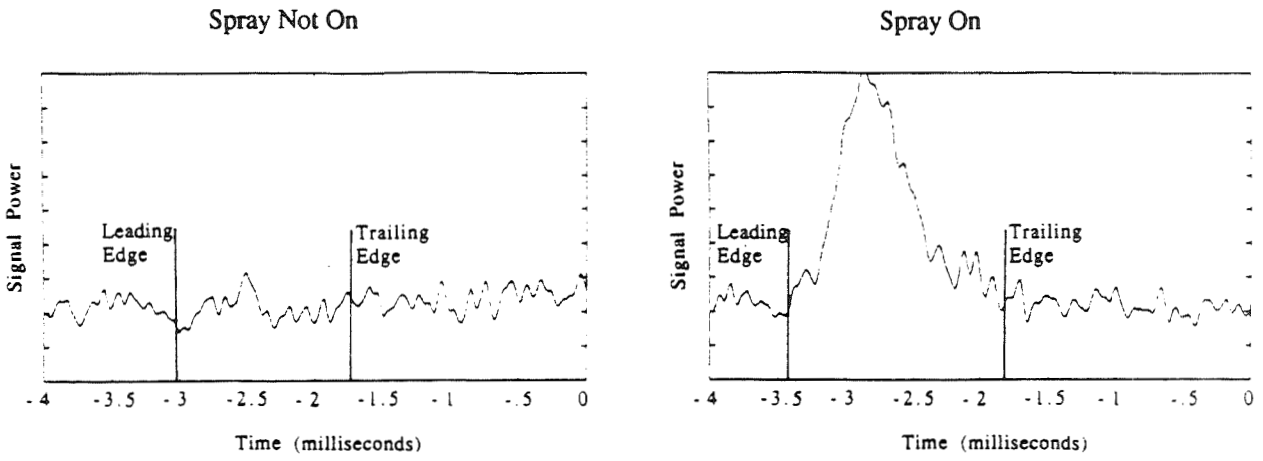


Fig. 13. IR Profiles at T=-15°C with and without water spray

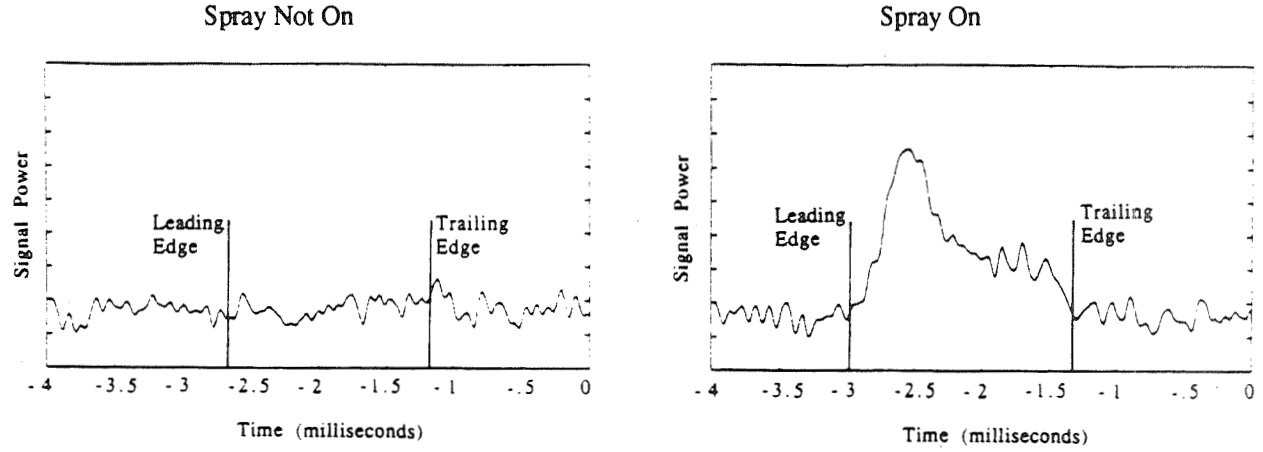


Fig. 14.IR Profiles at T=-6.8°C with and without water spray

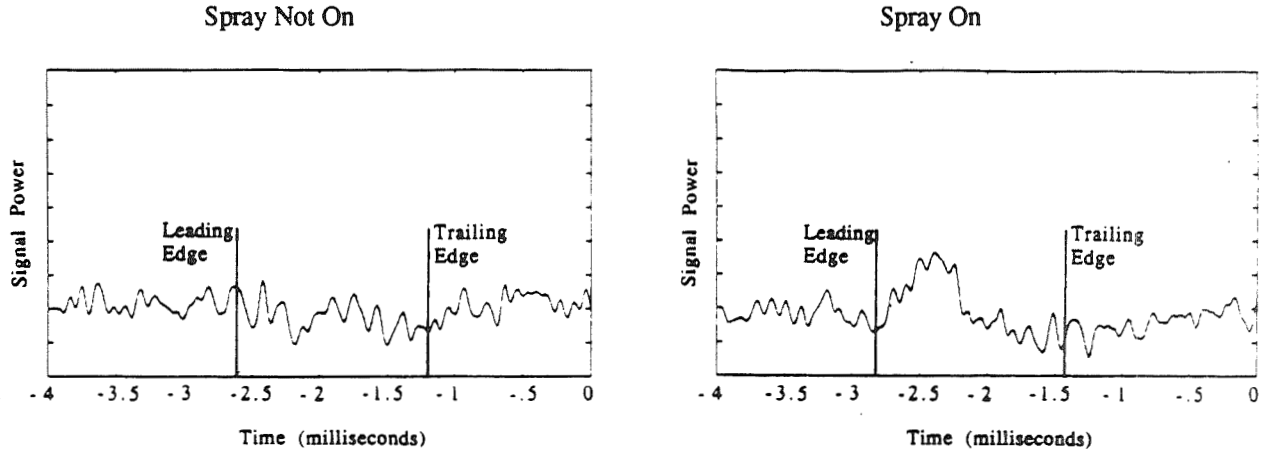


Fig. 15. IR Profiles at T=-3°C with and without water spray

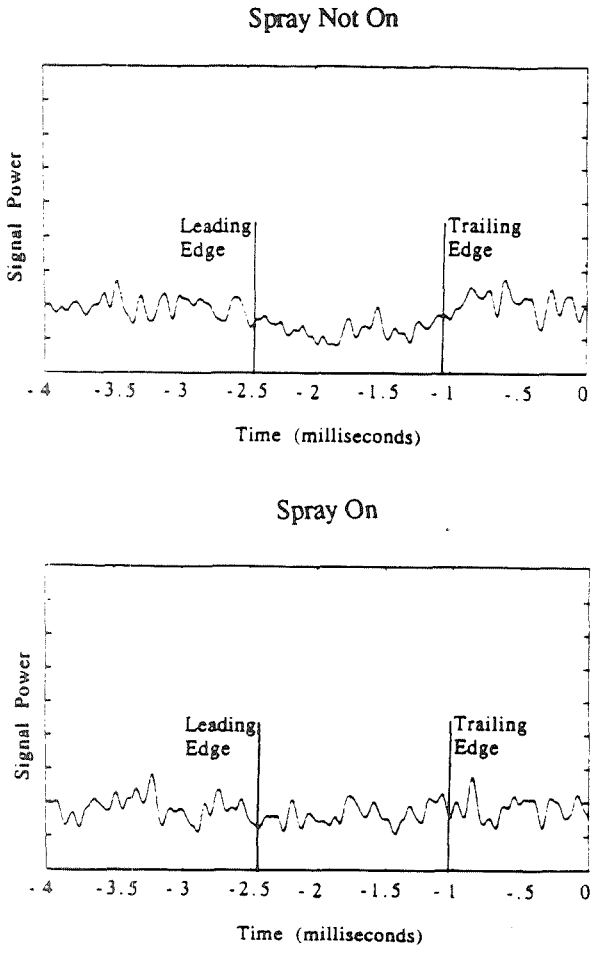


Fig. 16. IR Profiles at T=-0.4°C with and without water spray

4. Implementation Issues

While the above experiments indicate the potential of remote IR ice detection there are still several implementation issues to be considered. The rotor model did not include blade flapping or significant blade bending. These effects, in a helicopter application, will make it more difficult to track the blade location, and to keep it in focus. This may be addressed by use of a wide depth of field, or use of an auto focus system referenced to the blade leading edge.

Another issue to be considered is placement of the pyrometer on the helicopter. It will be necessary to mount the pyrometer so that it can view the blades, or possibly the tail rotor. Although small amounts of frost on the lens did not significantly degrade the signal in the rotor tests, it will be necessary to prevent the lens of the pyrometer from icing, frosting or becoming wet.

Signal processing is also an implementation issue, although it should be fairly straightforward. The signal must be amplified and synchronized with the rotor. It must be filtered, and then the profile compared to known icing conditions to determine the presence of ice. A simple leading edge to trailing edge temperature difference method may prove adequate. This data must be presented to the pilot in a useful format.

5. Conclusion

The results from preliminary IRT experiments, and the follow up rotor tests indicate that passive infrared ice detection can be used to remotely detect ice accretion on airfoils and rotors by measuring the characteristic chordwise temperature profiles which result from the release of latent heat during freezing. The technique has promise as a technique for remotely monitoring the icing of helicopter rotor blades. A prototype, suitable for flight testing, has been developed and successfully tested on a small scale rotor.

Acknowledgements

This work was supported by the National Aeronautics and Space Administration under grant NGC 3-927. The authors would like to thank the B.F. Goodrich Company and the Douglas Aircraft Company for their cooperation during the IRT tests and Professor Ken Dewitt and Alan Yaslick for providing thermocouple data.

References

1. Heinrich, A. Ross, R. Zumwalt, G., Provorse, J., and Padmanabhan, "Aircraft Icing Handbook - Vol. I", DOT/FAA/CT - 88/8-1, March 1988.

2. Corley-Byrne, P.L., "Helicopter Icing and It's Measurement", Presentation at SAE subcommittee meeting AC-9C - Aircraft Icing Technology, October 1986.

3. Hansman, R., Yamaguchi, K., Berkowitz, B., and Potapczuk, M., "Modelling of Surface Roughness Effects on Glaze Ice Accretion", AIAA-89-0734, AIAA 27th Aerospace Sciences Meeting, January 1989.

4. Henry, R. and Guffond, D., "Application de la Thermographie Infrarouge a l'Interpretation d'Essais dans une Soufflerie Givrante", Societe Francaise des Thermiciens, January 1989.

5. Skebe, S., "Synchronous Thermography", 3rd Japan-China joint conference on Fluid Machinery, April 1990

Deterministic Multi-Zone Ice Accretion Modeling

K. Yamaguchi* and R.J. Hansman, Jr.**
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA

M. Kazmierczak†
Department of Mechanical, Industrial, and Nuclear Engineering
University of Cincinnati
Cincinnati, OH

Abstract

A deterministic surface roughness transition model was proposed and implemented in the Multi-Zone version of the LEWICE ice accretion prediction code for glaze ice conditions. The transition model links the smooth/rough surface roughness transition with the laminar to turbulent boundary layer transition. The initial transition location is determined by boundary layer stability in the smooth laminar region. The subsequent upstream migration of the transition region is caused by a surface water bead formation process and controlled by the surface water flux out of the laminar/smooth region. The effectiveness of the surface roughness transition model was tested by comparing predictions of the deterministic Multi-Zone LEWICE code with experimental ice accretions for cylinder and airfoil geometries. Good agreement was achieved between the deterministic Multi-Zone LEWICE predictions and the experimental ice shapes. Some numerical difficulties were encountered due to limitations of the flow field code used in LEWICE. The glaze ice shapes were found to be sensitive to the laminar surface roughness and bead thickness parameters which controlled the transition location. The ice shapes were found to be insensitive to the turbulent surface roughness.

* Research Assistant, Student Member AIAA
** Associate Professor, Associate Fellow AIAA
† Assistant Professor, Member AIAA

1. Introduction

1.1. Background

Glaze icing presents the most difficult challenge for aircraft ice accretion modeling. In glaze icing, which normally occurs at temperatures near freezing or at high liquid water contents, there is insufficient convective heat transfer to remove all of the latent heat of freezing of the impinging supercooled water droplets. Consequently, the local ice accretion rate is controlled by the local convective heat transfer.

The local convective heat transfer from a surface is known to be strongly dependent on the ice surface roughness and the boundary layer behavior.[1,2] Because of the importance of heat transfer on the ice accretion rate, the surface roughness and it's influence on boundary layer transition becomes an important factor in modelling glaze ice accretion. Current analytical models such as LEWICE generally assume that the surface roughness is uniform, and the effective sand grain roughness, k_s, is used as an input parameter in the code.[3] The magnitude of the roughness parameter is normally determined empirically by comparison to experimental ice accretions. The erratic performance of glaze ice accretion models and the empirical manner in which the surface roughness is treated indicate the need for a more deterministic treatment of the surface roughness.

1.2. Experimental Observations

In prior experiments, detailed observations of accreting ice surface roughness were made at several icing facilities.[4] In one series of experiments conducted at the B.F. Goodrich Ice Protection Research Tunnel, a set of well defined ice shapes were recorded using high magnification microvideo cameras and a laser light sheet to accurately identify the plane of focus and the surface roughness.

These detailed photographic analyses of the accreting ice surfaces, along with the pioneering work by Olsen,[5] have revealed regions with distinct surface

roughness characteristics. The different surface roughness regions include a smooth region centered about the stagnation line, where surface water flow was observed for some cases, and a rough region downstream of the smooth region. The location of smooth to rough zones as a function of time have been studied in detail using a laser light sheet, which enabled an accurate observation of surface roughness and the behavior of the smooth to rough transition point.[2] In general, the rough to smooth transition point is observed to propagate towards the stagnation region with time. These observations are consistent with a model which couples the formation of water beads on the ice surface with the rough/smooth transition and the laminar/turbulent boundary layer transition.

1.3. Multi-Zone Model

Based on the experimental observations of ice formation in glaze ice regime, a Multi-Zone model, in which the accreting ice surface is divided into two or more discrete zones with varying surface roughness and water behavior, was proposed.[6]

In the simplest version of the Multi-Zone model, the surface is divided into two zones, the smooth zone and the rough zone. In the smooth zone, corresponding to the region observed to be centered about the stagnation point, the surface is smooth and uniformly wet with thin water film. Here, the convective heat transfer is insufficient to remove all of the latent heat of freezing, and surface water runback has been observed, indicating that the simple water runback model used in the original LEWICE implementation may be valid.[7]

In the rough zone, corresponding to the region found downstream of the smooth zone, the surface was observed to be considerably rougher. Here, the heat transfer is enhanced due to the increased roughness and the turbulent boundary layer. Experimental observations have indicated that in many cases there is no water runback in the rough zone.

1.4. Initial Implementation of the Multi-Zone Model in LEWICE Ice Accretion Prediction Code

In order to evaluate the effectiveness of the Multi-Zone model, a simple two zone version of the model was implemented in the LEWICE code, based on the hypothesis that the smooth/rough transition location coincided with the laminar/turbulent boundary layer transition location. Because no deterministic model for the propagation of the transition point was available, the transition location between the smooth and rough regions was input to the code from the experimental data in the preliminary implementation.

Comparisons between the preliminary Multi-Zone model with the original LEWICE code and experimental ice shapes indicated that significant improvements in glaze ice accretion prediction could be obtained with the Multi-Zone model using rough to smooth transition data from experimental observations.[6] These results indicated the need for a deterministic model of the rough to smooth transition behavior which could be incorporated in the Multi-Zone LEWICE code.

2. Deterministic Surface Roughness Transition Model

In order to implement the deterministic surface roughness model it was necessary to develop criteria for the initial location and subsequent propagation of the boundary between the smooth and rough zones. This smooth/rough transition has been investigated experimentally in several icing test facilities.[6,8] Based on the experimental studies a model of the surface roughness transition behavior of glaze ice has been proposed.

2.1. Initial Smooth/Rough Transition Location

Experimental observations indicated that the Reynolds number dependence of the *initial* location of surface roughness transition location correlated with the Reynolds number dependence of the boundary layer transition location. Therefore, it was hypothesized that the initial surface roughness transition is caused by the boundary layer transition. In this case, the initial smooth/rough transition will occur at the laminar/turbulent boundary layer transition point. Physically, it is thought that the low heat transfer in the laminar region causes a low freezing fraction which results in a relatively smooth surface water film. In the turbulent region, however, the increased heat transfer causes rapid freezing resulting in higher surface roughness.

It should be noted that the boundary layer transition location depends on the stability both of the laminar boundary layer which is influenced by the Reynolds number and of the stagnation region surface roughness. Increasing either tends to move the transition point toward the stagnation point. While the stagnation region is generally observed to be smooth, waves in the water film or small scale roughness can influence the initial location of the laminar/turbulent and hence the smooth/rough transition location.

2.2. Bead Formation and Propagation of the Transition Location

As discussed above, the smooth/rough transition point is typically observed to propagate toward the stagnation region with time. It has also been observed that higher surface water flow causes the rate of movement of the transition to increase.[8] Based on the

smooth/rough transition behavior and detailed observations of the surface water, it is hypothesized that the enhanced heat transfer in the turbulent region causes sufficient freezing to partially dry the surface and causes bead formation of unfrozen water which flows downstream from the laminar region. The bead formed at the interface of smooth and rough regions results in the increased roughness observed in the rough zone.

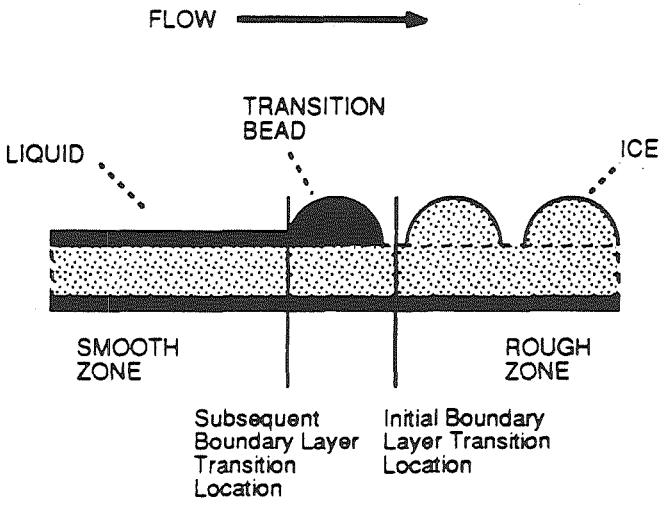


Fig. 1. Bead formation and boundary layer transition location

The mechanism for the propagation of the transition location is depicted in Fig. 1. The initial smooth/rough transition occurs because of the boundary layer transition as discussed above. As water flows from the upstream laminar region, a water bead is formed at the interface. The increased roughness of the bead causes the boundary layer to "trip" and transition to turbulent at the upstream edge of the bead. The, now turbulent, boundary layer enhances the heat transfer in the region where the bead has formed. With the increased heat transfer, the bead freezes forming rough ice. As the surface dries, beads begin to form further upstream. This process will repeat and the transition point will propagate toward the stagnation region as observed experimentally. By increasing the surface water flux, the rate of formation and growth of the interfacial beads are increased. This causes the observed increase in upstream propagation of the coupled rough/smooth-laminar/turbulent transition point with increasing surface water flux.

In some cases, splashing of impinging water drops has been observed to occur in the stagnation region. The splashing is thought to be limited to the thin water film in the smooth region. If significant splashing occurs, it will reduce the surface water flux into the transition region thereby slowing the propagation of transition.

2.3. Implementation in LEWICE Ice Accretion Prediction Code

Based on the above hypothesis that smooth/rough transition location coincides with the laminar/turbulent boundary layer transition location, a more physically realistic Multi-Zone model was implemented in the LEWICE code. Initially, a simple two-zone version of the Multi-Zone model was implemented through boundary layer transition, where the smooth and rough zones are considered to coincide with the regions of laminar and turbulent boundary layers.

For the initial time step, equivalent roughness element heights for both the laminar and turbulent zones are required as inputs. The laminar roughness height k_l, corresponding to the waviness of the laminar region water layer, is used to calculate a critical Reynolds number based on surface roughness size. This critical Reynolds determines the initial location of boundary layer transition.[2,9] Increasing the laminar region roughness will move the transition point forward. The turbulent roughness k_t is used to calculate the convective heat transfer coefficient which increases with roughness. By using two roughness element sizes, one for laminar region and one for the turbulent region, it is possible to control the transition location and turbulent heat transfer independently. The laminar roughness k_l is then used to determine the laminar to turbulent boundary layer transition location. The turbulent roughness k_t is used to calculate the heat transfer coefficient in the turbulent region.

This method more closely emulates the physical situation where two different roughnesses have been observed in the smooth and rough zones. In the laminar region, the roughness height corresponds to the roughness of the uniform water film. In the turbulent region, it corresponds to the roughness size observed in the rough region.

In the subsequent time steps, the location of smooth/rough transition was calculated based on the physical model depicted in Section 2.2. By assuming all the surface water mass flux is deposited into a bead at the boundary layer transition location, a relationship can be developed between surface water mass flux and the size of the bead formed. From mass conservation,

$$x = \frac{\dot{m}\,\Delta t}{\rho_\omega b}$$

where x is the distance covered by the bead along the surface, b is the bead thickness, $\dot{m}$ is the mass flux calculated by LEWICE, Δt is the time step used to run LEWICE, and ρ_ω is the density of water. With this

method, the bead thickness b, or the size of the bead, must be input. Based on the experimental observations of the surface roughness transition location, the thickness of the water film was estimated to be about 2 mm for the cases considered. An example of the transition propagation calculated using this method is compared with experimentally observed data in Fig. 2 for a 1" diameter cylinder at a temperature of -7°C, velocity of 64.3 m/s, LWC of 0.8 g/m^3, and MVD of 12 microns. It was found that for this particular case, the transition was well modeled using this method.

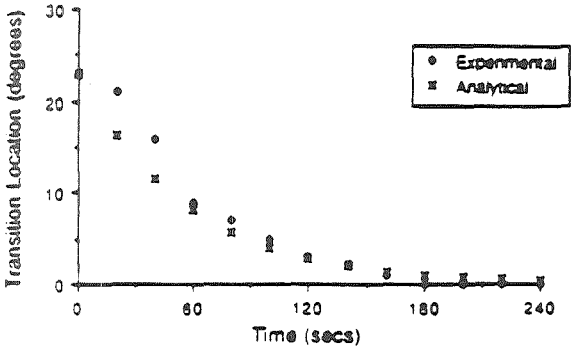


Fig. 2. Experimental and analytical surface roughness transition location on a 1" diameter cylinder with an assumed bead thickness of 2mm T = -7°C, V = 64.3 m/s, LWC = 0.8 g/m^3, MVD = 12 microns)

Because small time steps are used for some runs with LEWICE, a feature was added to the surface roughness transition model, where the surface runback water was allowed to "collect" at the interface for more than one time step, until there was enough surface water to form a bead. Once enough water has "collected" at the interface, the bead is added as frozen mass, and the boundary layer transition is set upstream of the bead.

3. Ice Shape Comparisons on Cylinders

In order to verify the effectiveness of the deterministic surface roughness transition model , ice accretion predictions for simple cylinders calculated by Multi-Zone LEWICE were compared to experimental ice shapes.

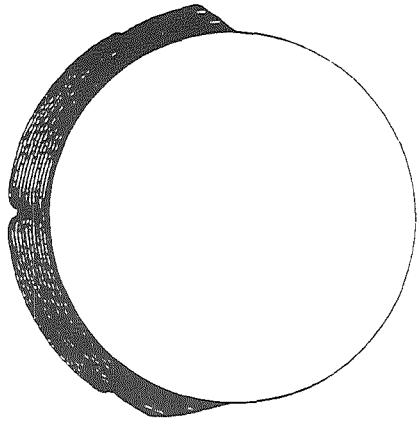
3.1. Initial Implementation of the Deterministic Model

An example of the initial implementation of the deterministic surface roughness transition model is shown in Fig. 3 for a 1" diameter cylinder at a temperature of -7°C, a velocity of 64.3 m/s, a LWC of 0.8 g/m^3, and a MVD of 12 microns. Based on observations of the surface roughness, the laminar roughness was set to k_l = 0.1mm, the turbulent roughness was set to k_t = 1mm, and the bead thickness was set to b = 2mm. In order to reduce the computational requirements during development of the transition model, the flow field was not recalculated for each time step. However, the ice accretion, heat transfer, and surface roughness components were updated at 20 second intervals. From Fig. 3 it can be seen that the Multi-Zone code with the deterministic surface roughness transition model is a significant improvement over the original Single-Zone LEWICE prediction which is shown in Fig. 3 for reference.

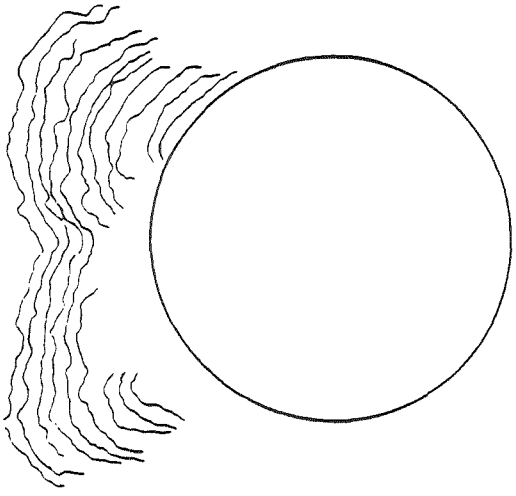
3.2. Effects of Flow Code Updates

Based on the positive initial results of the surface roughness transition model in the Multi-Zone code, a set of more complete runs were conducted with full updating of the flow field every time step. A number of difficulties were encountered in this process due primarily to numerical problems and limitations in the simple inviscid panel flow code used in LEWICE. One difficulty which occurred was a numerical instability resulting from the strong coupling between the flow velocity at the surface and the convective heat transfer. Due to the nature of the paneling method, the flow velocity at the surface exhibits high velocity spikes. Because the heat transfer increases with velocity, these spikes result in enhanced ice growth regions. In subsequent time steps, the flow is accelerated around these growths resulting in even further growth and instability. In the physical situation, regions of enhanced growth or roughness may occur but will not result in unstable growth due to lateral surface heat transfer (i.e. between adjacent elements) which is ignored in LEWICE. Lateral heat transfer would result in smoothing the heat transfer spikes. This was resolved in the model by including a smoothing routine in the local heat transfer calculation.

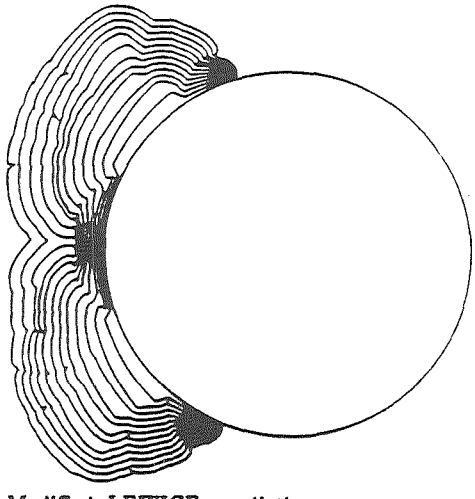
An example of the predicted ice accretion with flow updating and heat transfer smoothing is shown in Fig. 4 for the same conditions simulated in Fig. 3. It can be seen that some irregularity due to local heat transfer enhancement is still present and that the growth rate in the stagnation region is not as high as the experimentally observed case. The low stagnation growth rate is thought to be due to an underestimation of the flow velocity at the surface in the stagnation region after horn formation. Again, the inviscid panel code does not accurately reflect the flow in the concave stagnation region and the strong coupling between surface flow velocity and the local heat transfer coefficient results in a lower accretion rate. It should be noted that the local heat transfer coefficient is much more strongly dependent on the details of the flow near



a. Original LEWICE prediction.



b. Experimental ice shape.



c. Modified LEWICE prediction.

Fig. 3. Comparison of experimental and predicted ice shapes for a 1" cylinder at 20 sec. intervals. (T = -7°C, V = 64.3 m/s, LWC = 0.8 g/m^3, MVD = 12microns, k_l = 0.1mm, k_t = 1mm, and b = 2mm)

the surface that the local droplet impingement efficiency for cloud size droplets.

Based on a comparison between Figs. 3 and 4, it appears that while both simulations do an adequate job, the single flow field case more closely replicates the experimentally observed ice accretion. However, care must be exercised in drawing conclusions from such comparisons between experimental and simulated ice accretions due to uncertainties in the experimental icing parameters and the limited number of cases available for comparison.

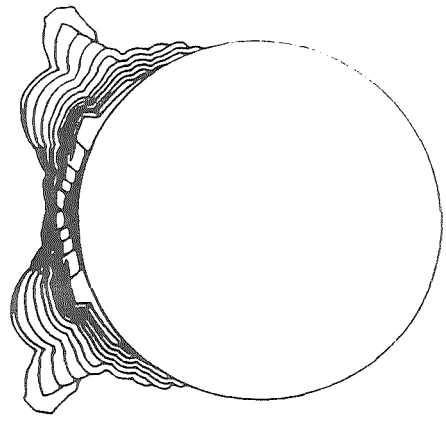
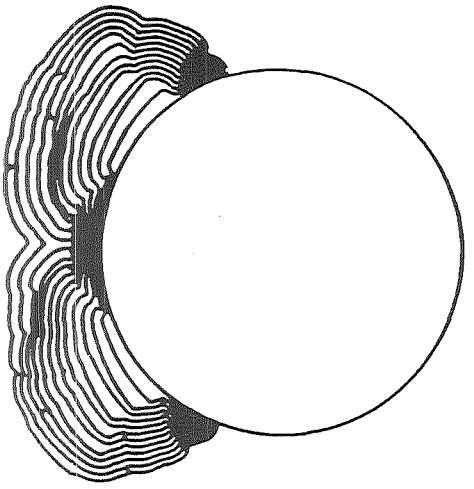


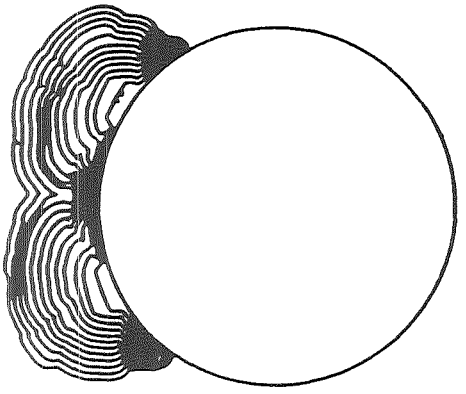
Fig. 4. Example of predicted ice accretion with flow updating for a 1" cylinder. (T = -7°C, V = 64.3 m/s, LWC = 0.8 g/m^3, MVD = 12 microns, k_l = 0.1mm, k_t = 1mm, and b = 2mm)

3.3. Effects of Other Factors

The effects of other input parameters on the predicted ice shapes were studied. It was found that increasing the thickness parameter b determines decreased the rate at which the transition location propagated toward the stagnation point. The turbulent roughness size k_t was found to only weakly influence the ice shape. This can be seen in Fig. 5, where the turbulent roughness size is increased from 1.35mm to 2.5mm for the conditions run in Section 3.1. Increasing the turbulent roughness size slightly increases the ice accretion in the rough region and limits the angular extent of the horns, but the overall ice shapes are remarkably similar. These factors are discussed more in detail in the next section, with examples of ice accretions on an airfoil.



(a) k_t = 1.35mm



(b) k_t = 2.5mm

Fig. 5. Comparison of ice accretions for two values of turbulent roughness size (T = -7°C, V = 64.3 m/s, LWC = 0.8 g/m^3, MVD = 12 microns, k_l = 0.1mm, and b = 2mm)

4. Ice Shape Comparisons on Airfoils

The effectiveness of the deterministic Multi-Zone model for an airfoil geometry was tested on a NACA 0012. Runs were conducted to investigate the effect of the input parameters; namely, the laminar roughness size k_l, the turbulent roughness size k_t, and the bead thickness parameter b, on the predicted ice shapes. A series of 12 simulations were performed on a 30 cm chord NACA 0012 airfoil at 0° angle of attack. The icing conditions were the same in all of these runs: V = 29 m/s, LWC = 0.5 g/m^3, T_{amb} = 260.55°K, and MVD = 20 microns. These conditions correspond to the Example 1 from the LEWICE User's Manual[9] which is based on the 120 sec. experimental ice accretion measured by Gent et al.[10] and is shown in Fig. 6. The ice profiles and flow fields were calculated every 20 seconds and the simulations terminated at two minutes.

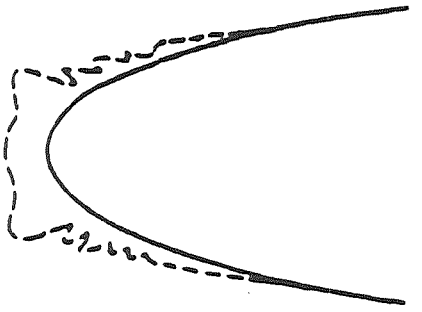
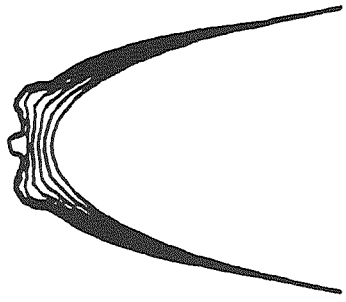


Fig. 6. Experimental ice accretion on a 30 cm chord NACA 0012 (α = 0°, T = -12.6°C, V = 129 m/s, LWC = 1 g/m^3, and MVD = 12 microns)

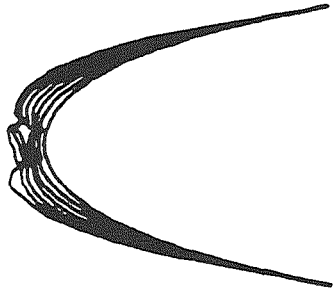
4.1. Effect of the Laminar Roughness Size

In order to investigate the effect of the laminar roughness size, k_l, was varied between 0.05mm to 0.35mm in the three runs, shown in Fig. 7. The other parameters were held fixed, with k_t = 0.35 and b = 1.0 mm. As shown in Fig. 1, the predicted ice shapes were strongly affected by the laminar roughness size. The simulation with the largest value of laminar roughness (Fig. 7a) produced the most "horn-shaped" ice structure (characteristic of glazed ice formations) while the runs with smaller value of laminar roughness (Figs. 7b, 7c) produced rougher, more "rime-type" ice accretions.

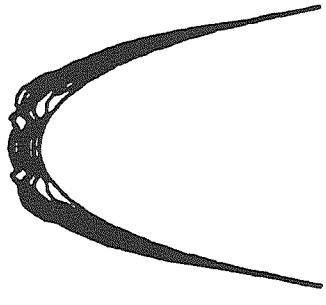
The difference in final ice profiles is attributed to changes in the *initial location* and the *subsequent motion* of the boundary layer transition point. It was found that the larger the laminar roughness size, the closer was the initial transition point to the stagnation region. For example, for a laminar roughness of 0.05 mm (Fig. 7c), the initial boundary layer transition point occurred at a distance seventeen segments (approximately 8.5 cm) downstream from the stagnation point and moved upstream to the ninth segment (approximately 4.5 cm) by T = 120 seconds. For the case with a laminar roughness of 0.10 mm (Fig. 6b), the boundary layer transition point started at segment 7 (approximately 3.5 cm) and moved to the stagnation region by t = 100 sec.). For Fig. 7a, having the largest value of the laminar roughness, virtually all of the airfoil surface was turbulent from the initial time step. It is believed that the increase in heat transfer in the turbulent region coupled with impingement effects causes greater ice growth and accounts for the horned structure.



(a) k_t = 0.35mm



(b) k_t = 0.10mm

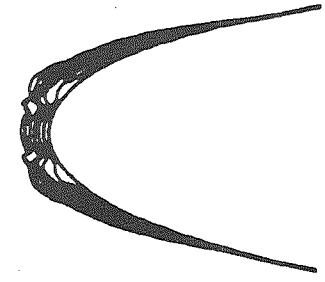


(c) k_t = 0.05mm

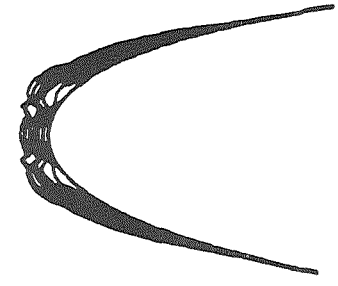
Fig. 7. Laminar roughness size influence on 120 sec. ice profiles

4.2. Effects of Turbulent Roughness Size

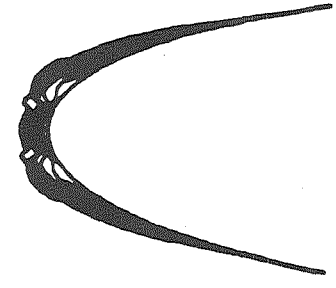
The effect of turbulent roughness size on ice shapes was investigated in the runs shown in Fig. 8. The laminar roughness and bead thickness parameter were the constant for all three cases, with values of 0.05 mm and 1.0mm, respectively, but the turbulent roughness was increased (k_t = 0.55mm) in Fig. 8b, and decreased (k_t = 0.15mm) in Fig. 8c, relative to the value used in Fig. 8a. Comparing the predicted ice profiles at t = 120 sec. shows that, as in the cylinder example in Section 3.3, the turbulent roughness size did not have much effect on the ice shapes. It is important to note that the initial transition location and the transition point motion were practically identical for all of these runs. The ice profiles are similar because all the water impinging on the surface in the turbulent region was



(a) k_t = 0.35mm



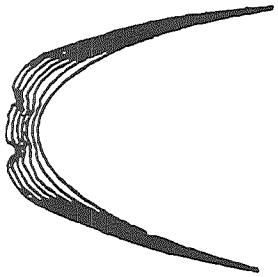
(b) k_t = 0.55mm



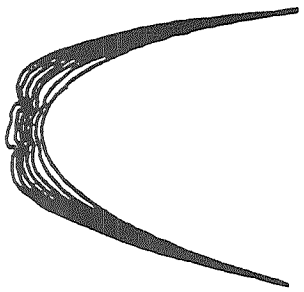
(c) k_t = 0.15mm

Fig. 8. Turbulent roughness size influence on 120 sec. ice profiles

frozen on impact. Therefore, although the larger roughness size increased the turbulent heat transfer, there was enough heat transfer, even with the smaller roughness sizes, to remove all of the latent heat of freezing. This is consistent with experimental observations at similar conditions where no surface water runback was found in the rough regions.[5] It also indicates that since all of the impinging water is frozen in the rough region, the amount of impinging water determines the amount of ice accreting in this region.



(a) k_t increases 0.1mm every 20 sec.



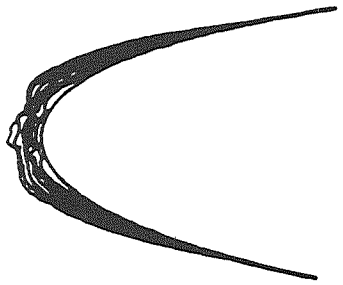
(b) k_t increases 0.2mm every 20 sec.

Fig. 9. Effect of increasing turbulent roughness size with time on 120 sec. ice profiles

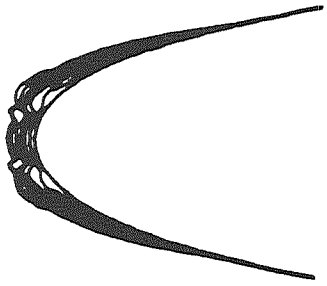
Further evidence that indicates that the turbulent roughness has little impact on the resulting ice shape is shown in Fig. 9. These runs do not employ the deterministic model but, rather, are based on the simpler Multi-Zone model where the transition movement was controlled manually. The laminar roughness and the transition movement were identical in Figs. 9a and 9b, but the turbulent roughness differed. Initially, the turbulent roughness at the beginning of each simulation had the same value for both cases but changed with time at different rates. The turbulent roughness in Fig. 9a increased 0.1mm every 20 seconds but was ramped at twice this rate in Fig. 9b. Despite the differences in turbulent roughness, the resulting ice profiles are very similar.

4.3. Effects of the Bead Thickness Parameter

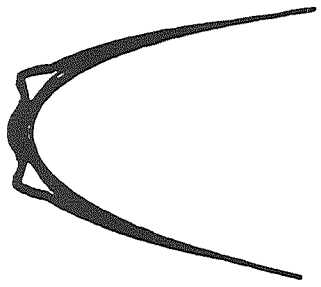
The final parameter investigated in this study was the bead thickness parameter b. The bead thickness is related to the size of the bead believed to be formed at the interface between the smooth and rough zones as discussed in Section 2.2. Its value determines the rate at which the transition location propagates toward the



(a) Bead thickness parameter = 0.5mm



(b) Bead thickness parameter = 1.0mm



(c) Bead thickness parameter = 2.0mm

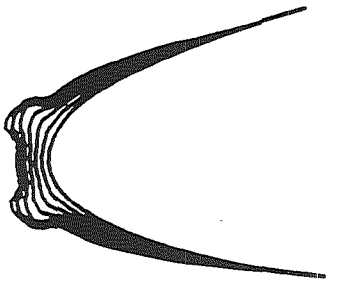
Fig. 10. Bead thickness parameter influence on ice profiles

stagnation point based on the amount of surface water flux. The effect of the thickness parameter on the ice shape was investigated in Figs. 10a, 10b, and 10c using the deterministic model with bead thickness values of 0.5mm, 1.0mm, and 2.0mm respectively. The laminar roughness and turbulent roughness were fixed and set to the same values used in Fig. 7c (k_l = 0.05mm, k_t = 0.35mm). Figure 10 compares the ice profiles at the same time and shows that the bead thickness parameter has major impact on the ice shape. (Note: because of a numerical instability, the Fig. 9c run was terminated at 100 seconds).

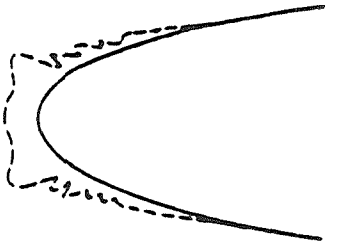
The profile changes drastically as the thickness parameter changes. A rime-type conformal ice accretion forms for the run with smallest value of bead thickness parameter (Fig. 10a), whereas horns clearly appear in the simulation with the largest bead thickness parameter (Fig. 10c). Since the initial transition location is the same in all of these runs (same laminar roughness), the difference in final ice shape profile is attributed to the *rate* at which the transition point propagates towards the stagnation point. The rate of propagation varies inversely with the bead thickness parameter. For thickness parameter = 0.5mm (Fig. 10a), the transition point moves upstream very quickly with time to the stagnation region. For thickness parameter 1.0mm (Fig. 10b), the transition point also moves continuously upstream with time but at a slower rate. For the largest value of thickness parameter, 2.0mm, the transition point has very little movement.

4.4 Experimental Comparison

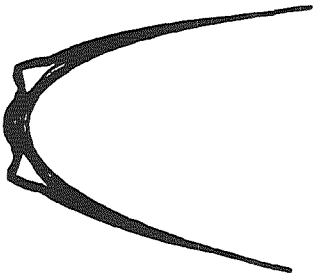
In Fig. 11, the experimental ice shape is compared with the original Single-Zone LEWICE and the Multi-Zone LEWICE result from the case run in Fig. 9c. It can be seen that the Multi-Zone LEWICE reproduced the experimental ice shape somewhat better than the original LEWICE code. It should be noted that this case was used to calibrate the effects sand grain roughness k_s and was Example 1 in the original LEWICE manual. It is considered one of the best cases of agreement between the original Single-Zone LEWICE and an experimentally measured glaze ice accretion. Thus, it appears that the Multi-Zone model works well on airfoil geometries with a more complete physical model and more realistic values of the laminar and turbulent surface roughness. Further comparisons are needed, however, over a range of different icing conditions to establish the robustness of the method.



(a) Original Single-Zone LEWICE



(b) Experimental Accretion



(c) Multi-Zone LEWICE

Fig. 11. Ice shape comparison for 30 cm chord NACA 0012

5. Conclusions

A deterministic surface roughness transition model was proposed and implemented in the Multi-Zone version of the LEWICE ice accretion prediction code for glaze ice conditions. The transition model links the smooth/rough surface roughness transition with the laminar to turbulent boundary layer transition. The initial transition location is determined by boundary layer stability in the smooth laminar region. The subsequent upstream migration of the transition region is caused by a surface water bead formation process and controlled by the surface water flux out of the laminar/smooth region. The effectiveness of the surface roughness transition model was tested by comparing predictions of the deterministic Multi-Zone LEWICE code with experimental ice accretions for cylinder and airfoil geometries. Good agreement was achieved between the deterministic Multi-Zone LEWICE predictions and the experimental ice shapes. Some numerical difficulties were encountered due to limitations of the flow field code used in LEWICE. The glaze ice shapes were found to be sensitive to the laminar surface roughness and bead thickness parameters which controlled the transition location. The ice shapes were found to be insensitive to the turbulent surface roughness.

Acknowledgements

This work was supported in part by the National Aeronautics and Space Administration and the Federal Aviation Administration under Grants NAG-3-666 and NGL-22-009-640. The work was also supported by the National Science Foundation Presidential Young Investigators Program, Award No. 8552702. Michael Kazmierczak was supported by the Ohio Aerospace Institute Investigators Program. The authors wish to thank Mark Potapczuk for his technical contributions.

References

1. Achenbach, E., "The Effect of Surface Roughness on the Heat Transfer from a Circular Cylinder to the Cross Flow of Air," International Journal of Heat and Mass Transfer, Vol. 20, 1977, pp. 359-369.

2. Yamaguchi, K. and Hansman, R.J., "Improved Ice Accretion Prediction Techniques Based on Experimental Observations of Surface Roughness Effects on Heat Transfer," MIT Aeronautical Systems Laboratory Report, ASL-90-5, 1990.

3. MacArthur, C.D., "Numerical Simulation of Airfoil Ice Accretion," AIAA Paper 83-0112, Jan. 1983.

4. Hansman, R.J., Yamaguchi, K., Berkowitz, B., and Potapczuk, M., "Modelling of Surface Roughness Effects on Glaze Ice Accretion," AIAA Paper 89-0734, Jan. 1989.

5. Olsen, W., and Walker, E., "Experimental Evidence for Modifying the Current Physical Model for Ice Accretion on Aircraft Surfaces," NASA TM-87184, 1986.

6. Yamaguchi, K. and Hansman, R.J., "Heat Transfer on Accreting Ice Surfaces," AIAA Paper 90-0200, Jan. 1990.

7. Messinger, B.L., "Equilibrium Temperature of an Unheated Icing Surface as a function of Airspeed," Journal of the Aeronautical Sciences, Jan. 1953, pp.24-42.

8. Hansman, R.J., and Turnock, S.R., "Investigation of Surface Water Behavior During Glaze Ice Accretion," AIAA Paper 88-0015, Jan. 1988.

9. Ruff, G.A. and Berkowitz, B.M., "Users Manual for the NASA Lewis Ice Accretion Prediction Code (LEWICE)," NASA C. R. 185129, 1990, p. 116.

10. Gent, R.W., Markiewicz, R.N., and Cansdale, R.T., "Further Studies of Helicopter Rotor Ice Accretion and Prediction," Paper 54, 11th European Rotorcraft Forum, Sept. 1988.

Alert Generation and Cockpit Presentation for an Integrated Microburst Alerting System

Craig Wanke* and R. John Hansman, Jr.**
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts USA

Abstract

Alert generation and cockpit presentation issues for low-level wind shear (microburst) alerts are investigated. Alert generation issues center on development of a hazard criterion which allows integration of both ground-based and airborne wind shear detection systems to form an accurate picture of the aviation hazard posed by a particular wind shear situation. A methodology for testing of hazard criteria through flight simulation has been developed, and has been used to examine the effectiveness and feasibility of several possible criteria. Also, an experiment to evaluate candidate graphical cockpit displays for microburst alerts using a piloted simulator has been designed.

Nomenclature

AD	approach degradation parameter (Eqn. 1)
C_L	aircraft lift coefficient
C_D	aircraft drag coefficient
d	glideslope indicator deviation
D	aircraft drag
F	F-factor (Eqn. 3)
g	gravitational acceleration
h	altitude
J	microburst impact parameter (Eqn. 2)
L	aircraft lift
m	aircraft mass
T	aircraft thrust
V	airspeed
W_x	horizontal wind velocity (tailwind positive)
W_h	vertical wind velocity (updraft positive)
x	groundtrack distance
α	angle of attack
γ	wind-relative flight path angle
ρ	correlation coefficient

* Research Assistant
** Associate Professor, Associate Fellow AIAA

1. Introduction

Low altitude wind shear is the leading weather-related cause of fatal aviation accidents in the U.S. Since 1964, there have been 26 accidents attributed to windshear resulting in over 500 fatalities.[1,2] The localized intense downdrafts known as microbursts are the most dangerous form of wind shear, and pose a serious hazard to aircraft during takeoff or approach. For this reason, wind shear hazard detection and alerting system development is a very active area of research (Fig. 1). The variety of systems under development for wind shear detection and ground-to-air datalink, combined with the advent of electronic cockpit instrumentation, allow many options for implementation of an integrated wind shear warning system. The development of the optimal system configuration requires work in two major areas: (1) alert generation methodology and (2) information transfer and presentation to flight crews.

2. Background and Problem Definition

2.1 Alert Generation

Several low-level wind shear detection systems, both ground-based and airborne, are currently operational or under development. The Terminal Doppler Weather Radar (TDWR) system is currently under final stages of development and is scheduled for initial deployment in 1992. The Low Level Windshear

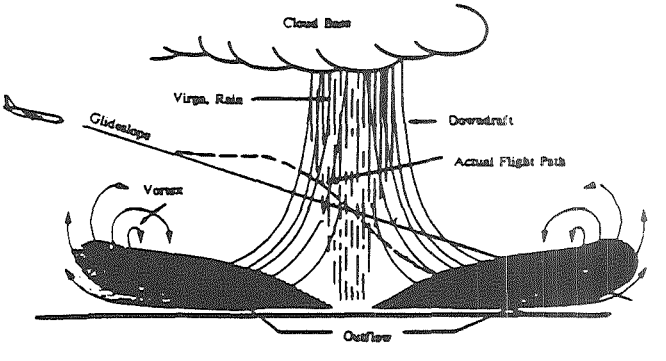


Figure 1. Aircraft encounter with a microburst on approach

Alert System (LLWAS), a system of on-airport anemometers, is currently operational at many U.S. airports and is being expanded in some locations and integrated with TDWR in others. Also operational are airborne reactive systems, which provide warnings after a wind shear has been encountered. Several airborne look-ahead systems, including infrared, Doppler weather radar, and Doppler laser radar (lidar) systems are in the developmental stage. In the near future, almost any combination of the above systems may be available in a given situation.

As a result, the first task for alert generation is to combine data from several systems, which use different measurement techniques, into an estimate of the hazard posed by a particular wind shear event. If the detected hazard exceeds a pre-determined hazard threshold, an alert must be generated and disseminated to the affected aircraft. For safety reasons, the number of missed alerts must be minimized. For pilot confidence in the alerting system to remain high and for minimal disruption of airport operations, the alert should only occur in cases where the wind shear truly poses a threat. One essential element necessary to satisfy these requirements is accurate assessment of the wind shear hazard.

Accurate assessment of wind shear hazards, in turn, requires use of an effective "hazard criterion." The hazard criterion should be a quantity which can be directly measured or inferred from available measurements, and which accurately reflects the danger to aircraft posed by the low-level wind shear event being measured. An example of a microburst hazard criterion (currently used by the TDWR system) is "total divergence," which equals the maximum headwind change that could be experienced by an aircraft flying directly through the microburst. Due to the variety of different wind shear sensors available, accuracy is not the only requirement for good hazard criteria. To allow accurate combination ("fusion") of data from multiple sensors, a balance must be struck between 1) the most easily measured and accurate hazard criterion for each sensor used, and 2) use of the fewest different hazard criteria possible to make data fusion easier.

2.2 Information Transfer and Presentation

Once an alert has been generated, timely and effective presentation of the alert to the flight crew is required. A severe microburst event which occurred during the 1988 TDWR evaluation at Denver Stapelton airport illustrates this point.[3,4] On July 11, 1988 several microbursts of increasing intensity were detected by the TDWR system. Alerts were generated and disseminated via standard radio communications to five aircraft on approach. Only one of these aircraft aborted the approach prior to penetrating the area of microburst activity. The others, despite the alerts, did not recognize the threat until after penetration of the microbursts, resulting in one near accident and in all of the aircraft making missed approaches. Even though the alerts were accurate and produced well in advance of the approaches, verbal dissemination of the alerts proved ineffective and the advance warning was lost.

The message from this event is that alert dissemination and presentation needs to be carefully designed to allow effective alerting. Two new technologies will aid this process: 1) digital ground-to-air datalinks such as the FAA's Mode-S system and 2) the electronic flight instrumentation systems (EFIS) in modern transport aircraft. Preliminary work has been done to compare the effectiveness of verbal, alphanumeric, and graphical modes of presenting uplinked wind shear alerts. Graphical alerts were found to result in improved decision-making, lower crew workload, and were preferred by flight crews.[5,6]

2.3 Research Focus and Approach

The work presented in this paper focusses on two of the major issues presented above: 1) wind shear hazard criteria, and 2) evaluation of prototypical graphical cockpit displays for integrated wind shear alerts. A methodology for testing of hazard criteria through flight simulation has been developed. Preliminary runs have been performed for a set of possible hazard criteria, including that currently used by the TDWR system. The values of the possible hazard criteria for a set of modeled microburst windfields have been correlated with the flight path degradation experienced by a simulated aircraft flying through those windfields to obtain an estimate of usefulness.

The prototypical cockpit displays for wind shear alerts will be tested with a piloted flight simulation study. The experimental design has been completed, and active line pilots are being solicited as subjects.

3. Evaluation of Microburst Hazard Criteria

3.1 Approach

The danger to aircraft posed by a microburst is not a directly measureable quantity. A microburst "hazard criterion" is a directly measureable quantity which provides an indication of the aviation hazard. Therefore, the assumption made for this experiment was that *a good hazard criterion should correlate well with the impact of a microburst on an aircraft trajectory.*

In order to evaluate hazard criteria based on this assumption, 1) a nominal microburst encounter situation, and 2) a quantitative measure of the effect of

a microburst on an aircraft trajectory were defined. The nominal condition was chosen to be a microburst encounter on final approach. In addition, it was assumed that the pilot or autopilot was unaware of the presence of the microburst, and hence attempted only to maintain the glidepath with a "normal" control strategy. This nominal case is used only to allow the comparison of hazard criteria under a consistent situation; it is intended that the results will be extended to include other situations such as takeoffs.

For the nominal case (approach), either dropping below the glideslope or loss of airspeed must be considered hazardous conditions. To quantify this, an "Approach Degradation" (AD) parameter (similar in concept to those used in References 7 and 8) was defined which combines both glideslope deviation and airspeed loss:

$$AD \equiv \sqrt{\left(\frac{d_{low}}{d_{full\ scale}}\right)^2 - \left(\frac{V_{ref} - V}{V_{ref} - V_{stall}}\right)^2} \qquad (1)$$

where d_{low} = "dots" below glideslope

In developing this measure, it was assumed that a critical situation develops if either (1) the aircraft drops two "dots" (0.7°) below the glideslope*, or (2) the aircraft's airspeed drops to the stall airspeed. Thus, either of these situations was defined as a "unit hazard." The root-mean-square of the fractional hazards due to falling below the glideslope or losing airspeed is therefore a point measurement of the approach degradation.

The angular error vs. glideslope measurement was chosen rather than altitude error to emphasize the greater danger due to being below glideslope at lower altitudes. Also, deviations above glideslope or velocity changes above trim are not included in the computation of AD; the terms inside the parentheses in the above equation can only take positive values.

To evaluate the impact of the overall windfield on the approach, this parameter is integrated over the entire approach to produce the overall "microburst impact" parameter J.

$$J \equiv \int [AD]\, dt \qquad \text{over entire approach} \qquad (2)$$

* Two "dots" of glideslope deviation means that the pilot's glideslope deviation indicator has reached full scale in the low direction, which is the equivalent of falling 0.7° below the 3° glideslope as measured from the runway threshold.

A series of aircraft simulation runs were performed, using this nominal condition and a series of high-resolution modeled microburst windfields. The value of the microburst impact parameter was computed for each run. A set of possible microburst hazard criteria were then computed for each of the modeled windfields. The hazard criteria were then evaluated by calculating the linear correlation coefficient between each of the hazard criteria and the microburst impact parameter for the complete set of simulation runs. Note that "correlate well" as used to define a good hazard criterion has been specialized to mean good linear correlation for the purposes of this analysis. Linear correlation is used because it is 1) desirable for simplicity in applying the hazard criterion and 2) easily evaluated by computing the correlation coefficient. Two complete sets of runs were performed, using the same windfields but different autopilot models, in order to show the insensitivity of this analysis to the control strategy employed.

3.2. Candidate Hazard Criteria

The following set of hazard criteria was chosen for examination in the simulation runs.

1) Total divergence, ΔU (headwind-to-tailwind shear across the event)
2) Mean shear, ΔU/ΔR, defined as total divergence divided by the shear distance (the distance between the headwind and tailwind peaks)
3) Peak F-factor (defined below)
4) Largest F-factor exceeded for a given distance (1500, 300, 450, 6000 foot lengths were tested)
5) Largest mean F-factor over a given distance (1500, 3000, 4500, 6000 foot lengths were tested)

The total divergence, as mentioned above, is the criterion currently used in TDWR operational evaluations for both alert thresholds and intensity reporting. A measurement of greater than 30 knots divergence by the radar triggers a microburst alert (with the divergence reported as an "X knot loss") and a measurement of greater than 20 but less than 30 knots is given as "wind shear with loss."[9] Mean shear is an alternative criterion, which could be easily computed from TDWR measurements; it is obtained by dividing the divergence value by the radial distance across which it is measured.

The "F-factor" hazard criterion, proposed by researchers at NASA Langley Research Center[10], is based on the impact of a microburst windfield on the total energy (kinetic plus potential) of the aircraft. It is a direct measure of the loss of potential rate-of-climb (or loss of effective thrust-to-weight ratio) due to the

immediate windfield. It is dependent on the time rate of change in the aircraft frame of the tailwind velocity, the vertical wind velocity, and the aircraft airspeed. A derivation of F based on longitudinal aircraft equations of motion is available in References 10 and 11.

$$F \equiv \frac{\dot{W}_x}{g} - \frac{W_h}{V} \tag{3}$$

F-factor is a natural measurement for reactive sensors based on inertial instruments, since all of the parameters can be directly measured. Forward-looking infrared systems under development also derive F, computing it from the temperature gradient sensed along the aircraft flight path.[10] TDWR, however, must infer F since the vertical winds cannot be directly measured.

Hazard criterion 4 is computed by taking the largest F-factor value which is exceeded for a minimum of the specified distance along the flight path. This is analogous to having an airborne sensor system, focussed ahead of the aircraft, which generates an alert if the measured F exceeds a given threshold value for the specified distance. Since F is an instantaneous measurement, taking the maximum F over a distance is similar to an integration and indicates the magnitude of total performance loss over that distance. Similarly, hazard criterion 5 takes the largest F-factor computed from a running average over the specified distance ahead of the aircraft.

3.3 Aircraft/Microburst Interaction Simulation

The aircraft/microburst interaction simulation (Figure 2) was based on longitudinal point-mass equations of motion in wind axes, which allows flight in a vertical plane only with no short-period pitching dynamics. The aircraft data used (provided by NASA Langley Research Center, and used in Reference 11) was for a Boeing 737-100 aircraft, and included non-linear curve fits for CL and CD as functions of angle-of-attack, flap position, and gear position.

$$\dot{V} = \frac{T}{m}\cos\alpha - \frac{D}{m} - g\sin\gamma - \dot{W}_x\cos\gamma - \dot{W}_h\sin\gamma \tag{4}$$

$$\dot{\gamma} = \frac{1}{V}\left[\frac{T}{m}\sin\alpha + \frac{L}{m} - g\cos\gamma + \dot{W}_x\sin\gamma - \dot{W}_h\cos\gamma\right] \tag{5}$$

$$\dot{x} = V\cos\gamma + W_x \tag{6}$$

$$\dot{h} = V\sin\gamma + W_h \tag{7}$$

$$\alpha = \theta - \gamma \tag{8}$$

$$C_L = f(\alpha, \text{flaps, gear position}) \tag{9}$$

$$C_D = f(\alpha, \text{flaps, gear position}) \tag{10}$$

A multivariable controller for airspeed and flight path angle was then designed for this model. This inner-loop controller took the form of a fully coupled proportional-plus-integral cascade compensator. In addition, an outer loop was closed around flight path angle to track the glideslope. Two versions of the outer loop (glideslope tracking) compensator were used. The baseline compensator was a simple gain, and produced a fairly slow response. The second version was a dynamic compensator which produced a faster response, and tended to weight glideslope deviations more heavily than airspeed deviations. The "slow" controller had a natural frequency of 0.1 rad/sec, while the "fast" controller had a natural frequency of 0.4 rad/sec.

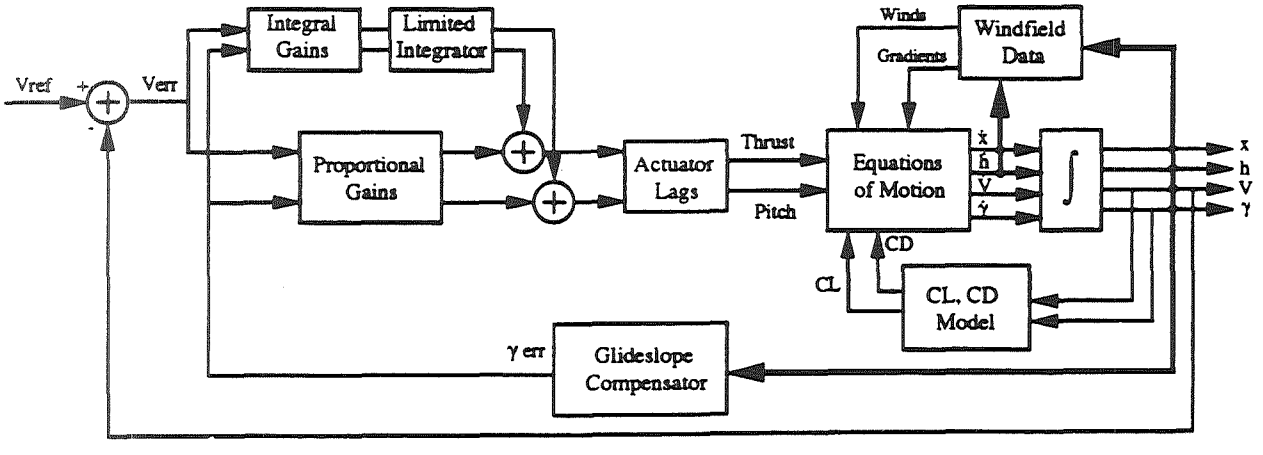


Figure 2. Simulation Block Diagram

The windfield data used was generated by the Terminal Area Simulation System (TASS)[12], and is a simulation of a complex multiple microburst event which occurred at Denver-Stapelton airport on July 11, 1988. This event caused one near accident and a total of five aircraft to make missed approaches, and has been extensively documented in Reference 3. Data from five times during this event was available, and 12 paths through various parts of the windfield at various times were selected for analysis. These paths all penetrate the approximate center of at least one microburst, and vary in total divergence (at 300ft AGL) from 42 to 70 knots. Shear distances (measured from peak headwind location to peak tailwind location) vary from 6592 feet up to 20000 feet. Figure 3 illustrates one of the horizontal velocity fields from the TASS model, and some of the paths used for the approaches.

A typical microburst windfield (as encountered in one of the simulation runs) is shown in Figure 4. The aircraft first encounters a performance-increasing headwind. This is followed by a downdraft and a rapid transition from headwind to tailwind, both of which tend to drive the aircraft below glideslope and reduce airspeed. These effects are aggravated if the pilot or autopilot is unaware of the microburst, and reduces thrust during the headwind portion of the event to maintain glideslope tracking.

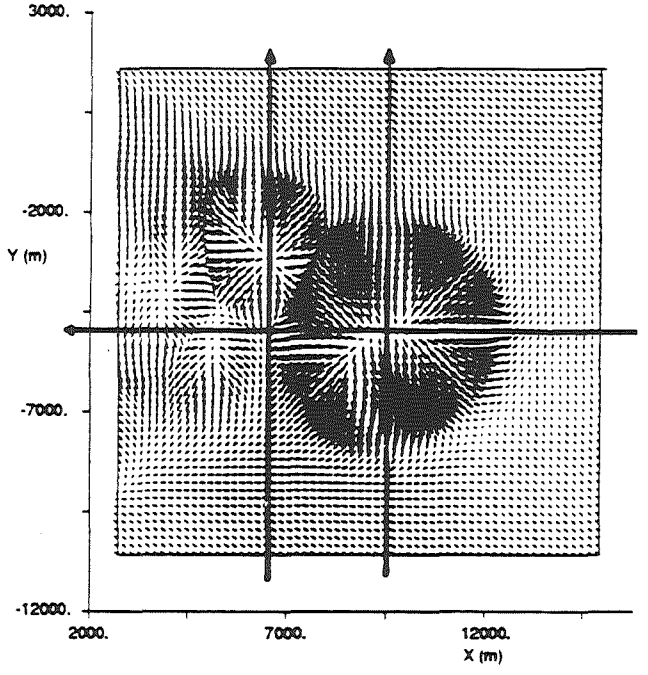


Figure 3. TASS windfield for 7/11/88 at 2212.75 UTC. This is a vector plot of the horizontal plane winds at 271 feet AGL. The lines indicate some of the paths used in the simulation runs (windfields 2, 3, and 5).

For each windfield, an approach was made such that the aircraft penetrated the core of the microburst at approximately 1.5 nm (at an altitude of 400 to 500 feet) from the desired touchdown point. The aircraft was trimmed for approach at 140 knots with landing gear extended and flaps set at 25°. The available thrust was limited to approach maximum, which results in a maximum thrust-to-weight ratio of approximately 0.17. As stated above, it is important to note that the assumption was made that the pilot or autopilot was unaware of the presence of the microburst, and hence attempted only to maintain the glidepath with a "normal" control strategy.

The actual simulation was run on a Sun 3/80 workstation using MATRIXx SYSTEM_BUILD software by Integrated Systems, Inc. The altitude and airspeed histories for a sample run are presented in Figure 5.

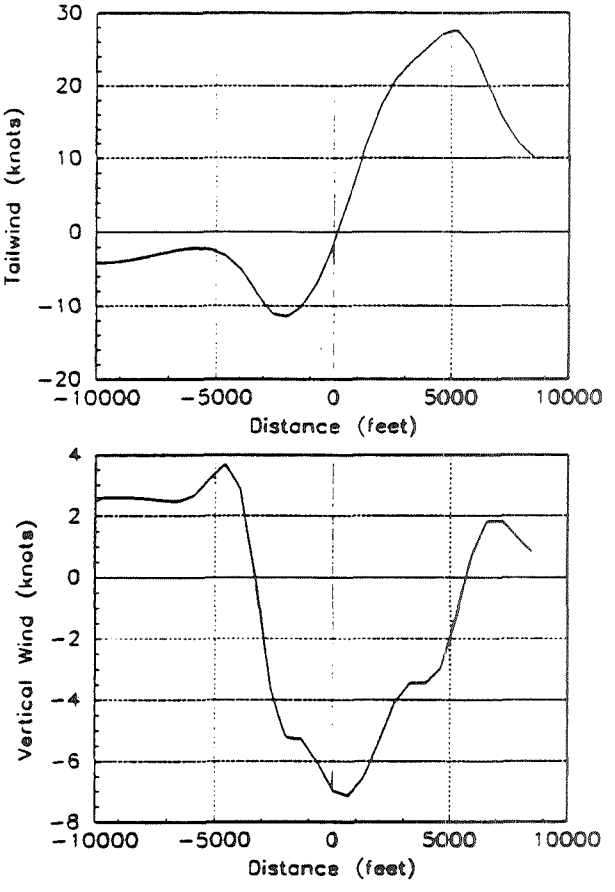


Figure 4. A typical microburst wind profile. These winds were encountered by the aircraft during one of the simulated approaches.

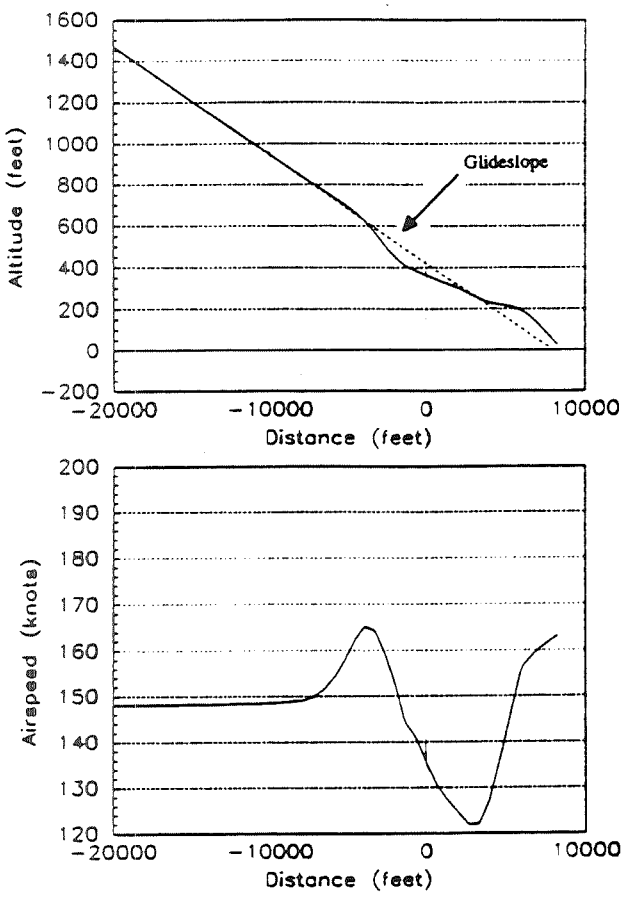


Figure 5. Sample altitude and airspeed histories for a simulation run

3.4 Results and Discussion

The primary result of each simulation run was the final value of J, the microburst impact parameter. The hazard criteria were then applied for each windfield. Divergence and mean shear were computed from a 300 foot AGL path through the windfield along the approach path to simulate a TDWR-like measurement. F-factor parameters were computed from the actual F-factor history experienced by the aircraft in the simulation. A linear correlation analysis was then performed for each of the hazard criteria vs. J Table 1 presents the results for the set of 12 runs using the baseline "slow" glideslope control strategy.

The criteria which showed the best correlation ($\rho > 0.81$) for this set of runs were mean shear, largest F exceeded over 1500 or 3000 feet, and largest mean F over 3000 to 6000 feet. The peak F-factor along the flight path does not correlate as well. Measuring the largest F over 6000 feet correlates very poorly. This is

Table 1. Correlation coefficients between hazard criteria and microburst hazard parameter

Hazard Criterion	Correlation coefficient, ρ
Divergence	0.486
Mean shear	0.875
Peak F	0.711
Largest F exceeded over:	
1500 feet	0.811
3000 feet	0.858
4500 feet	0.779
6000 feet	0.445
Largest mean F over:	
1500 feet	0.743
3000 feet	0.791
4500 feet	0.801
6000 feet	0.790

logical, since the distance over which critical values of shear and downdraft were experienced was generally less than 6000 feet. Total divergence, the criterion currently used by TDWR, did not correlate very well ($\rho = 0.486$). A comparison of the divergence and mean shear results is shown in Figure 6.

The greater effectiveness of mean shear (over divergence) is not surprising, since the wind shear threat is due to loss of energy from both rapid change in horizontal wind velocity and from vertical wind velocity (downdraft). The mean shear criterion is an improvement over divergence in two ways: 1) it provides a better measure of the effect of changing horizontal winds, since due to the dynamics of the aircraft, the rate of headwind change is as critical as the amount of headwind change, and 2) due to mass continuity, the magnitude of the downdraft is related to the rate of headwind change, and therefore the mean shear value also accounts in some fashion for the downdraft component.

The varying results for the criteria based on F demonstrate a problem with using a point measurement to predict hazard. Clearly, F must be significant over some distance/time to be truly hazardous. Microbursts, however, have some wind profile shape similarity. If all microburst windfields were geometrically similar, the peak F would basically define the hazard posed by a given microburst. Since this is not the case, measuring/averaging F over a distance seems to be an effective strategy, yielding some information about integrated energy loss as well as point hazard. The problem lies in selection of the appropriate threshold length (for criterion 4) or averaging window length (for criterion 5). Results for several of the F-based criteria are shown in Figure 7.

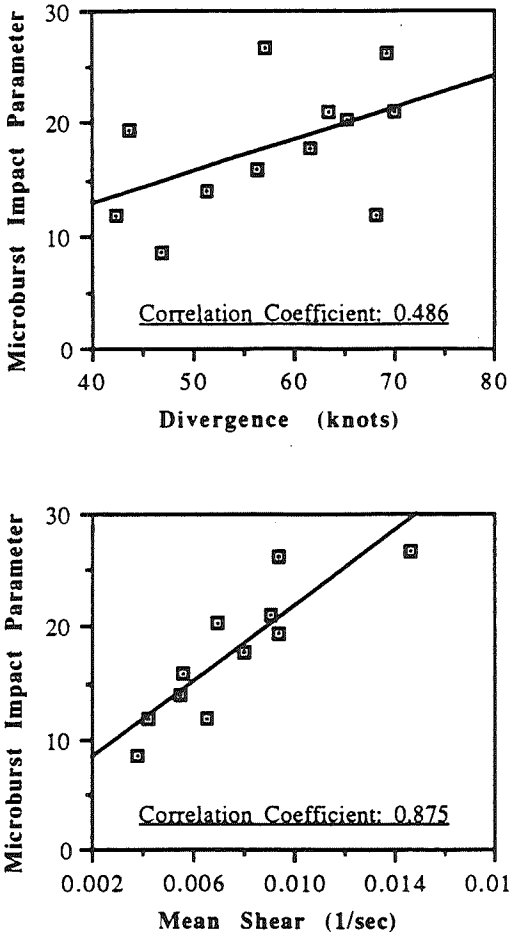


Figure 6. Microburst impact parameter vs divergence and mean shear hazard criteria

For the "largest F exceeded" criterion, 3000 feet proved to be the best distance, although 1500 and 4500 foot distances also gave good correlation. Using the averaging technique, it was found that all the distances tested worked fairly well, with the best results for 4500 feet. Neither of these criteria were very sensitive to the distance chosen. The "largest F exceeded" criterion produced better results for the shorter distances, and is simpler to evaluate, which indicates that it may provide a better, earlier warning when used with an airborne look-ahead sensor. It is important to note, however, that the results for the different distances are most likely somewhat dependent on the model windfield used. Also, the advantage of criterion 4 over criterion 5 is not very significant in view of the limited number of runs performed.

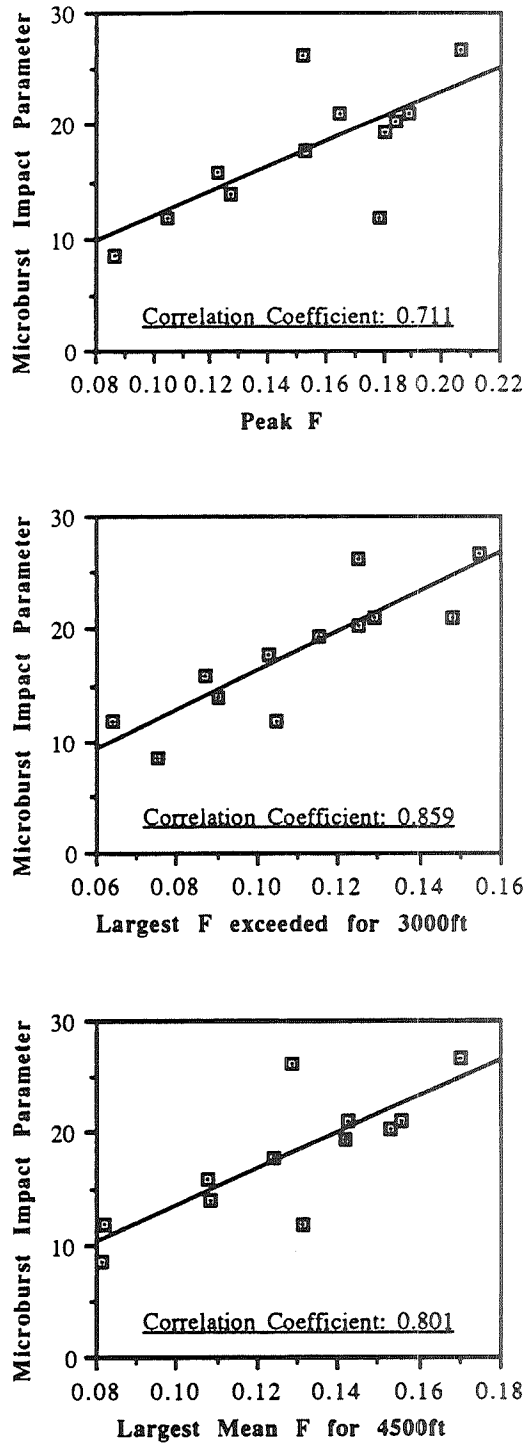


Figure 7. Microburst impact parameter vs. several F-based hazard criteria

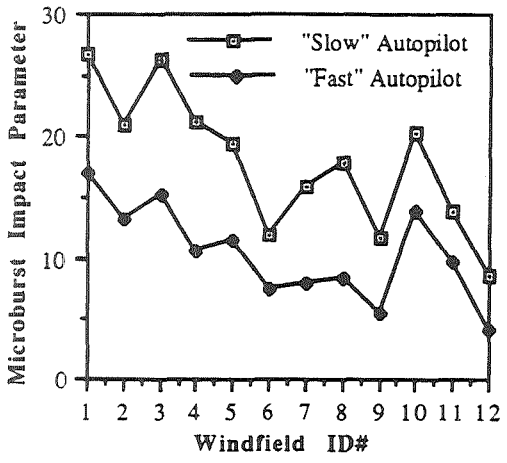


Figure 8. Microburst impact parameter for each windfield, for two different glideslope controllers

In order to examine the effects of control strategy on the above results, an additional set of runs was performed, using the "fast" glideslope controller with the same windfields. This faster controller resulted in lower values of J for all of the approaches (Figure 8). However, the correlation coefficients for the hazard criteria were not significantly different than those from the first set of runs (Table 2). This indicates that, although the final magnitude of J (and hence the hazard) depends on the control strategy employed, hazard criteria which correlate well with the microburst impact parameter for one control strategy will still correlate well for another strategy. A similar effect would be seen when comparing results for aircraft with differing performance capabilities.

Table 2. Correlation coefficients for runs with "fast" glideslope controller

Hazard Criterion	Correlation coefficient, ρ
Divergence	0.491
Mean shear	0.845
Peak F	0.814
Largest Mean F (4500ft)	0.888

3.5 Analysis Limitations

Before drawing conclusions from the above analysis, some limitations should be discussed. The most important is that all the microburst windfields analyzed came from a single day (in fact, all within 10 minutes) due to the limited amount of high-resolution windfield model data available. Although a fairly well distributed set of microburst intensities and sizes were available in this data set, the results may be biased by some feature of the TASS model or by characteristics particular to the meteorological conditions on the day simulated. This analysis will be extended to include additional data sets as they become available.

A further limitation is the use of a single aircraft model of a fairly small transport aircraft. Use of a different transport aircraft model or a different pilot or autopilot model would clearly result in different values of J. The higher-bandwidth control strategy, for example, lowered the final J value for all runs. This did not, however, significantly alter the correlation results, since the difference in approach tracking due to the different control strategy was roughly the same for all microburst encounters. Using a different transport aircraft model would have a similar effect. It is also interesting to note that the only control tradeoff available from an energy standpoint is whether airspeed control or altitude control is weighted more heavily. The microburst impact parameter includes both of these deviations. The extension of this analysis to include general aviation aircraft is also under consideration; GA aircraft have much slower approach speeds and considerably different dynamic characteristics. Also, as discussed previously, these results are specific to the nominal case chosen (approach), and need to be extended to cover takeoff situations as well.

In addition, the evaluation of the hazard criteria did not include measurement limitations. For example, the measurement of divergence and mean shear at 300 feet AGL was done to "simulate" a ground-based radar like TDWR, but does not account for measurement details such as averaging across the radar beamwidth, for example. Similarly, F was taken from the aircraft history after the simulation run, and not from a simulation of a forward-looking infrared or other airborne sensing system.

Finally, it should be noted that in some past wind shear related accidents (the crash of Delta 191 at DFW in 1985, for example) there were short-scale pitching and rolling motions of large enough amplitudes to considerably affect the controllability and performance of the aircraft. The disturbances which caused these oscillations are not generally included in windshear models (due to their small length scales) and are not visible in dual or triple doppler radar wind measurements (due to finite resolution). These disturbances are also not included in J or in any of the hazard criteria.

3.6 Implementation Issues

The second requirement for a good hazard criterion, as stated in Section 2.1, is that it must be measureable by the available wind shear sensors. Based on the

correlation coefficients, mean shear and largest F-factor exceeded over 3000 feet were the best criteria. For a ground-based radar like TDWR, which cannot directly measure vertical velocity and hence cannot directly measure F, mean shear is a natural measurement. The TDWR system currently generates accurate values of both divergence and shear distance. Airborne infrared systems now under development measure a temperature gradient at some point ahead of the aircraft, which is converted to a point measurement of F. Using the F-based criterion would be natural in this case, since the F-over-a-distance alert criterion reduces to the IR system sensing an F above a designated threshold for a designated period of time. Current airborne reactive sensors measure F through a low-pass filter, which is similar to averaging F over a distance. Clearly the measurement technique used has a strong bearing on which hazard criterion is appropriate.

Once the criterion has been established, the question that remains to be answered is: what value of the hazard criterion should trigger an alert? This question is difficult, since it requires evaluation of how much wind shear hazard is "acceptable" and how much is a threat. The analysis technique used above cannot answer this question. The hazard criteria evaluated above will most likely correlate similarly with the microburst impact parameter if other aircraft configurations, types, or autopilots (such as in Figure 8) are used, but there will be a scaling difference. In other words, applying a good hazard criterion to two different microbursts will indicate which one is a greater threat, but will not indicate that both might be hazardous to a heavily loaded aircraft with a low excess thrust-to-weight ratio, while neither may be hazardous to a nearly empty aircraft. Alternatively, two aircraft with different autopilots (or with two pilots of differing skill) might exhibit a similar difference in hazard threshold. A margin of safety must be included in definition of the alerting threshold, and all classes of aircraft to which the alert will be issued need to be considered.

4. Planned Piloted Simulator Study

4.1 Goals

The goal of this study is to evaluate candidate graphical cockpit displays for wind shear (microburst) alerts on aircraft equipped with an Electronic Flight Instrumentation System (EFIS). For this evaluation, graphical microburst alerts will be displayed on the Electronic Horizontal Situation Indicator (EHSI), a moving-map type navigational display.

The experiment is designed to evaluate the following:

1) General pilot reactions to the use of the EHSI for microburst alert presentation and to some baseline graphical display formats
2) The effect of displaying either only "threatening" microbursts or both threatening and less intense microbursts on pilot decision making and missed approach planning
3) The effect of reflectivity on pilot decision-making (i.e. will the subjects react differently to "dry" microburst events than to "wet" events)
4) Pilot reaction to presenting "fused" alerts (where the sensors used are not evident) vs. presenting discrete alerts (where a different representation is used for airborne sensors than for ground-based sensors)
5) The proximity to a hazard at which the subjects feel a missed approach must be initiated.

The results of this evaluation will be used to determine critical design items for implementation of integrated graphical wind shear alerts.

4.2 Experimental Methodology

The experiment will be performed with the MIT Advanced Cockpit Simulator, a part-task simulation of an advanced transport aircraft equipped with both an EFIS and a Flight Management Computer (FMC). The simulator, an upgraded version of that used in the modes of presentation comparison discussed in Section 2.2, includes all primary electronic flight displays, display controls, autoflight systems, and a partial simulation of the FMC.

A sample candidate EHSI display format for microburst alerts is shown in Figure 9. Remotely detected microbursts are presented as solid red circles or "bandaid" shapes in the appropriate location on the map display. The method by which the microbursts are being detected is transparent to the pilot on this display; an alternative display in which airborne remote sensor information is distinguished from ground-based information will also be tested. Reactive alerts, triggered by penetration of a microburst event, will be displayed as a flashing red circle directly around the "ownship" symbol. In one methodology, only microbursts which have been judged "threatening" are displayed. Another display option is to also show "secondary" events, which are measured areas of wind shear (weak microbursts) which do not exceed the hazard threshold. These could be displayed as open red circles or bandaid shapes to distinguish them from hazardous microbursts.

The evaluation will be performed in two different settings: 1) static display presentation, and 2) real-time simulated approaches. The static display presentation will be used to obtain pilot input on the general display format and display options in a non-flying situation. The simulated approaches will test the display in flight situations. A set of scenarios have been designed to obtain both quantitative data and pilot opinions on each of the items on the above list. For example, the same threatening microburst situation will be flown using several different display options, and differences in the pilots' handling of the situation will be observed.

Active line pilots who are currently flying EFIS/FMC-equipped transport aircraft will be recruited for this experiment. The flying portion of the experiment will be performed in conjunction with an electronic approach plate experiment, which will attenuate the anticipation caused by repeated microburst alerts. Scenarios with non-threatening microbursts present and with no microbursts present will also be included, to prevent the subjects from developing a systematic way of handling microburst threats.

5. Conclusions

Issues related to the development of an alerting system for microbursts in the terminal area have been studied. Two specific issues have been addressed: 1) identification of a good criterion for microburst hazard assessment, and 2) evaluation of graphical cockpit displays for presentation of microburst alerts.

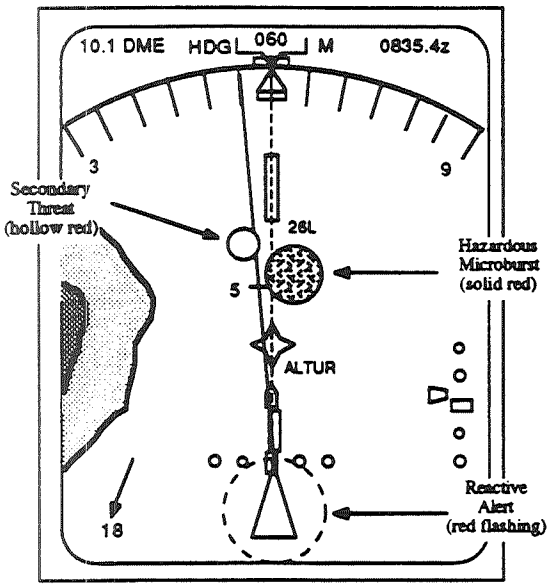


Figure 9. Candidate EHSI display format

A methodology for evaluating microburst hazard criteria has been developed, based on batch flight simulations of aircraft on approach through microburst windfields. A measure of the impact of a microburst on an aircraft's approach trajectory has been defined, which includes both airspeed losses and glideslope deviations. Hazard criteria are evaluated by calculating the linear correlation coefficient between values of the hazard criteria and the values of the microburst impact parameter for several approaches through different microburst windfields. A preliminary evaluation of a selected set of hazard criteria has been performed. The results indicate that the total headwind change along the aircraft flight path (referred to as "total divergence") does not correlate well with the approach degradation caused by microbursts. However, the "mean shear," computed by dividing the total divergence by the distance over which it occurs, correlates very well with approach degradation. Also, application of the "F-factor" hazard criterion (a measure of the aircraft energy loss caused by the windfield) over certain distances proved to be an effective predictor of approach degradation. Finally, it was found that the correlation results obtained from runs using two different autopilot models were not significantly different.

An experiment has been designed to evaluate options for graphical cockpit display of microburst alerts. The experiment, to be performed on a part-task simulator with electronic flight instrumentation, will obtain pilot feedback and performance data on candidate graphical display options.

Acknowledgements

This work was supported by the MIT Lincoln Laboratory and the National Aeronautics and Space Administration (NASA) under contract BARR-10-119, and the Federal Aviation Administration and NASA under grant NGL-22-009-640. The authors would like to thank Steve Campbell at MIT Lincoln Laboratory for information on the TDWR system and general advice and support, David Hinton at NASA Langley Research Center for the aircraft simulation model data, input on the use of F-factor, and suggestions for design of the candidate graphical display format, and Fred Proctor of MESO, Inc. for the data from the TASS model.

References

1. National Research Council, *Low Altitude Wind Shear and Its Hazard to Aviation*, National Academy Press, 1983.

2. Wolfson, M.M., "Characteristics of Microbursts in the Continental United States," *The Lincoln Laboratory Journal*, Vol. 1, No. 1, Spring 1988.

3. Schlickenmaier, Herbert W., "Windshear Case Study: Denver, Colorado, July 11, 1988," Federal Aviation Administration, DOT/FAA/DS-89/19, 1989.

4. Hansman, R. J., and Wanke, C., "Cockpit Display of Hazardous Windshear Information," AIAA Paper 89-0808, January 1989.

5. Wanke, C., Chandra, D., Hansman, R. J., and Bussolari, S. R., "A Comparison of Voice and Datalink for ATC Amendments and Hazardous Wind Shear Alerts," 4th International Symposium on Aviation and Space Safety, Toulouse, France, 20-22 November 1990.

6. Wanke, C., and Hansman, R. J., "Hazard Evaluation and Operational Cockpit Display of Hazardous Windshear Information," AIAA Paper 90-0566, January 1990.

7. McCarthy, J., Blick, E. F., and Bensch, R. R., *Jet Transport Performance in Thunderstorm Wind Shear Conditions*, NASA CR-3207, 1979.

8. Turkel, B. S., Kessel, P. A., and Frost, W., *Feasibility Study of a Procedure to Detect and Warn of Low-Level Wind Shear*, NASA CR-3480, 1981.

9. Merritt, M. W., Klingle-Wilson, D., and Campbell, S. D., "Wind Shear Detection with Pencil-Beam Radars," *The Lincoln Laboratory Journal*, Vol. 2, No. 3, 1989.

10. Bowles, R. L., "Reducing Windshear Risk Through Airborne Systems Technology," 17th Congress of the International Council of the Aeronautical Sciences, Stockholm, Sweden, September 9-14, 1990.

11. Hinton, David A., *Relative Merits of Reactive and Forward-Look Detection for Wind-Shear Encounters During Landing Approach for Various Microburst Escape Strategies*, NASA TM-4158, DOT/FAA/DS-89/35, February, 1990.

12. Proctor, F. H., *The Terminal Area Simulation System - Volume I: Thearetical Formulation; Volume II: Verification Cases*, NASA CR-4046 and CR-4047, DOT/FAA/PM-86/50, 1987.

Cockpit Display of Hazardous Weather Information

R. John Hansman, Jr.* and Craig Wanke†
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts U.S.A.

Abstract

Information transfer and display issues associated with the dissemination of hazardous weather warnings are studied in the context of windshear alerts. Operational and developmental windshear detection systems are briefly reviewed. The July 11, 1988 microburst events observed as part of the Denver TDWR operational evaluation are analyzed in terms of information transfer and the effectiveness of the microburst alerts. Information transfer, message content and display issues associated with microburst alerts generated from ground based sources (Doppler Radars, LLWAS and PIREPS) ars evaluated by means of pilot opinion surveys and part task simulator studies.

1. Introduction

Technological advances in ground-to-cockpit datalink capability, information display, and hazardous weather detection create the possibility for new and improved methods of informing flight crews about weather hazards. However, the availability of increased information and multiple modes of communication also lead to problems of system integration. Issues including the selection, transfer, and presentation of information must be addressed in the development of advanced systems for the display of hazardous weather information. In addition, design procedures, centered around the needs of the flight crew and the capabilities of the available equipment, should be applied.

The display and information transfer issues related to advanced windshear alerting systems in the terminal area have been chosen as an initial point of focus. This problem was chosen both to investigate general issues related to the dissemination of hazardous weather information and to focus on specific issues of a critical near term need. Windshear in the terminal area is one of the most dangerous weather-related problems faced by aviation today.[1]. The real-time detection of windshear hazards is a very active field of research,[2,3,4] and thus provides a useful testing ground for issues related to advanced data uplink and display of hazardous weather information.

2. Background

2.1 Terminal Area Windshear

Low-altitude windshear is the leading weather-related cause of fatal aviation accidents in the U.S. Since 1964, there have been 26 accidents attributed to windshear resulting in over 500 fatalities [1,5]. Low-altitude windshear can take several forms. Macroscopic forms, such as gustfronts caused by colliding warm and cold air masses, can generally be predicted and avoided. However, the small intense downdrafts known as microbursts are far more dangerous and difficult to detect.

Microbursts begin with a cool downdraft formed at the base of a cumulus or cumulonumbus cloud. If the downdraft is strong enough to impact the surface, it spreads out radially and creates an small area (1 to 4 km in diameter) of intense windshear. Such conditions typically last for short periods (10-30 min), but can be very dangerous to aircraft at low altitudes, particularly on takeoff or final approach Initially, the aircraft experiences a strong headwind, which causes a momentary increase in lift. Next, the aircraft enters an area of downdraft, and then a sharp tailwind. This combination results in loss of effective airspeed and corresponding loss of lift. (Fig. 1). It may also serve to destablilize the flight trajectory. The resulting performance loss can in some cases be sufficient to result in ground impact. In addition, microbursts can be accompanied by strong edge vortices, which can further destabilize the aircraft. Most fatal windshear accidents have been attributed to microbursts.[5]

An additional factor which makes microbursts particularly dangerous is that they are generally not obvious either visually or to standard airborne weather radar. Microbursts have been observed to occur both during periods of severe rain or during periods of little or no low-altitude precipitation. For meteorological and instrumentation purposes, it is convenient to distinguish between 'wet' and 'dry' microbursts. Dry microbursts, more common in the western U.S., can sometimes be detected by the presence of curling clouds of dust on the ground or vertical cloud shafts known as 'virga'. Wet microbursts cannot generally be distiguished from benign rain cells with radar reflectivity information.

Microbursts have been observed with intensities greater than most aircraft could be reasonably expected to survive. Avoidance is the best way to handle a windshear hazard. This indicates a need for reliable remote detection, allowing the flight crew adequate advance warning to plan and execute a maneuver to avoid microburst penetration.

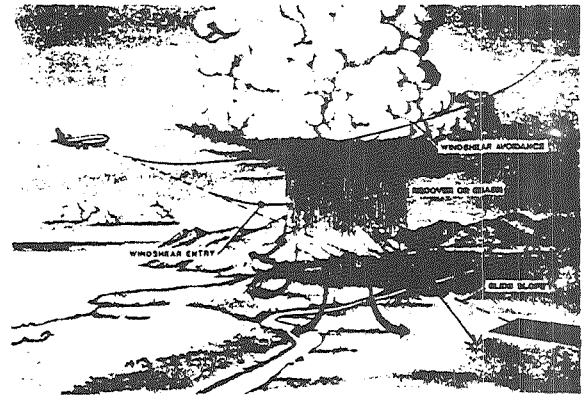


Fig. 1: Microburst windshear encounter on approach (from Ref. 2).

*Associate Professor, Associate Fellow AIAA
†Research Assistant, StudentMember AIAA

2.2 Microburst Detection

2.2.1 Current Procedures

Current procedures for microburst detection and warning center around the Low-Level Windshear Alert System (LLWAS), Pilot reports, and improved pilot education through efforts such as the FAA's *Windshear Training Aid*.[6] LLWAS is a system of anemometers currently in service at most major U.S. airports designed to measure shifts in wind speed and direction within the airport perimeter. Although capable of detecting macroscopic phenomena such as gustfronts, the anemometer spacing is larger than the characteritic surface dimension of many microbursts, and thus LLWAS remains fairly ineffective for detection of microburst windshear. The *Windshear Training Aid* states that "If an LLWAS alert (triggered by wind speed and/or direction differential) occurs, it indicates the presence of something shear-like, though not necessarily indicative of magnitude or location. However, the absence of an alert does not necessarily indicate that it is safe to proceed!".[6]

Pilot reports (PIREPS) of windshear provide the most reliable data. The availability of PIREPS necessarily requires that an aircraft penetrate a microburst, which is not desirable; but the information, unlike LLWAS, provides conclusive evidence of a windshear hazard for subsequent aircraft. It is desirable to integrate PIREPS with any sensor data available in future windshear detection systems.

The *Windshear Training Aid* itself is designed to inform pilots and controllers about windshear, primarily how to recognize and avoid or recover from microburst encounters. Avoidance is practiced through the use of LLWAS information, weather reports, and visual clues. In past accidents, these clues have been largely ignored; increased windshear training emphasis is being used to increase pilot awareness of these events. However, high pilot workload in the terminal area and the relative rarity of hazardous windshear makes it difficult for crews to fully assimilate the evidence of windshear before penetration.

2.2.2 Emerging Windshear Detection Technologies

To meet the need for improved windshear warning, new systems for detection are under development. Both airborne and ground-based systems are under consideration. Airborne look-ahead systems are still primarily experimental: candidate technologies include doppler radar, doppler lidar, and infrared radiometry. [2,3,4] To be an effective, dependable windshear avoidance tool, an airborne system must be able to detect windshear ahead of the aircraft to a range of 1 - 3 km, thus typically providing 15 to 45 seconds of warning. Also, the sensor should work for either wet or dry microbursts with enough resolution to adequately measure size and intensity. None of the methods mentioned have yet fully demonstrated these capabilities in flight.

Ground-based remote sensing technology is much more developed. LLWAS and PIREPS often yield useful data, but are not always available or accurate. Ground-based doppler radars have been succesfully demonstrated for microburst detection (JAWS, Huntsville, Denver) [5] and have an advantage over airborne systems in terms of ground clutter suppression, size and power. Experiments performed at Huntsville, AL in 1986 and at Denver in 1987 and 1988 have shown impressive results (Table 1). The predominance of wet microbursts at Huntsville and dry microbursts in Denver shows the versatility of the ground-based doppler radar. The ability of such systems to integrate data aloft with wind measurements near the surface allows for earlier forecasting of microburst locations and outflow strengths.

MICROBURST DETECTION

Data	Probability of detection* $\Delta V<20$ m/s	$\Delta V \geq 20$ m/s	Total	Probability of false alarm
Huntsville 1986	88%	100%	91%	5%
Denver 1987	90%	99%	92%	5%
Combined	90%	100%	92%	5%

GUST-FRONT DETECTION

Data	Probability of detection $\Delta V \leq 15$ m/s	$\Delta V > 15$ m/s	Probability of false alarm
Denver 1987	81%	93%	5%

* ΔV = net wind change in shear region (only events with ΔV values greater than 10 m/s are scored.)

Table 1:Doppler radar windshear detection results [7]

The demonstrated capability of ground-based doppler radar for windshear detection and forecasting makes it the most viable system for near term use for microburst avoidance information. The combination of doppler radar, improved LLWAS systems and PIREPS makes an integrated ground-based system the primary focus for system integration and automated datalink issues.

2.3 Ground-to-Air Data Transfer

Digital ground-to-air data transfer is an area under active development. Several methods of digital ground-to-air data transmission are currently or nearly available. ACARS, a privately-sponsored system for the uplink and downlink of digital information related to commercial aviation, is currently in use by many major airlines. It provides a high-speed alphanumeric datalink for flight management information, helping to relieve congestion on crowded ATC voice frequencies. With the addition of satellite relays, ACARS coverage will extend to most international commercial air routes. The first satellite transmissions are expected to begin in the third quarter of 1989 in the Pacific Ocean region [8].

Another system slated for near-term deployment is the FAA's Mode-S surveillance datalink. Mode-S is an extension of the altitude encoding Mode-C transponder in the ATC Radar Beacon System allowing message delivery from ATC to individual aircraft. Each individual message can carry 48 useful bits of information, and the time for the interrogation beam to scan the entire coverage area is 4 to 12 seconds. Messages can be also be linked in groups of up to 4 frames or sent as a longer Extended Length Message with less urgency.

In the long term, the Aviation Satellite Communications System (SatCom) is being developed. The goal is a standardized worldwide system for digital voice and data communications, based on nine existing satellites in geosynchronous orbit.[9] Other systems such as digital ATIS or enroute weather channels are also envisioned for future development.

2.4 Information Transfer Issues in the 1988 Denver TDWR Evaluation

An event which illustrates many of the information transfer issues occurred during the 1988 Terminal Doppler Weather Radar (TDWR) operational demonstration at Stapelton International Airport in Denver. On July 11, a period of severe microburst activity occurred. It is instructive to evaluate the warnings and responses of the five aircraft which initiated and abandoned approaches immediately prior to the closure of the airport.

2.4.1 TDWR Set Up

In the 1988 operational evaluation, the TDWR radar was operated by the MIT Lincoln Laboratory and located at the Buckley AFB southeast of the Stapelton airport. Microburst and gustfront alerts were generated from the doppler weather radar data by an automatic algorithim and confirmed on line by an NCAR meterologist. The alerts were then sent by ground line to the Stapelton control tower and the terminal radar approach control (TRACON).

The information was displayed in two formats in the control tower. The local tower controller, who had primary responsibility for the dissemination of microburst alerts, had an alphanumeric display (Fig. 2) which could present either TDWR or LLWAS information in the same format. This was done to minimize the transition between periods of TDWR and LLWAS-only operation. The tower supervisor and the TRACON also had the geographical situation display which is shown in Fig.3. This color display presented the locations of microbursts, gustfronts and precipitation on a plan view of the runway configuration. In addition, LLWAS wind vectors are displayed. In the tower a local controller working arrivals to Runways 26L and 26R would have to cross the tower cab to have access to the geographical situation display.

Type of wind shear	Runway	Threshold winds	Wind shear Headwind change (kts)	Location
	CF	190 16 G 25		
MBA	35 LD	160 22	50-	RWY
MBA	35 RD	180 5	25-	RWY
MBA	35 LA	030 23	55-	1 MF
	35 RA	180 10	60-	3 MF
MBA	17 LA	180 5	25-	RWY
MBA	17 RA	160 22	55-	RWY
	17 LD	180 10	60-	RWY
MBA	17 RD	030 23	55-	RWY

Fig. 2: Example of controllers alphanumeric display (from Ref. 7).

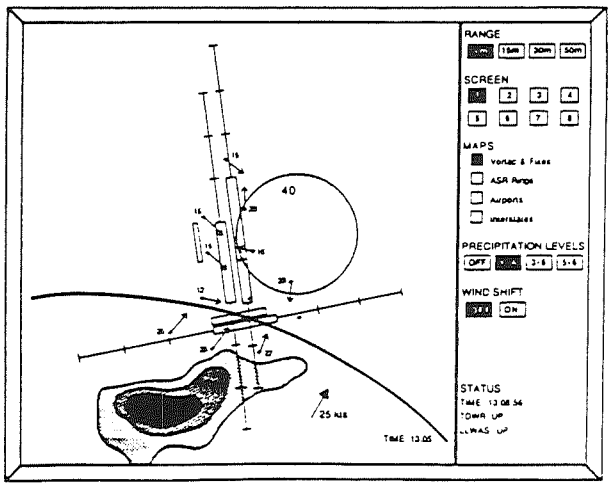


Fig. 3; Geographical situation display used in the control tower and TRACON (from Ref. 7).

2.4.2 July 11, 1988 Scenario

The period of intense microburst activity began at the Stapelton airport shortly after 2200 UTC. At this time arriving aircraft were landing on runways 26L and 26R. Departing aircraft were using runways 35L and 35R. On the arrival ATIS, aircraft were informed of a convective SIGMET for the eastern Colorado area,and that the doppler radar windshear detection demonstration was in progress. After 2203 UTC the ATIS was updated to include "low level windshear advisories in effect".

The evolution of the microburst event can be seen in the geographical situation displays presented to the tower supervisor at at 2201, 2207 and 2212 UTC (Fig.4). At 2201 UTC there

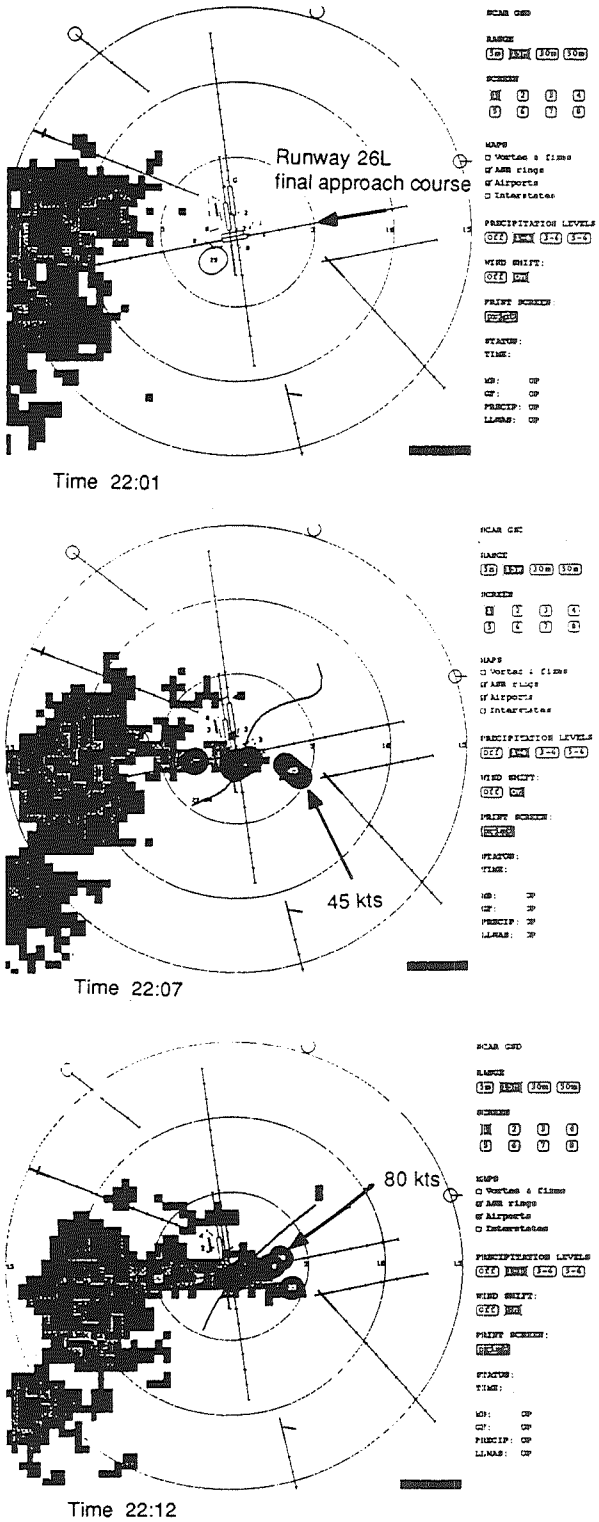


Fig. 4: Geographical situation display plots showing the evolution of the severe microburst event of July 11.

was an area of precipitation southwest of the airport and a region of 25 kt windshear within 2 miles of the airport center. By 2207 UTC. a gustfront had developed over the airport with some light precipitation. Several microbursts had developed with the gustfront including a 45 kt headwind to tailwind cell located on the approach to runways 26L and 26R. By 2212 UTC the microburst had increased in strength to 80 kts and the precipitation had increased. This microburst event continued at high intensity to 2222 UTC when it began to abate. Windshear values of 30 kts were still being measured at 2230.

The altitude versus time plots generated from Mode C transponder replies for the 5 aircraft which initiated approaches between 2207 and 2214 UTC are shown in Fig. 5. Also shown are the times at which microburst alerts were given to the aircraft and the time of reported missed approach. All aircraft which penetrated the microburst reported intense windshear. Transcripts of the verbal microburst alerts given to each aircraft by the local tower controller are presented in Table 2. It is unknown if there were any microburst alerts issued to these aircraft by the TRACON approach controller. However, the fact that 4 of the 5 aircraft elected to continue the approach indicates that this was unlikely.

2.4.3 Implications of the July 11 Experience

Several issues important to the development of microburst alerting systems are apparent from this data. The variability in aircrew interpretation of microburst warnings can be seen by comparing the response of aircraft A to that of aircraft B. The aircraft were approaching parallel runways and were issued virtually identical alerts within 30 seconds of each other. Aircraft A elected to immediately abandon the approach based on the microburst alert and visual observations of a descending rain shaft. This aircraft never penetrated the primary microburst area. Aircraft B elected to continue the approach, penetrated the microburst, and descended to within 100 ft of the runway threshold before executing a missed approach.

Another issue which arises from the data is the delay between the generation and the voice transmission of the alert to the aircraft by ATC. Fig.6 plots the delay to alert for each aircraft based on the first TDWR generated microburst alert at 22:06:17 UTC and the assumption that no alerts were given to these aircraft by the TRACON. It can be seen that the shortest delay was approximately 60 seconds and that a delay of 350 seconds was encountered for the last aircraft to report to the tower (Aircraft E). The delays in excess of 100 seconds are likely a result of the effort to make the TDWR alerts apear like LLWAS alerts. The primary windshear alert responsibility therefore rested with the tower controller who did not have contact with the aircraft until they were at the outer marker. It does appear, however, that a minimum delay of approximately 60 seconds can be expected for the dissemination of verbal alerts even if the aircraft is in contact with the controller who has alerting responsibility.

A third issue which arises is that the initial microburst alert for each aircraft was imbedded within a routine landing clearance message. The routineness of the message may have resulted in a lack of urgency associated with the alert. This possible lack of urgency coupled with the high cockpit workload which occurs at the outer marker may have contributed to the difficulty some crews had in fully assessing the magnitude of the hazard. It is also worth noting that the tower controller relied primarily on the alphanumeric display. It is interesting to consider whether his level of urgency may have increased if he had access to the geographical situation display and could have more easily visualized how the situation was developing.

The final point which comes out of the analysis is the importance of PIREPS. Both the flight crews and the tower controller were more likely to react conservatively to the microburst alert after several aircraft had gone around and

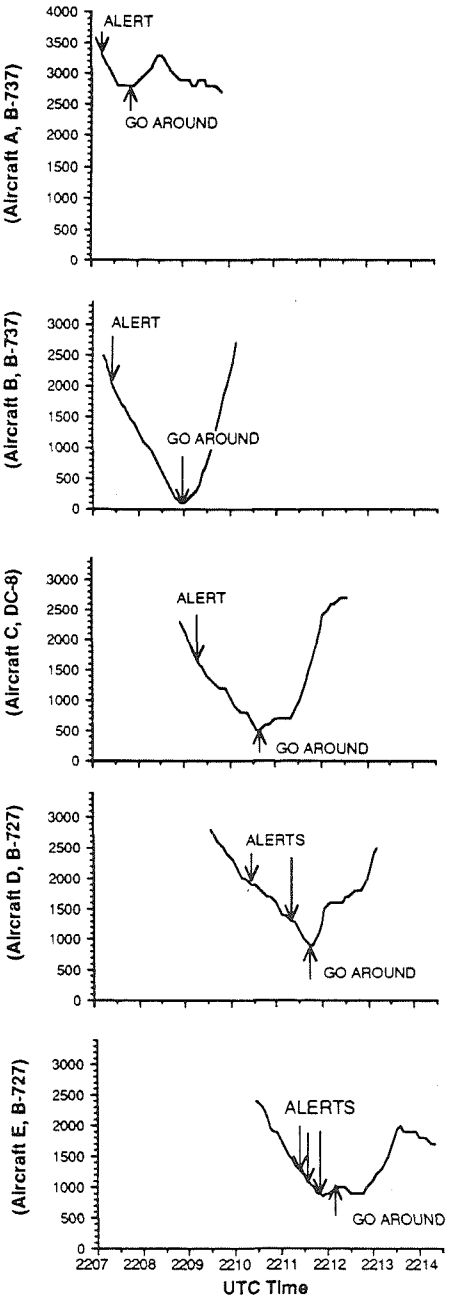


Fig. 5: Plots of aircraft height (AGL) versus time.

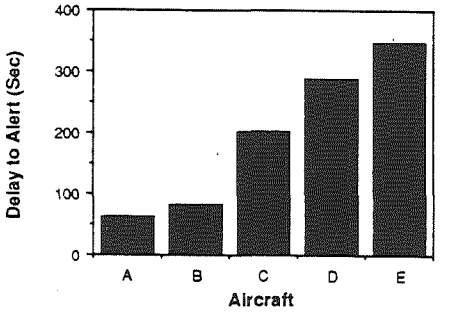


Fig. 6: Delay between first microburst alert and transmission of alert to aircraft.

	Aircraft A
22:07:15	"Aircraft A, Denver tower, runway two six right, cleared to land. Microburst alert, centerfield wind two two zero at none, a forty knot loss, one mile final as reported by machine, no pilot report."
	Aircraft B
22:07:35	"Aircraft B, Denver tower, runway two six left cleared to land. Winds two one zero at five, a forty knot loss, one mile final microburst alert, not substantiated by aircraft."
	Aircraft C
22:09:35	"Aircraft C heavy, Denver tower, microburst alert, threshold wind one four zero at five, expect a fifty knot loss, two mile final, runway two six left, cleared to land."
	Aircraft D
22:11:05	"Aircraft D, caution have turbulence from the heavy DC-8. He is going around. We have a microburst alert, threshold winds, zero nine zero at three. Expect a seventy knot loss on a three mile final."
22:11:45	"Microburst alert, runway two six. Threshold wind, one five zero at five, expect an eighty knot loss on a three mile final."
	Aircraft E
22:12:05	"Aircraft E, microburst alert, threshold wind one six zero at six, expect an eighty knot loss on a three mile final, say request."

Table 2: Transcripts of verbal microburst alerts issued to each aircraft.

reported wind shear. This, coupled with the increasing microburst intensity, explains why the later aircraft initiated their missed approaches at higher altitudes than aircraft B which had no PIREP information to confirm the microburst alert.

3. Research on Windshear Detection and Warning in the Advanced ATC Environment

3.1 Problem Statement

The integration of ground-based information sources with digital datalinks such as Mode-S shows great potential for the accurate prediction and delivery of microburst windshear alerts with minimal delay. Fig. 7 illustrates possible information flow configurations for such a system. The multiple potential data paths are dependent on the acceptable degree of automation. Clearly, the delay between detection and alert is minimized with a fully automated process whereby computer algorithms determine alerts from PIREPS, TDWR, and LLWAS data and use Mode-S to directly distribute them. However, putting the controller in the loop to some degree would help filter false alarms and more efficiently control the destination of the data.

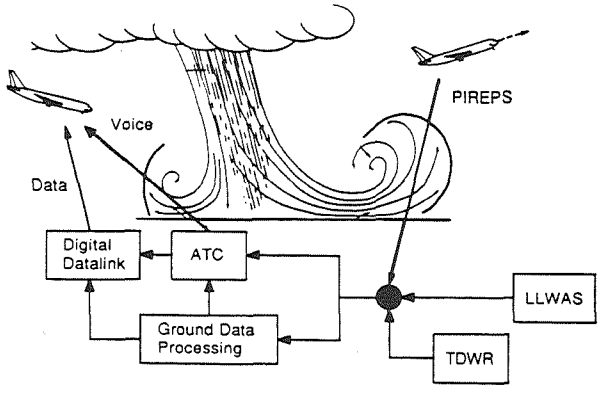


Fig 7: Possible windshear data distribution in the advanced ATC environment.

A number of other information issues also require consideration. The content, timing, transmission, and presentation of windshear information all need to be determined. Automated links such as Mode-S are subject to bit limitations and reliability considerations. This has an impact on message content and distribution. The timing and priority of alerts must be considered to get maximum efficiency and to generate the least possible confusion. The high workload of both controllers and flight crews during terminal area operations adds a further measure of difficulty. Finally, the varying levels of instrument sophistication in civil aircraft must be considered. The advanced moving map displays in modern transport category aircraft allow for development of user-oriented graphical presentations, while many general aviation aircraft have no visual display capability.

3.2 Investigations

Several investigations are being performed to address the issues discussed above. Flight crew opinion surveys are being used to obtain user input on a number of factors. Data is sought on current operational issues such as LLWAS and other available windshear information sources, as well as pilot perceptions of the microburst threat. Also, issues of data transmission and presentation are addressed. In addition, some issues are being addressed through flight simulation studies. A simple experiment based on a general aviation simulator was conducted to compare voice communication with graphical data presentation modes. Also, a part-task simulation of the Boeing 757/767 has been developed in order to do more sophisticated investigation into optimization of graphical warning formats, information content and delivery timing, and the effect on pilot workload.

3.3 Results

3.3.1 Current Windshear Procedures

For user input on current windshear alert systems and requirements for future systems, a pilot opinion survey is being conducted. A preliminary sample of 20 United Airlines line and training pilots has been completed, and a further distribution of 250 is underway. Initial results show several consistent trends. It is almost universally agreed (94%) that microbursts pose a major safety hazard to transport aircraft. Fifty-three percent of the respondents have had what they considered to be a hazardous windshear encounter; most incidents occurred at DEN, a United operations hub. When posed the question "Currently available windshear alert data is sufficient for safe operation in the terminal area," only 17% of the respondents agreed, while 56% disagreed. All but one of the pilots felt that "...a system to provide aircrews with better and more timely windshear alerts is necessary." The results clearly indicate that flight crews are not completely confident in currently available data and would be very receptive to improvements. Figure 8 shows the pilots' average ranking of possible sources of windshear information.

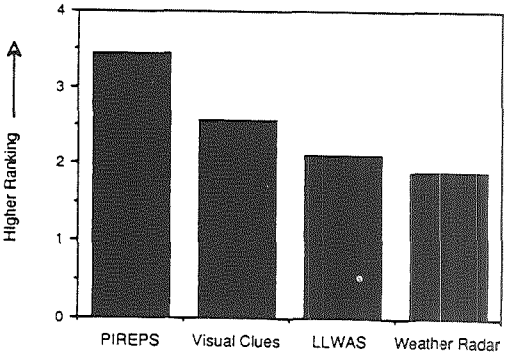


Fig. 8: Pilot ranking of windshear information sources.

Significantly, PIREPS and visual clues are both considered more useful for windshear avoidance than LLWAS alerts. Yet, neither pilot reports or visual clues are always available; this underlines the need for a detection system which can reliably provide some degree of advance warning.

3.3.2 Modes of Information Presentation

Because of the high workload in terminal area operations, it is important to consider the manner in which information is presented to the flight crew. This was illustrated by the Stapelton incident; even though data was available, it was difficult to effectively communicate it to the flight crew. There are several possible modes of information presentation in the cockpit: *voice*, *alphanumeric*, or *graphical*. Issues to be considered include crew workload, preferences, and the capabilities of the aircraft instrumentation. The widespread use of CRT displays in modern transport aircraft, for example, opens up new possibilities for totally automated graphical information displays. Moving map displays, such as the Electronic Flight Instrumentation System (EFIS) used on the Boeing 757-767 generation of aircraft, are good candidates for display of critical weather information.

Responses from the pilot survey indicate that pilots are receptive to graphic displays. (Fig. 9). The specific suggestion of integrating windshear information with an EFIS-type moving map display was strongly supported. Also of interest was the preference of ATC voice alerts over alphanumeric links or ATIS information. Comments received indicated that the low ranking of ATIS was due to the time between updates.

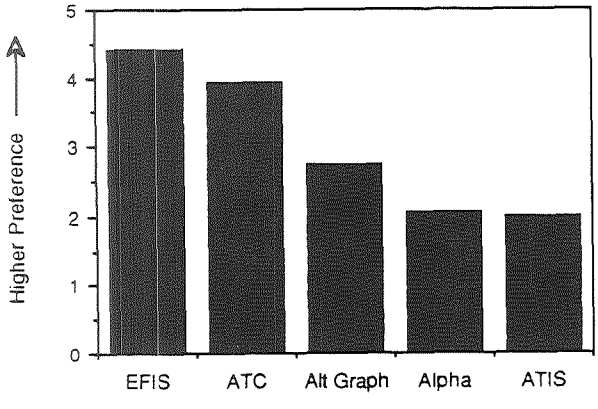


Fig. 9: Pilot rankings of possible relay/presentation modes of windshear information from the ground

A preliminary experiment has been conducted with a general aviation simulator to compare the efficiency of voice and graphical modes of presentation. Eight GA pilots with 210 to 1,700 total flight hours were tested. The scenario involved a microburst which appeared during an ILS approach, when the aircraft reached the outer marker. Avoidance of the microburst required a non-standard missed approach. The information was presented by voice, on a runway-fixed graphic display of microburst position, and on a graphic display showing both the microburst and aircraft positions. The data (Table 3) shows the effectiveness of the graphic displays. Avoidance improved significantly with the graphic displays, even though the same information was presented at the same time in each case.

Presentation Type	Avoidance Rate
Voice (JAWS format)	43%
Runway-Fixed Graphical Display: Microburst position only	62%
Runway-Fixed Graphical Display: Microburst + Aircraft position	94%

Table 3: Results of experiment with general aviation simulator and computer graphic display.

These results considered, the incident at Stapelton Airport serves as an illustration of the problems of voice communication. Crew and ATC workload in terminal phases of flight is high, leading to possible confusion and error. The simulation indicates that even under fairly light workload, the difficulty involved in fully interpreting the microburst threat from a voice warning can mean the difference between avoidance and penetration. Further evaluation of communication modes, including a variety of alphanumeric and graphic formats, will be performed with the part-task 757/767 simulator.

3.3.3 Message content and timing

The issue of what data is necessary and when it should be presented is important for either voice or digital transmission. In either case, a limited amount of information can be contained, and the timing must be determined to give the crew maximum awareness while minimizing the increase in workload. An initial viewpoint can be obtained from the pilot surveys. The responses indicate that location and intensity of microbursts are clearly the most important information items. Size, microburst movement, and intensity trends are of secondary importance, and shape data is generally felt to be inconsequential.

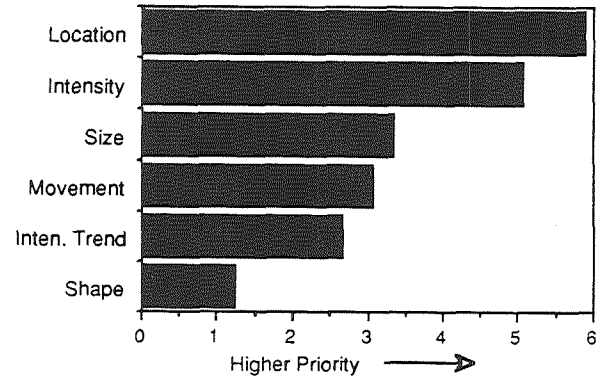


Fig. 10: Pilot ranking of microburst information by importance

The issues of what constitutes a hazardous microburst, who should be informed, and when are more difficult to resolve. The surveys are less clear in this case; the most common response was that aircraft should be alerted as soon as microbursts are detected anywhere around the airport vicinity. A few pilots defined a particular phase of flight, i.e. at the outer marker, when cleared for approach, or immediately upon entering the terminal area, as the best point for delivery of microburst alerts.

In response to a question about threshold shear levels, there was general agreement that a windshear advisory should be issued for approximately 10 knots of head-to-tail shear and a warning for 15 knots of shear. Also, it was almost unanimously expressed that decisions about the threat posed by windshear in a particular situation should be made entirely by the pilot, and the controller's role should be to maintain safe separation during

avoidance maneuvers. However, it remains to be determined what locations and intensities of microbursts actually constitute a threat in the view of the pilot. It is impractical to plan on distribution of all available windshear information in raw form to all aircraft in a congested terminal area. Some 'threshold hazard level' needs to be defined, based not only on the windshear intensity of the microburst, but including other factors such as the microburst and aircraft locations, aircraft altitude, and desired flight path.

3.4 Current Research

Research to resolve these issues is being conducted with the part-task 757/767 simulation shown in Fig. 11. The simulation uses an IRIS 2400T graphics computer , an autopilot control panel, and an EFIS control panel to duplicate the electronic instrumentation and flight dynamics of the aircraft. Data from TDWR experiments is used to generate simulated airborne weather radar returns and the windfield over the airport.

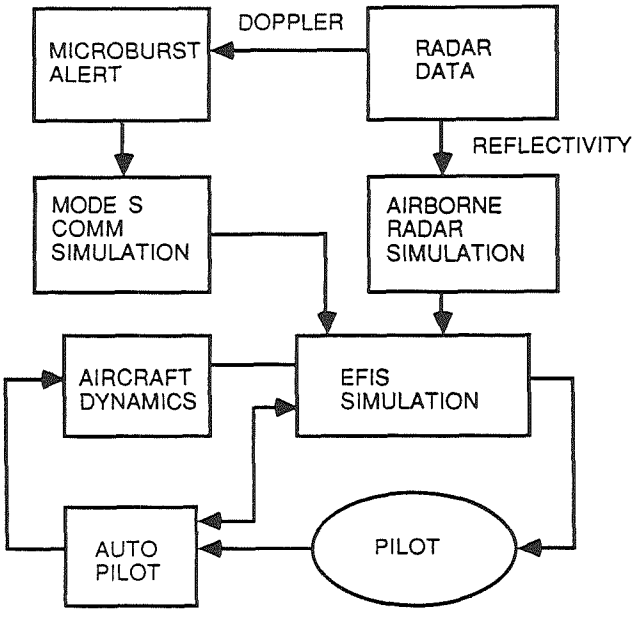


Figure 11: Part-Task 757/767 EFIS Simulation

The initial simulations are based on data provided by NCAR and the Lincoln Laboratory TDWR evaluations. The Stapelton incident is very well documented, and serves as a model for scenario construction. With a suitable sidetask, workload levels will be properly adjusted to get a reasonable range of pilot responses. Once the simulation is validated and a range of scenarios developed, issues of information format can be explored. Simulations of Mode-S transmissions with varying alphanumeric and graphic alert formats can be added, as well as voice communications, and the differences in pilot decision making and reaction time can be measured. If the results are commensurate with the results of the earlier general aviation simulations, more specific tests can be performed. These will center around more specific information issues, such as warning content, timing, and display formats.

4. Conclusion

Based on the above, the following points can be made:

- Technological advances in weather sensors and information transfer will allow development of sophisticated hazardous weather detection and alert systems. Design guidelines, centered around the end user and the available equipment, need to be applied to these systems.

- Microburst windshear is an weather hazard of particular concern and hence provides a good test case for development of user-oriented weather alert displays. Pilot surveys indicate that currently available detection and alert systems are not adequate, and a system for advance detection and alert is needed.

- The events which occurred during the TDWR operational evaluation on July 11, 1988 were analyzed in the context of information transfer issues. The observations included the following: Variability of pilot response to similar microburst alerts. The verbal relay of microburst alerts was found to induce delays. The inclusion of microburst alerts with other routine messages was thought to reduce the sense of urgency of the alerts. Finally, PIREPS have been found to be extremely important in validating the TDWR alerts to the user.

- A review of the current state of microburst detection technology and the analysis of the TDWR operational evaluation leads to the conclusion that the integration of ground-based doppler radar, LLWAS and PIREPS, and of a digital datalink such as Mode-S is the most viable near-term system for reliable advance warning of windshear.

- A simple flight simulator study has indicated that display of windshear information with a graphical display of aircraft and microburst position can result in significantly greater microburst awareness and greatly improve the probability of avoidance when compared with standard voice transmission.

- An opinion survey of air carrier pilots was conducted. Pilots feel that PIREPS and visual clues are the best currently available methods for microburst detection, while LLWAS and airborne weather radar are less effective. Also, pilots were receptive to the idea of displaying windshear information on an EFIS display, preferring the EFIS to ATC voice communications. Alphanumeric information and ATIS were rated poorly for transmission of windshear alerts.

- When asked about microburst alert information content, pilots specified that microburst location and intensity were the most important items, followed by size, movement, and intensity trend information.

- Research is currently in progress to further explore the issues involved. A part-task Boeing 757/767 simulation has been developed to address issues of warning content, timing, and a selection of alphanumeric and graphical display formats.

Acknowledgments

This work was supported by the MIT Lincoln Laboratory under contract BARR-10-119, and the Federal Aviation Administration and the National Aeronautics and Space Administration under grants NGL-22-009-640 and NAG-1-690. Data from the 1988 TDWR operational evaluation was provided by Dr. Wayne Sand and Cleon Biter of the NCAR Research Applications Program. The authors would like to thank the respondents, Rick Brown, and United Airlines for assistance with the pilot opinion surveys.

References

1. National Research Council, *Low Altitude Wind Shear and Its Hazard to Aviation*, National Academy Press, 1983.

2. Targ, R., and Bowles, R.L., "Investigation of Airborne Lidar for Avoidance of Windshear Hazards," AIAA Paper 88-4658, Sept. 1988.

3. Bracalente, E.M., Britt, C.L., and Jones, W.R., "Airborne Doppler Radar Detection of Low Altitude WIndshear," AIAA Paper 88-4657, Sept. 1988.

4. Adamson, H.P., "Airborne Passive Infrared System for the Advance Warning of Low-Level Windshear and Clear Air Turbulence: 1988 In-Service and Theoretical Work," AIAA Paper 88-4659, Sept 1988.

5 Wolfson, M.M., "Characteristics of Microbursts in the Continental United States," *The Lincoln Laboratory Journal*, Vol. 1, No. 1, Spring 1988, pp. 49-74.

6. Federal Aviation Administration, *Windshear Training Aid*, 1987.

7. National Center for Atmospheric Research, "Terminal Doppler Weather Radar (TDWR): A Briefing Paper," July 1, 1988.

8. Private Communication with McEachen, A. of Aeronautical Radio, Inc., August 1988.

9. Rucker, R.A., Flathers, G.W., "The Future for Aeromobile Digital Communications," *IEEE/AIAA 8th Digital Avionics Systems Conference*, October, 1988.

OHIO UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY 1990-1991

Robert W. Lilley
Avionics Engineering Center
Department of Electrical and Computer Engineering
Ohio University
Athens, Ohio

SUMMARY OF RESEARCH

This twentieth year of the Joint University Program (JUP) saw continued progress by students, faculty and staff in four major areas. Brief reports are given in this section.

- The study of spectrum-efficient methods for transmitting weather information to aircraft has resulted in definition of an improved amplitude-and-phase modulation process which permits re-utilization of voice channels for both voice and data. The system is being implemented for testing.

- Multi-sensor navigation systems introduce increased flexibility and performance for aviation, at the expense of increased complexity. Insuring a high level of system reliability and integrity requires that faults not only be detected, but also isolated to specific system elements, so that the remaining capabilities of the system may be used with confidence. An algorithm has been developed, and will be tested using combined GPS/INS systems.

- The JUP has served as a vehicle for reporting work accomplished in evoked-potential vision-tracking experiments to determine the engineering parameters of this input and control method. Successful detection of an operator's visual selection of objects has been achieved, and experiments continue to determine resolution and moving-target tracking capabilities.

- Application of GPS in an interferometric mode permits accurate measurement of differential motion; aircraft attitude may be determined using GPS only, with multiple antennas. The JUP team carried out the first known flight tests of this application.

Joint University Program participants are encouraged to report technical results in the open literature. The 1990-1991 papers and presentations are listed in the annotated bibliography on the following pages.

ANNOTATED BIBLIOGRAPHY OF 1990-91 PUBLICATIONS

1. Kline, P. A.: Availability of Positioning and Receiver Autonomous Integrity Monitoring for the Global Positioning System. Proceedings of the 3rd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 1990.

The Minimum Operational Performance Standards (MOPS) are being developed for airborne supplemental navigation equipment using the Global Positioning System (GPS). This is a unique task in the sense that the receiver must provide integrity assurance independent of that supplied by GPS. Independent integrity assurance, or receiver autonomous integrity monitoring (RAIM), requires at least one redundant GPS measurement. Therefore, the availability of GPS depends on the availability of RAIM. A computer simulation is presented for determining RAIM availability for GPS which incorporates a GPS coverage model and a Markov reliability model. The Markov model is used to assign state probabilities to GPS failure scenarios. Up to six simultaneous satellite failures are considered. A parametric analysis is presented to determine the effect on GPS RAIM availability. Some parameters that are included in the analysis are the maximum allowable Horizontal Dilution of Precision (HDOP), the Mean Time To Repair (MTTR) for failed GPS satellites, and the addition or omission of the altimeter as an additional measurement.

2. Farrell, J. L., and Van Graas, F.: That All-Important Interface. Proceedings of the 3rd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 1990.

The goal of nav systems integration should include growth flexibility, to accommodate the future addition of data sources and operations not envisioned in original configurations or plans. That aim is seriously compromised when information available from existent sources is deficient in content, form, timeliness, or precision. Unfortunately this is a common occurrence, not an occasional oversight; information is typically conveyed in ways that became standard long before modernization. Prime examples are attitude (expressed in terms of the familiar roll-pitch-heading convention) and velocity components in single precision.

These and other instances of accepted procedures are reviewed, along with an illustration of how various practices impose fundamental but completely unnecessary limitations on achievable integration performance. In accordance with available means of correcting all deficiencies, which have been widely known for years, straightforward measures are proposed whereby standards can be updated. The related issue of mounting location for critical elements on nonrigid structures is also briefly addressed. In all cases under consideration, the intent is to eliminate impediments to true system integration.

3. Van Graas, F.: In Flight Demonstration of Hybrid GPS/LORAN RAIM. Proceedings of the National Technical Meeting of the ION, Phoenix, AZ, January, 1991.

A receiver autonomous integrity monitoring (RAIM) scheme has been implemented in a real time, prototype hybrid GPS/LORAN receiver. A four-channel GPS receiver and an eight-channel Long Range Navigation (LORAN-C) receiver are used to obtain raw GPS and LORAN pseudorange measurements. The measurement data is sent to a microcomputer for position and integrity determination. Navigation data is presented to the pilot on a standard course deviation indicator with a data renewal rate of up to one update per second. Provisions are made to simulate signal malfunctions in real time by injecting step or ramp failures in the pseudorange measurements. A least squares fault detection algorithm is presented for the detection of slowly growing measurement errors. The performance of the hybrid GPS/LORAN receiver has been demonstrated through actual flight tests. One of the flight tests is summarized in the paper.

4. Braasch, M. S.: A Signal Model For GPS. Published in NAVIGATION: Journal of the Institute of Navigation, Vol. 37, No. 4, Winter 1990-91.

As the development of GPS continues, there will be an increasing need for a software-centered signal model. This model must accurately generate the observed pseudorange that would typically be encountered. The observed pseudorange varies from the true geometric range because of range measurement errors, which stem from a variety of hardware and environmental factors. In this paper, these errors are classified as either deterministic or random, and, where appropriate, their models are summarized. Of particular interest is the model for Selective Availability, which was derived from actual GPS data. The procedure for determination of this model, known as system identification theory, is briefly outlined. The synthesis of these error sources into the final signal model is given, along with simulation results.

5. Van Graas, F., and Braasch, M. S.: GPS Interferometric Attitude and Heading Determination: Flight Test Results. Proceedings of the 47th Annual Meeting of the Institute of Navigation, Williamsburg, VA, June 1991.

Attitude and heading determination using GPS interferometry is a well-understood concept. However, efforts have been concentrated mainly in the development of robust algorithms and applications for low dynamic, rigid platforms (i.e. shipboard). This paper presents results of what is believed by the authors to be the first realtime flight test of a GPS attitude and heading determination system. The system is installed in Ohio University's Douglas DC-3 research aircraft. Signals from four antennas are processed by an Ashtech 3DF 24-channel GPS receiver. Data from the receiver are sent to a microcomputer for storage and further computations. Attitude and heading data are sent to a second computer for display on a software generated artificial horizon. Demonstration of this technique proves its candidacy for augmentation of

aircraft state estimation for flight control and navigation as well as numerous other applications.

6. Van Graas, F., and Farrell, J.: Receiver Autonomous Integrity Monitoring (RAIM): Techniques, Performance and Potential. Proceedings of the 47th Annual Meeting of the Institute of Navigation, Williamsburg, VA, June 1991.

The goal of receiver autonomous integrity monitoring (RAIM) for the Global Positioning System (GPS) is to provide a sufficient level of navigation integrity for all phases of flight based on the signals transmitted by the GPS satellites only. Integrity requirements for airborne use of GPS are reviewed. This is followed by the description of a baseline fault detection algorithm which is shown to be capable of satisfying tentative integrity requirements. The related issue of testing the fault detection algorithm is also briefly addressed. Preliminary performance results for the baseline fault detection algorithm are presented, along with the potential of RAIM techniques for achieving GPS integrity.

7. Skidmore, T.: The Electroencephalographic Human-Computer Interface. Ph.D. Dissertation, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, June 1991.

This dissertation discusses the primary features of an electroencephalographic (EEG) human-computer interface. The paper begins with an overview of the human-computer interface problem and covers issues such as safety, cost, and complexity of technology. A detailed description of the equipment used throughout the research is given. This description includes the electrodes and their placement, the neurodata amplifier, the analog-to-digital converter, the analysis software, and the host computer. The foundation for realizing an electroencephalographic human-computer interface is based primarily on two possible paradigms. The first is a direct thought-controlled interface in which an attempt is made to correlate the EEG data with conscious thoughts of the subjects. Topics discussed in this area include raw-data analysis, chaos theory, autoregressive modelling, biofeedback and asymmetry spectral analysis. The second and most promising method for achieving an interface involves the use of visual evoked potentials. The primary evoked-potential issues presented are verification of the driving response, frequency separation, transient analysis and object multiplicity.

A HYBRID MODULATION FOR THE DISSEMINATION OF WEATHER DATA TO AIRCRAFT

Dennis M. Akos
Ohio University
Athens, Ohio

SUMMARY

Ohio University is continuing to conduct research to improve its system for weather data dissemination to aircraft. The current experimental system transmits compressed weather radar reflectivity patterns from a ground based station to aircraft. Although an effective system, the limited frequency spectrum does not provide a channel for transmission. This introduces the idea of a hybrid modulation. The hybrid technique encodes weather data using phase modulation (PM) onto an existing aeronautical channel which employs amplitude modulation (AM) for voice signal transmission. Ideally, the two modulations are independent of one another. The planned implementation and basis of the system are the subjects of this paper.

BACKGROUND

A system for weather data dissemination to aircraft was developed at Ohio University to improve weather uplink service to general aviation aircraft. This system obtained weather radar reflectivity patterns from the National Weather Service via telephone lines. Weather radar reflectivity patterns are utilized since the transmission of such images has been identified as a method of improving aircraft safety. The image is digitized, compressed using compact codes and run length encoding, modulated using Quadrature Phase Shift Keying (QPSK), and transmitted over a Very High Frequency (VHF) aeronautical channel. In the aircraft, data is demodulated and processed so that the image can be displayed (ref. 1). This system was implemented successfully using a data rate of 2400 bits/second which provides a mean image time of ten seconds per image.

This system offers a potential improvement over the inadequate weather uplink service now in use. The remaining obstacle concerns its placement in the already overcrowded frequency spectrum. A possible solution may be obtained through the use of a hybrid modulation, utilizing both amplitude and phase modulation on the same channel. Currently, voice communication between ground and aircraft is accomplished using amplitude modulation of a VHF carrier. Ideally, this carrier can be phase modulated with the weather data with no interference between the two modulations.

HYBRID MODULATION

A signal using both amplitude and phase modulations can be expressed by:

$$A_c \cdot [1 + k_v\, m(t)] \cdot \cos[2\pi f_c t + \theta(t)]$$

Where m(t) is the amplitude modulated voice signal and $\theta(t)$ is the phase modulated data. Ideally, these two modulations are independent of one another, but due to the band and amplitude limiting necessary for transmission, an interference mechanism is introduced (ref. 2). Generally, a phase modulated signal retains a constant amplitude. However, when this signal is filtered so that its higher frequency energy is removed, envelope variations result which directly interfere with the AM voice in a hybrid modulation. It is important to select a phase modulation which is spectrally efficient, thereby containing less high frequency energy. One such phase modulation method is Minimum Shift Keying (MSK). Research regarding an AM voice/MSK data hybrid signal has shown that the system is feasible under the following constraints: (1) data modulated at a maximum of 2400 bits/second; (2) AM modulated by voice with an index of modulation limited to 0.7; and (3) standard VHF AM communications equipment for both transmitter and receiver (refs. 1-2). It is important to note that the hybrid modulation can utilize the same data rate as the current experimental system.

CURRENT RESEARCH

Although the hybrid AM/MSK system has been studied in theory, it has yet to be implemented. This implementation is the basis of the current research. A Digital Signal Processing (DSP) chip, the TMS320C30 by Texas Instruments, will be used to generate the quadrature and in-phase components of the MSK signal based on digitized weather radar data. The components will be mixed with a carrier generated by a standard VHF AM transmitter for transmission of the hybrid signal. Aboard the aircraft a second DSP chip will be placed in parallel with the AM voice receiver so that data can be demodulated and processed for image display. The proposed implementation of this experimental system is shown in figure 1.

A computer simulation was developed in order to verify to degradation of the AM voice signal due to the envelope variations of the filtered PM signal. The simulation uses an actual voice signal so that the degradation can be judged subjectively. This allows for the testing of the maximum data rates based on voice degradation. The simulation is done by sampling an analog voice signal and using random data bits to generate AM/PM modulated signals. These signals are

then passed through two filters, one representing a transmission filter and a second representing the receiver filter. Each sample is then amplitude demodulated to recover the distorted voice samples which are then transformed back into an analog signal for subjective judgment.

Possible alternatives to MSK are being researched also. There exists an entire class of continuous phase modulations which are spectrally more efficient than MSK. Among these are Gaussian MSK (GMSK), Tamed Frequency Modulation (TFM), and Generalized TFM (GTFM) (refs. 3-5). These modulation methods employ a pre-modulation filter to reduce high frequency spectral energy. With a more compact power spectral density, the filtering required for transmission should produce less envelope variation than if MSK modulation is used. The voice degradation simulation should be useful in verifying these assumptions. If these modulation methods prove to be superior to MSK, the overall transmitter design need only be changed by reprogramming the DSP chip, a fairly simple procedure. Much research has yet to be done on these alternative phase modulations, especially regarding determination of the bit error rate associated with each in a hybrid modulation.

CONCLUSION

This paper has described a spectrally efficient technique of combining both an amplitude and phase modulation on the same carrier for the dissemination of weather data to aircraft. The technique, which has been shown to be feasible theoretically, is currently being implemented in an experimental system. This system will allow weather data reflectivity patterns along with voice information to be transmitted using the same frequency spectrum now reserved solely for voice communications. It should also be noted that existing AM receivers can still be used with no modification to demodulate the voice signal. The system also can be used to transmit any other relevant information which might be useful in a data-link.

REFERENCES

1. Parker, C. B.: A Technique for the Automated Dissemination of Weather Data to Aircraft. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, June 1989.

2. Benelli, G. and Fantacci, R.: An Integrated Voice-Data Communication System for VHF Links. IEEE Transactions on Communications, Vol. COM 31, No. 12, December 1983.

3. Murota, K.: GMSK Modulation for Digital Mobile Radio Telephony. IEEE Transactions on Communications, Vol. COM 29, No. 7, July 1981.

4. Jager, F. and Dekker, C. B.: Tamed Frequency Modulation, A Novel Method to Achieve Spectrum Economy in Digital Transmission. IEEE Transactions on Communications, Vol. COM 26, No. 5, May 1978.

5. Chung, K. S.: Generalized Tamed Frequency Modulation and its Application for Mobile Radio Communications. IEEE Journal on Selected Areas in Communications, Vol. SAC-2, No. 4, July 1984.

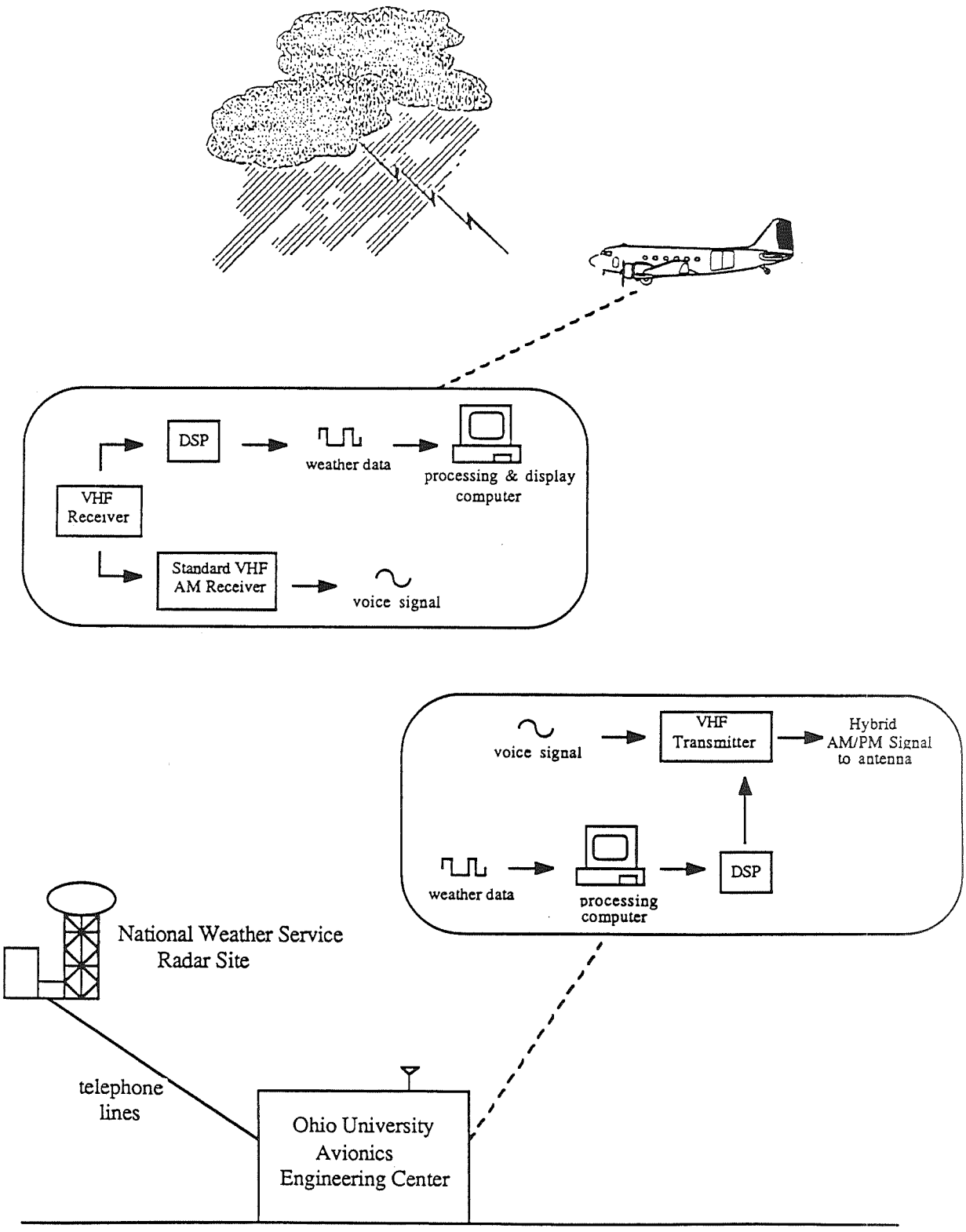


Figure 1. Proposed Weather Data Dissemination System

FAULT DETECTION AND ISOLATION FOR MULTISENSOR NAVIGATION SYSTEMS

Paul A. Kline and Frank van Graas
Ohio University
Athens, Ohio

SUMMARY

Increasing attention is being given to the problem of erroneous measurement data for multisensor navigation systems. A recursive estimator can be used in conjunction with a "snapshot" batch estimator to provide fault detection and isolation (FDI) for these systems. A recursive estimator uses past system states to form a new state estimate and compares it to the calculated state based on a new set of measurements. A "snapshot" batch estimator uses a set of measurements collected simultaneously and compares solutions based on subsets of measurements. The "snapshot" approach requires redundant measurements in order to detect and isolate faults. FDI is also referred to as Receiver Autonomous Integrity Monitoring (RAIM).

BACKGROUND

The objective is to detect and isolate sensor malfunctions which cause unacceptably large position errors using only inconsistency in the measurement data. Previously, FDI has been successfully applied to redundant inertial navigation systems (refs. 1-3). However, FDI can be used for any multisensor navigation system, including systems based on the Global Positioning System (GPS), Long Range Navigation (Loran-C) and the Global Navigation Satellite System (GLONASS).

A recursive estimator, such as the Kalman filter, uses the history of the user navigation state to form a predicted state estimate. If the difference between the predicted state estimate and the calculated state (based on a new set of measurements) is too large, a fault is declared. This approach is excellent for detecting step errors or rapidly growing ramp errors, and requires no redundant measurements. However, if a measurement ramp error with a small slope enters the system (caused by for instance an uncorrected clock drift in a GPS satellite), the Kalman filter may "smooth" the data rather than declare a fault. To solve this problem, a batch estimator is used since it does not depend on the history of the user state.

The "snapshot" batch estimator is based on a least squares solution which requires at least one redundant measurement for detection, and at least two redundant

measurements for isolation. Solutions calculated based on subsets of measurements can be compared to determine if inconsistency exists. Once a fault is declared, isolation is accomplished by applying detection techniques to subsets formed by leaving out one measurement at the time. This procedure is successful if no fault is found when the faulty measurement is omitted and a fault is declared for each subset containing the erroneous measurement. The focus of the remainder of this paper is on the characterization of "snapshot" batch estimators to perform FDI.

LEAST SQUARES FAULT DETECTION

One of the main input parameters to the fault detection algorithm is the alarm threshold, defined as the allowable horizontal radial error in the calculated user position. The ideal case would be to raise a flag only when this limit is exceeded and never raise a flag otherwise. However, since fault detection is performed in the presence of measurement noise and in a domain other than the solution space, it is only possible to detect a fault with a certain probability. Therefore, two undesirable events are possible - a false alarm and a missed detection. Two major parameters used in characterizing the performance of the fault detection algorithm are the probability of a false alarm (P_{FA}) and the probability of a missed detection (P_{MD}). As one might assume, it is desirable for these probabilities to be very small.

A fault is declared when a detection statistic exceeds a certain detection threshold (T_D). Two cases of a ramp error are shown in figure 1. In case I, T_D is breached before the alarm threshold is crossed, causing a false alarm. As the position error grows, the false alarm becomes a correct alarm. In case II, the alarm threshold is exceeded before the detection threshold is exceeded, resulting in a missed detection. Eventually T_D is crossed, causing a flag to be raised for a correct alarm. The normal operating state includes all circumstances where neither threshold is exceeded. For multisensor systems, this state should have a large probability of occurrence.

FAULT DETECTION ALGORITHM

A least squares approach can be used for fault detection. The linear relationship between the measurements and the user state is given by:

$$\underline{y} = H \underline{\beta} \tag{1}$$

where: $\underline{y}$ = measurement vector
$\underline{\beta}$ = user state vector
H = data matrix

The dimension of H is n-by-m, where n is the number of measurements and m is the dimension of the user state vector. The user state vector $\underline{\beta}$ consists of the user

position coordinates and other navigation state elements such as clock offset with respect to, for instance, GPS time, as required by the navigation solution.

Three cases exist:

1) $n < m$: Underdetermined system
2) $n = m$: Exactly determined system
3) $n > m$: Overdetermined system

Algorithms for managing the redundant measurements in case 3, an overdetermined system, form the basis of fault detection. In the presence of redundant signals, a parity equation can be derived from equation (1). First, a QR factorization is performed on the data matrix H (ref. 4):

$$H = QR \tag{2}$$

This factorizes H into an orthogonal matrix Q ($Q^TQ = I$) and an upper triangular matrix R. R contains (n-m) rows of zeros, reflecting that H includes data from redundant measurements. Substituting equation (2) for H in equation (1) yields:

$$\begin{aligned} \underline{y} &= QR\underline{\beta} \\ Q^T\underline{y} &= Q^TQR\underline{\beta} \\ Q^T\underline{y} &= R\underline{\beta} \end{aligned} \tag{3}$$

Let R be partitioned into an m-by-m upper triangular matrix U and (n-m) rows of zeros, denoted by 0. Partition Q^T conformably into Q_1 and Q_2.

$$\begin{pmatrix} Q_1 \\ --- \\ Q_2 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} U \\ --- \\ 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \tag{4}$$

The least squares solution is given by:

$$\underline{\beta} = U^{-1}Q_1\underline{y} \tag{5}$$

Note that U is nonsingular due to the way R is partitioned. The parity equation is:

$$Q_2\underline{y} = 0 \tag{6}$$

Since $\underline{y}$ contains measurement errors such as noise ($\underline{e}$) and measurement biases ($\underline{b}$), a parity vector ($\underline{p}$) can be defined as (replace $\underline{y}$ by $\underline{y} + \underline{e} + \underline{b}$):

$$\underline{p} = Q_2\underline{y} + Q_2\underline{\epsilon} + Q_2\underline{b}$$
$$\underline{p} = Q_2\underline{\epsilon} + Q_2\underline{b} \qquad (7)$$

Although the measurement noise and bias errors are not known, their components in parity space are given by equation (7). The parity vector can be used as a detection function for declaring faults.

PARITY SPACE AND ESTIMATION SPACE

As an example, consider a scenario where one redundant measurement is available. In this case, the parity vector becomes a scalar. The detection statistic is given by $|p|$, which is assumed to be normally distributed. Figure 2 shows the distribution of p when no bias exists in any of the measurements. A fault is declared when the detection threshold (T_D) is exceeded. Note that integrating the area under the tails outside T_D yields the probability of a false alarm in parity space.

Figure 3 illustrates the existence of a bias in one of the measurements. The distribution of the detection statistic p is shifted over, and the area under the curve within the limits of T_D is the probability of a missed detection in parity space.

If an alarm is raised in parity space, it can either be a correct alarm or a false alarm in estimation space. If no alarm is raised in parity space, it can either be normal operation or a missed detection in estimation space. Thus, the definitions of a false alarm and a missed detection are slightly changed. References 5 and 6 contain a detailed explanation of the relation between parity space and estimation space.

FAULT ISOLATION CONCEPTS

Let n be the number of available measurements. Then k detection functions are formed ($k = 1,\ldots,n$; one for each measurement), and an alarm is raised if any $|d_k| > T_D$. Once alarm status has been reached, the next step is to attempt isolation. At least two redundant measurements are required for this. The process uses the fault detection algorithm applied to all subsets created by leaving out one measurement at the time. By omitting the failed measurement, the detection functions for that subset should all lie within T_D. By omitting a healthy measurement, at least one detection function should exceed T_D.

The number of detection functions used for isolation is n(n-1). Let $d_{m,k}$ be the detection function for the k^{th} satellite with satellite m omitted (note that k=m does not exist). Successful isolation occurs when:

1) If m = failed satellite, all $|d_{m,k}| < T_D$

and 2) If m ≠ failed satellite, at least one $|d_{m,k}| > T_D$

Figure 4 shows the fault detection and isolation state diagram. From the diagram, it can be seen that the probability of a false alarm should be very small, because it results in either removing a healthy measurement or in system unavailability. The latter is very undesirable for a sole means of navigation system. Furthermore, the probability of a wrong isolation should also be minimized. The goal of the current research is to define the transitional probabilities of the state diagram given the required state probabilities for navigation systems.

CONCLUSIONS

A fault detection and isolation algorithm is presented for use in a multisensor navigation system. A state diagram has been developed which incorporates all important system states for the FDI process. Efforts continue on the quantification of state transition probabilities such that navigation system requirements for fault detection and isolation can be satisfied.

REFERENCES

1. Daly, K. C., Gai, E. G., and Harisson, J. V.: Generalized Likelihood Test for FDI in Redundant Sensor Configurations. Journal of Guidance and Control, Vol. 2, Jan.-Feb. 1979.

2. Gai, E. G., Adams, M. B., and Walker, B. K.: Determination of Failure Thresholds in Hybrid Navigation. IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-12, Nov. 1976.

3. Morrell, F. R., Bailey, M. L., and Motyka, P. R.: Flight Demonstration of Redundancy Management Algorithms for a Skewed Array of Inertial Sensors. Proceedings of the AIAA/AHS/ASEE Aircraft Design and Operations Meeting, AIAA Paper 88-4434, Atlanta, GA, September 7-9, 1988.

4. Golub, G. H., and Van Loan, C. F.: Matrix Computations, Second Edition. The Johns Hopkins University Press, Baltimore, MD, 1989.

5. Brenner, M.: Implementation of a RAIM Monitor in a GPS Receiver and an Integrated GPS/IRS. Proceedings of the Third International Satellite Division Meeting of the Institute of Navigation, September 1990.

6. Van Graas, F. and Farrell, J. L.: Receiver Autonomous Integrity Monitoring (RAIM): Techniques, Performance and Potential. Proceedings of the 47th Annual Meeting of the Institute of Navigation, Williamsburg, VA, June 10-12, 1991.

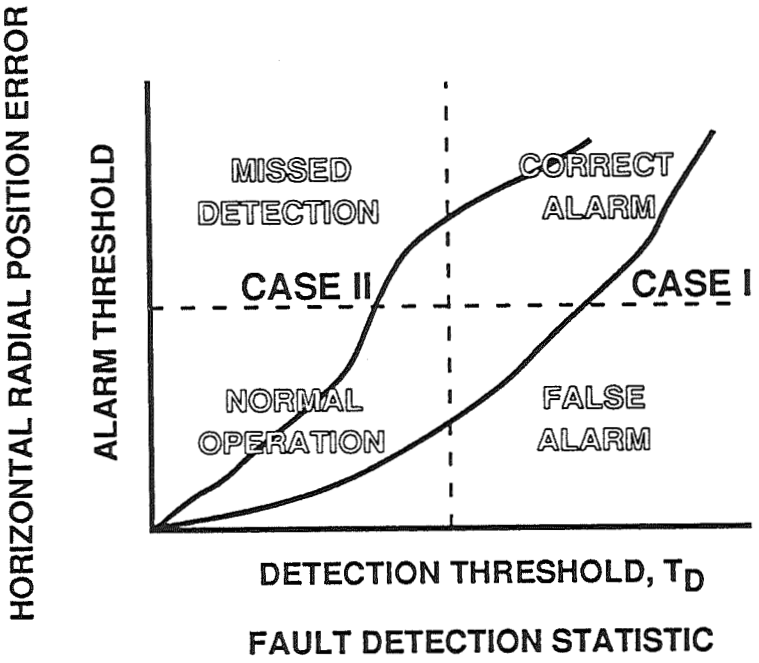


Figure 1. Example of the impact of slowly growing measurement errors on fault detection.

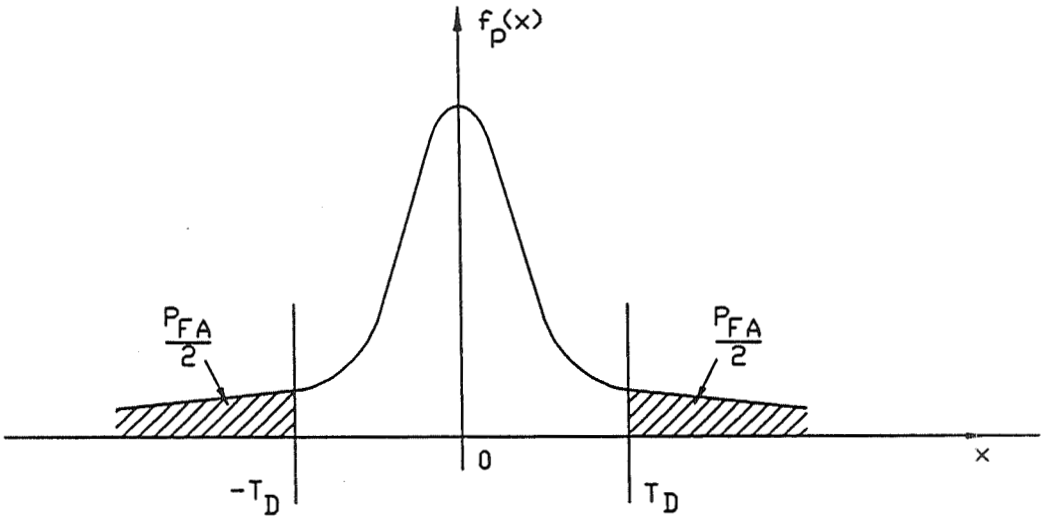


Figure 2. Probability Density Function for p - no measurement bias.

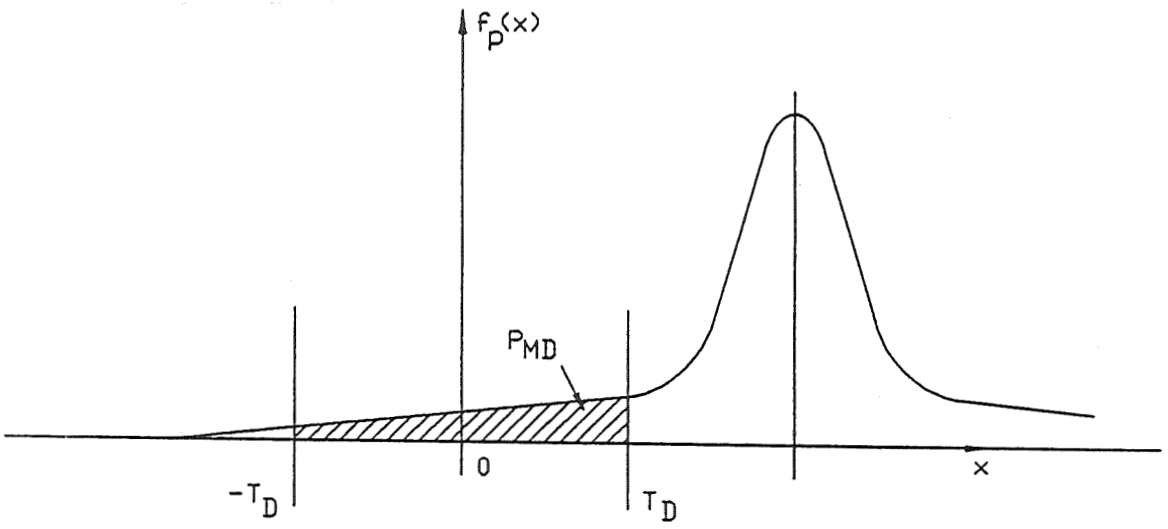


Figure 3. Probability Density Function for p - with measurement bias.

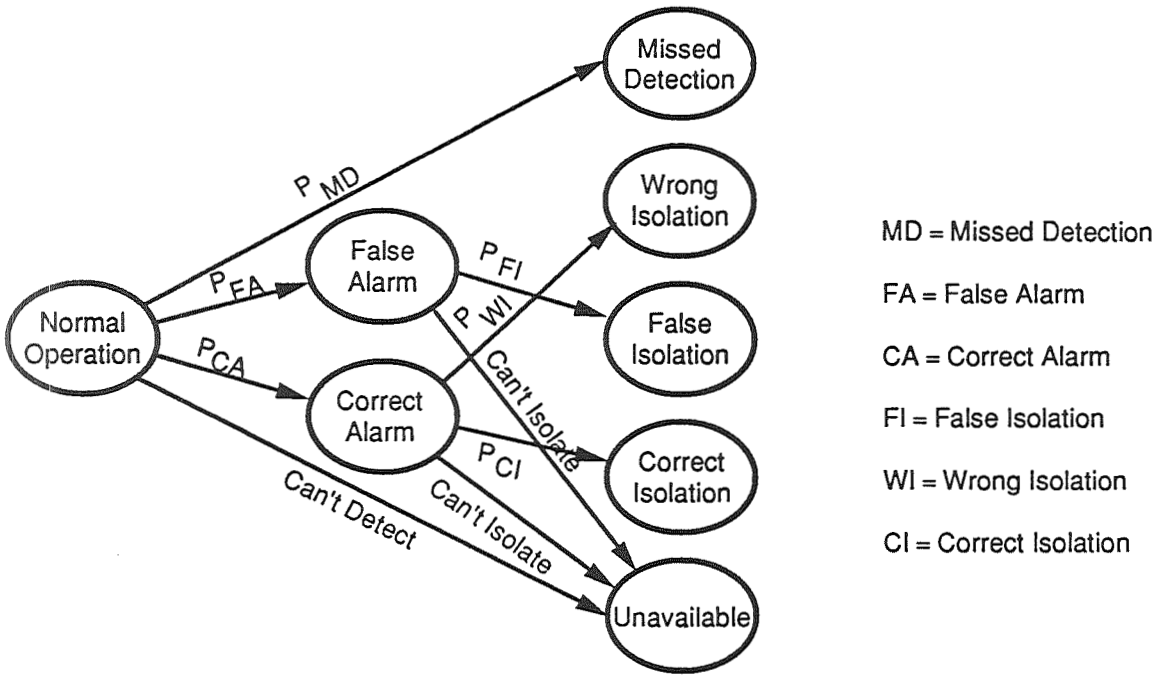


Figure 4. Fault Detection and Isolation State Diagram.

ESTABLISHING AN EVOKED-POTENTIAL VISION-TRACKING SYSTEM

Trent A. Skidmore
Ohio University
Athens, Ohio

SUMMARY

This paper presents experimental evidence to support the feasibility of an evoked-potential vision-tracking system. The topics discussed are stimulator construction, verification of the photic driving response in the electroencephalogram, a method for performing frequency separation, and a transient-analysis example. The final issue considered is that of object multiplicity (concurrent visual stimuli with different flashing rates). The paper concludes by discussing several applications currently under investigation.

INTRODUCTION

The primary use of visual evoked potentials (VEPs) has been in clinical applications (refs. 1-2). However, because VEPs involve conceptually simple instrumentation, are relatively safe, and inherently noninvasive, they are worth investigating for possible use in other applications. One such application would be an evoked-potential vision-tracking system. To demonstrate the feasibility of such a system, it will be shown that distinct frequency-separable components can be evoked in the scalp potentials recorded using conventional electroencephalographic (EEG) techniques.

THE STIMULATOR

A flexible stimulator was needed to investigate the issues involved in an evoked-potential interface. To realize the stimulator, a personal computer was used, equipped with the C programming language, a high-resolution video card and monitor. Several programs were generated that could flash a white object (either a square or arrow) on a black background at a specified rate. Computer hardware constrained the flashing rates to be integer submultiples of the 70.1 Hertz (Hz) screen update rate. This led to the four primary frequencies being 35.050, 23.367, 17.525, and 14.020 Hz. While slower flashing rates were possible, it was deemed that, at least for this initial investigation, a minimum of 14.020 Hz should be imposed. This was based primarily on the need to avoid the alpha band (8-13 Hz) and to minimize viewer irritation.

To ensure that the stimulator was performing as expected, data were collected using a photo darlington transistor. With the photo transistor taped to the monitor (stimulator), a 1" by 1" white square was flashed against a black background. For each of the chosen frequencies, the output of the photo transistor was seen to be a periodic train of pulses at the expected rate. Spectral analysis showed that each stimulating frequency produced a component at the fundamental as well as at higher harmonics. Thus, based on these findings, it was concluded that the stimulator was performing correctly, generating driving frequencies at the desired rates.

VERIFYING THE PHOTIC DRIVING RESPONSE

To verify that the evoked potentials were detectable in the EEG, data were collected under five test conditions. These included flashing the 1" by 1" white square (at a viewing distance of 36") for two, sixteen-second runs at each of the rates listed above (35.050, 23.367, 17.525 and 14.020 Hz). In addition, to form a control set, data were collected for a run in which the subject looked at an illuminated black screen with no flashing object. The EEG was recorded (at a rate of 256 samples/second) over the O_1 and O_2 locations, each referenced to C_Z, according to the International 10-20 System of Electrode Placement. Sample spectra from the 35.050-Hz and 23.367-Hz runs are shown in figure 1, with the stimulating frequencies circled. To generate these plots, a hybrid periodogram was calculated. This periodogram averaged the spectra from both electrode pairs over two separate sixteen-second runs. This was done to highlight the peaks and to effectively reduce the surrounding noise. It is, however, crucial to note that the desired frequency peaks were also visible in the unaveraged spectra. The periodogram was performed merely to enhance the photic driving effect.

FREQUENCY SEPARATION

To show that an evoked-potential interface is possible, it is important to verify that the four driving rates are distinguishable. To demonstrate this, sixteen seconds of O_1-C_Z data for each rate were passed through four separate band-pass filters. The filters were fourth-order, infinite-impulse response Butterworth filters implemented in the LabVIEW programming language. The bandwidth of each filter was 1 Hz, with a center frequency corresponding to each driving rate (35.050, 23.367, 17.525 or 14.020 Hz). The spectral output of two of these filters is shown in figure 2. The legend indicates the notation used for each photic driving frequency. In both plots it is clear that the particular driving frequency being tested produces the largest spectral output. For example, in the top graph, data from the four rates were passed through the 35.050-Hz filter. As can be seen, the run containing the 35.050-Hz driving frequency, marked with the solid squares, has the largest spectral component at the output of the 35.050-Hz band-pass filter. In every case, each driving frequency produced the largest response in its associated filter.

From these results it is apparent that band-pass filtering can be used to detect the dominant frequency in the steady-state evoked potential. While it is beyond the scope of this paper to discuss frequency

detection schemes in detail, it has been demonstrated that separation and detection of the dominant photic driving frequency are possible.

TRANSIENT ANALYSIS

To demonstrate the transient nature of the eye/brain system as the frequency of a single stimulating object is varied, it was necessary to devise a method for measuring time-dependent frequency changes. One manner of doing this is through the use of the spectrogram technique. To understand the operation of the spectrogram, suppose that a signal consists of the following components,

$$
\begin{aligned}
x(t) &= 0 && 0 < t < 4 \text{ seconds} \\
&= 4\cos(2\pi(10)t) && 4 < t < 8 \text{ seconds} \\
&= 3\cos(2\pi(20)t) && 8 < t < 12 \text{ seconds} \\
&= 2\cos(2\pi(30)t) && 12 < t < 16 \text{ seconds} \\
&= 1\cos(2\pi(40)t) && 16 < t < 20 \text{ seconds} \\
&= 1\cos(2\pi(30)t) && 20 < t < 24 \text{ seconds} \\
&= 2\cos(2\pi(20)t) && 24 < t < 28 \text{ seconds} \\
&= 3\cos(2\pi(10)t) && 28 < t < 32 \text{ seconds}
\end{aligned}
$$

and is sampled at 256 samples/second. A spectrogram of this signal is shown in figure 3. The plot consists of thirty-two one-second records or epochs. Each one-second record is the Fourier transform of a one-second block of time. The height of each record represents a linear measure of the spectral content during that particular time interval. The units on the horizontal axis are Hertz and the frequency binwidth (resolution) is 1 Hz. Observe from the example how the spectrogram plot "hides" lines of lower magnitude behind peaks that occur earlier in time. This convention allows time-dependent spectral trends to be observed while minimizing plot clutter.

The results of applying the spectrogram technique to demonstrate time-dependent spectral changes are highlighted in figures 4 and 5. Data for this test were collected when the stimulator was alternately flashing a 35.050-Hz and 23.367-Hz 1" white square on a black background. The stimulator initially displayed a brief message instructing the subject to watch the square. The stimulator then alternated between the two frequencies every five seconds. A total of forty-five one-second records are represented in the figures. The data was recorded over O_1 with respect to C_Z.

To generate the spectrogram and to effectively illustrate the transitions between frequencies, the data were first passed through a parallel-filter network. This network consisted of two band-pass filters centered at 35.050 Hz and 46.733 Hz. The bandwidth of each filter was 4 Hz. The 35.050-Hz filter was used to isolate the 35.050-Hz component and a 46.733-Hz filter was used to isolate the 23.367-Hz component. The reason for using the 46.733-Hz filter is twofold: For this subject, the 23.367-Hz stimulator actually produced more distinct spectral peaks at the first harmonic (46.733 Hz) than at the fundamental. Furthermore, because of the 1-Hz binwidth limitation of this particular test, the 46.733-Hz

filter captured the 23.367-Hz stimulator effect better than a 23-Hz filter.

To further illustrate the transient effects, consider figure 5. These graphs show two frequency-slices from the spectrogram of figure 4. The top graph illustrates the spectral changes in the 35-Hz bin. The bottom graph illustrates the spectral changes in the 47-Hz bin. The numbers placed directly on the graphs indicate the fundamental stimulating component in that time range. For example, for $1 \leq t \leq 5$ seconds, the stimulator was flashing 35.050 Hz. For $6 \leq t \leq 10$ seconds, the stimulator was flashing 23.367 Hz, and so on. From these plots it is clear that the evoked-potential method can track time-dependent spectral changes as perceived by the eye/brain system. Furthermore, it is apparent that harmonics of the stimulating frequency are also of importance and may be of use in certain detection schemes.

OBJECT MULTIPLICITY

Object multiplicity is the problem of having more than one flashing region within the field of vision. In order to achieve a viable interface, it is necessary for the eye/brain system to be able to distinguish between various flashing objects. To investigate this concept, a simple test scenario was devised. In this test, the monitor was made to flash two horizontal arrows, one at 35.050 Hz and the other at 23.367 Hz. Figure 6 shows the dimensions and spacing of the arrows. The viewing distance was again 36". Figure 7 shows the spectral results when the horizontal separation {s} of the arrows was 3.5". For the upper graph, the subject was instructed to look at the 35.050-Hz arrow (placed at screen right). As expected, there is a large spectral peak at this frequency and no significant 23.367-Hz component. The lower graph shows the spectrum generated when the subject was focused on the 23.367-Hz flashing arrow. Note that the dominant peak is again at the driving frequency under observation.

To further test the object-multiplicity issue, the arrows were moved to a separation of only 1.75". The results from this test were similar to those described above. As a final test, the arrows were flashed with no horizontal separation {s=0"}. Again, the driving frequencies were surprisingly distinguishable. Therefore, based on these results, the eyes appear to be able to focus on a region that is sufficiently small so as to allow this type of system to be used in a direct human-computer interface.

APPLICATIONS

There are many possible applications for an evoked-potential vision-tracking system. Two applications that are currently being investigated include a two-dimensional directional controller (fig. 8), and a moving-target tracking system (fig. 9). In the figure-8 scenario, the user observes and watches a stimulator that is displaying a four-arrow arrangement. Each arrow is flashed at a different frequency. This is done in order to evoke a frequency-separable photic driving response that is dependent upon which arrow the subject is looking at. For example, if the subject is focusing on the "up" arrow, flashing at a frequency of f_1 (flashes per second), then an f_1-Hz component should manifest itself in the EEG.

Subsequently, if the user changes focus from the "up" arrow to, say the "right" arrow, then the primary frequency being driven into the EEG will be changed from f_1 to f_2. Using a detector that would recognize these frequency changes, a user could systematically focus on a particular sequence of arrows to perform a two-dimensional control task, such as moving a computer's mouse.

Figure 9 illustrates one possible realization of a moving-target tracking system. For this, the user would watch a monitor (stimulator) displaying, for example, several moving airplanes, each flashing at different rates. Upon determining which airplane the user was focused on, the computer would highlight that plane. Based solely on the photic driving response of the highlighted object, the application would be capable of tracking the eye's movement as the airplane (target) moves across the screen.

CONCLUSIONS

Several important issues have been discussed that relate to the feasibility of an evoked-potential vision-tracking system. It has been shown that the photic driving response is apparent in the spectra of the EEG and is distinguishable through band-pass filtering. It has also been shown that the transient properties of the photic driving response are well defined, indicating that the eye/brain system is capable of discerning time-dependent frequency changes. Finally, it has been demonstrated that dual stimulating objects also appear to generate responses that are unique and separable. Based on these results, a conceptual foundation for a vision-based interface has been established.

ACKNOWLEDGEMENT

The work reported here was in part funded through NASA Grant NAG 1-940 and is the subject of a patent disclosure (No. 18/4/91) filed at Ohio University April 9, 1991.

REFERENCES

1. Barber, C. and Blum, T. eds.: Evoked Potentials III - The Third International Evoked Potentials Symposium, 1987.

2. Bodis-Wollner, I. ed.: Evoked Potentials, Annals of The New York Academy of Sciences, Vol. 388, 1982.

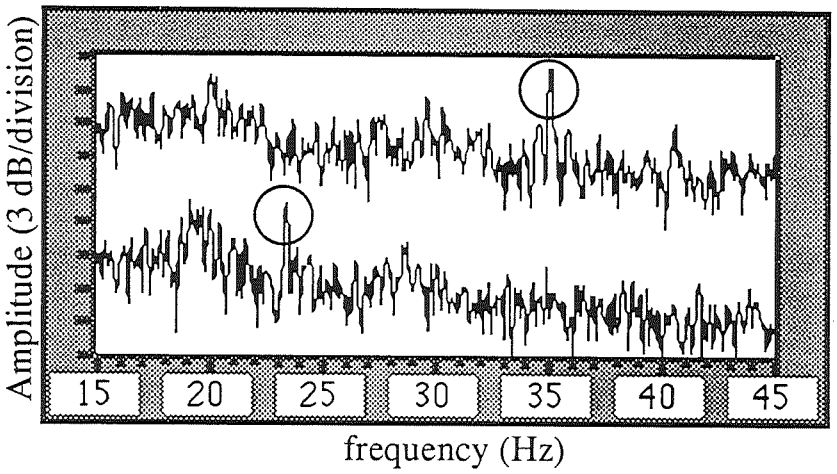


Figure 1. Photic Driving Response at 35.050 and 23.367 Hz.

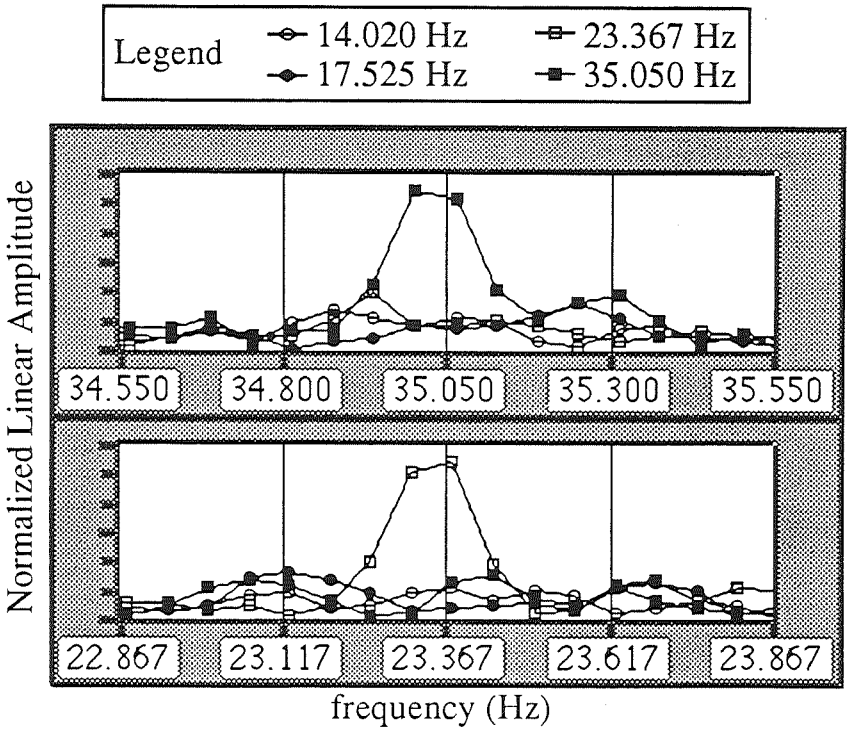


Figure 2. Frequency Separation Using Band-Pass Filtering.

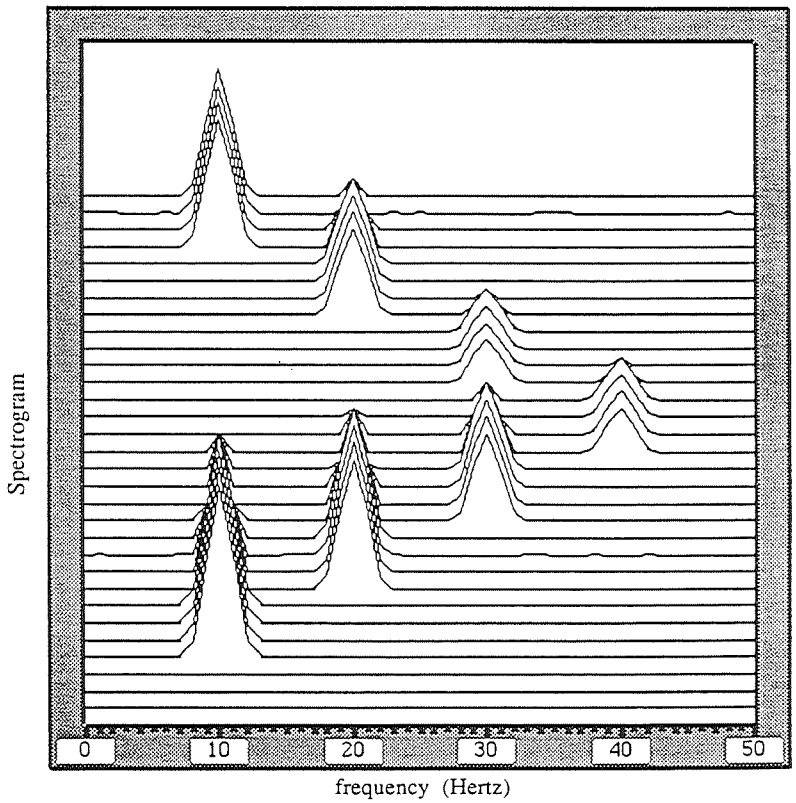


Figure 3. Spectrogram example.

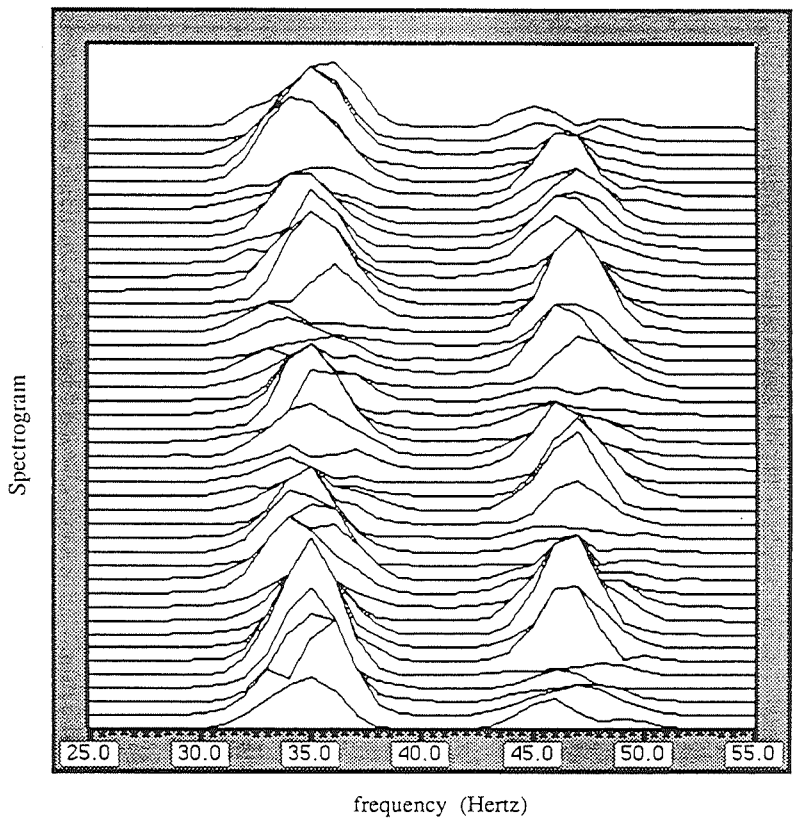


Figure 4. Spectrogram for the output of 35- and 47-Hz band-pass filters.

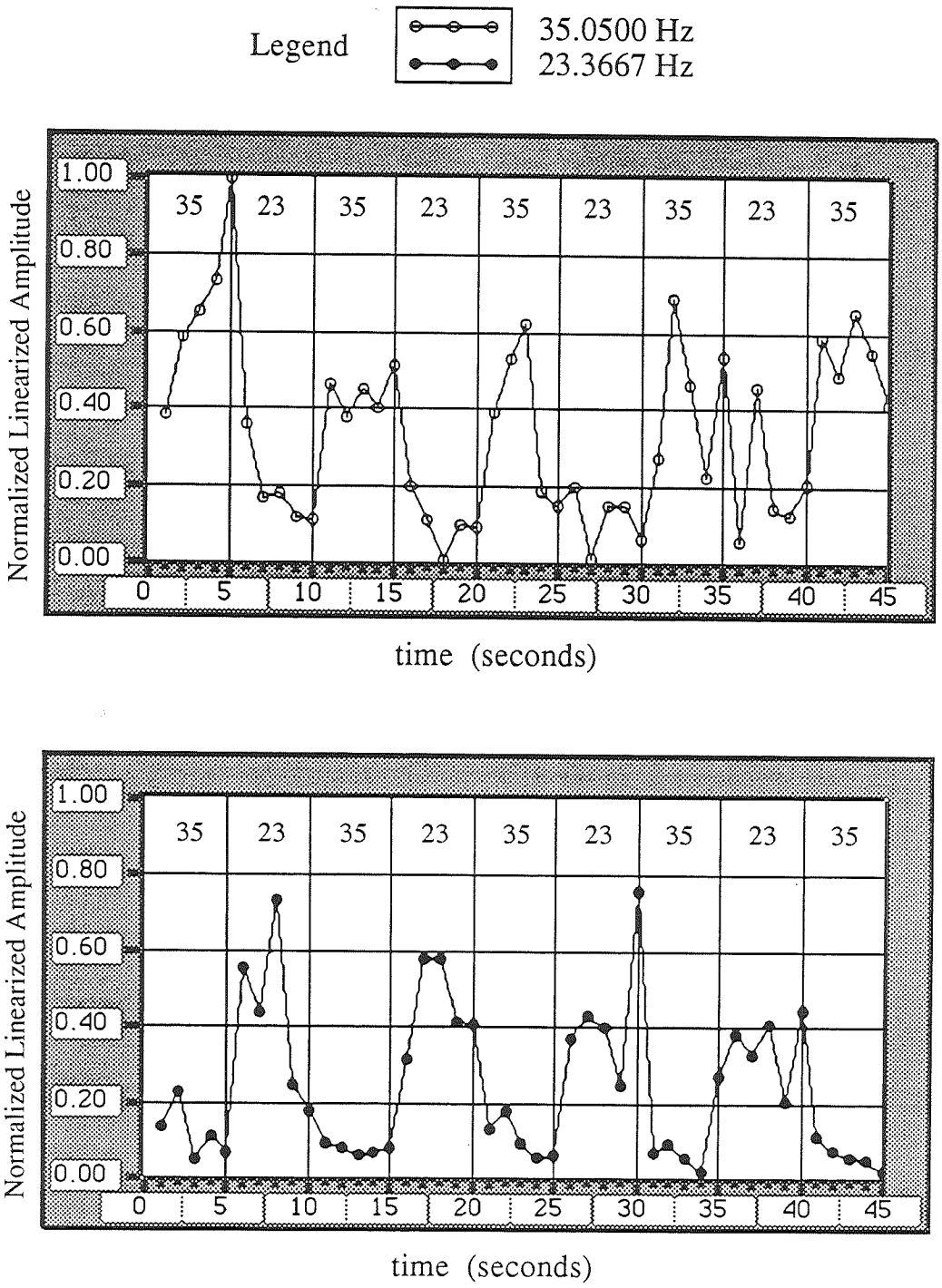


Figure 5. 35-Hz (top) and 47-Hz (bottom) spectrogram slices.

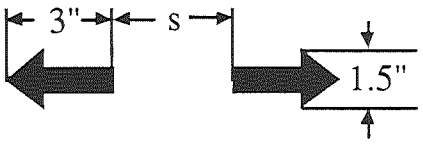


Figure 6. Object Multiplicity Test.

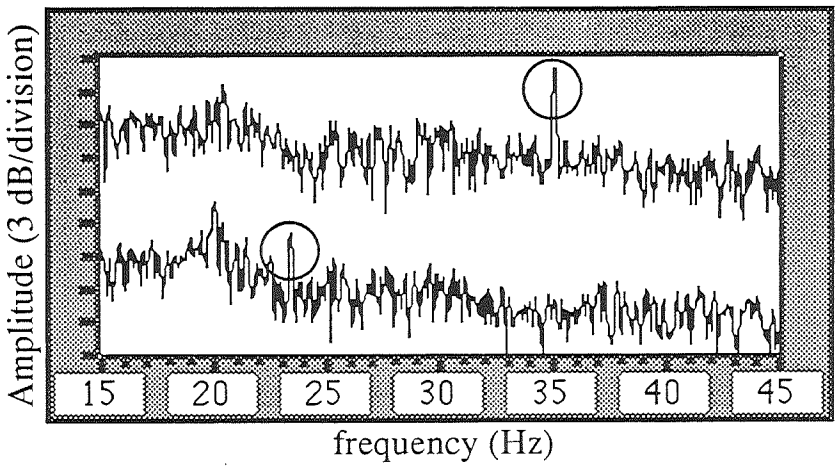


Figure 7. Object Multiplicity at 35.050 and 23.367 Hz.

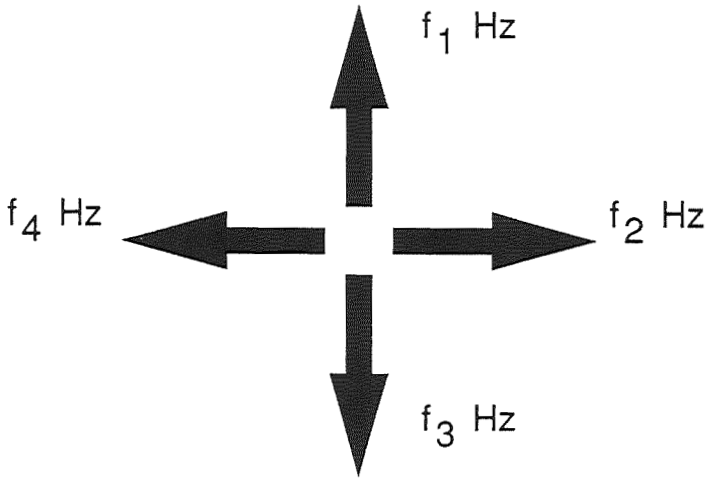


Figure 8. Four-arrow evoked-potential interface.

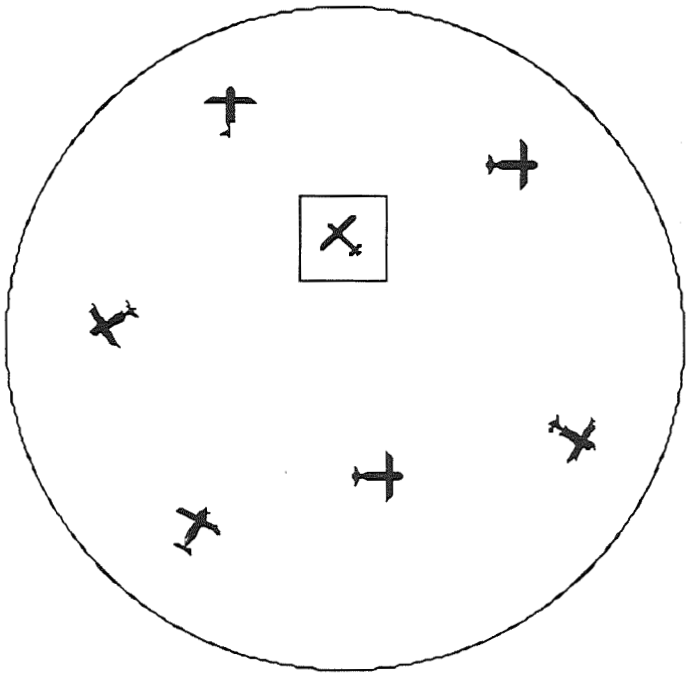


Figure 9. The target-tracking evoked-potential interface.

GPS INTERFEROMETRIC ATTITUDE AND HEADING DETERMINATION: INITIAL FLIGHT TEST RESULTS

Frank van Graas and Michael Braasch
Ohio University
Athens, Ohio

SUMMARY

Attitude and heading determination using GPS interferometry is a well-understood concept. However, efforts have been concentrated mainly in the development of robust algorithms and applications for low dynamic, rigid platforms (e.g. shipboard). This paper presents results of what is believed by the authors to be the first realtime flight test of a GPS attitude and heading determination system. The system is installed in Ohio University's Douglas DC-3 research aircraft. Signals from four antennas are processed by an Ashtech 3DF 24-channel GPS receiver. Data from the receiver are sent to a microcomputer for storage and further computations. Attitude and heading data are sent to a second computer for display on a software generated artificial horizon. Demonstration of this technique proves its candidacy for augmentation of aircraft state estimation for flight control and navigation as well as for numerous other applications.

INTRODUCTION

The Global Positioning System (GPS) will soon be capable of providing worldwide positioning and timing services to properly equipped users. In addition, interferometric techniques can be used to obtain realtime, highly accurate attitude and heading information from the GPS measurement data as well. GPS interferometry is a technique based on the measurement of phase differences of the GPS carrier signal observed at two or more antennas. Attitude and heading of a platform can be calculated from measurements with two linearly independent interferometers where each interferometer consists of two separate antennas.

GPS interferometry has been used extensively in precise static and kinematic positioning. Experimental results of static relative positioning using GPS interferometry were first reported in 1982 (ref. 1), showing subcentimeter accuracies for baseline lengths between 8 and 52 meters. Further stationary experiments were presented in 1983 (ref. 2). After that time, implementation issues were emphasized, such as the development of fast initialization algorithms (ref. 3). Realtime low dynamic shipboard tests were performed in 1988 (ref. 4), with reported angular accuracies on the order of several degrees for a 60-centimeter baseline. The first non-realtime determination of aircraft pitch and heading using one GPS interferometer was reported in 1989 (ref. 5), revealing pitch and heading uncertainties below 1 milliradian for a 23-meter baseline. Today, sub-milliradian accuracies,

even for 1-meter baselines, are attainable for rigid platforms using off-the-shelf equipment (ref. 6). In this paper, we present methods and results for realtime pitch, roll, and heading determination of a dynamic, flexible platform.

PRINCIPLES OF ATTITUDE DETERMINATION

In the formalism of rigid body kinematics, attitude is defined by the rotation transformation which relates a coordinate system fixed in space to a coordinate system fixed in the body (ref. 7). For the purposes of aircraft navigation, the space coordinate system typically is defined to be locally level. The positive x-axis is pointed towards north, the positive y-axis is pointed east and the positive z-axis is pointed down. Finally, since the transformation is concerned with rotation only (and not translation), the origin of the space coordinate system is coincident with the aircraft center of gravity. Similarly, the body (or aircraft) coordinate system also has its origin at the aircraft center of gravity. However, for the aircraft coordinate system, the positive x-axis is directed along the fuselage axis towards the nose, the positive y-axis points along the right wing (right according to the pilot), and the positive z-axis is directed towards the floor of the aircraft (ref. 8).

The two coordinate systems are related through the antenna locations. Under the rigid body assumption, the antenna locations remain fixed in the aircraft coordinate system. Their locations in this system are determined prior to the flight through standard GPS static surveying techniques. While in flight, the antenna locations in the space coordinate system are determined using GPS interferometry. Thus at any given epoch, the antenna locations in the two coordinate systems are related as follows:

$$[R]_s=[T][R]_a \tag{1}$$

where: $[R]_s$ is the matrix containing the locations of the antennas in space coordinates. $[R]_a$ is the matrix containing the locations of the antennas in aircraft coordinates. [T] is the rotation transformation matrix.

The transformation is a 3x3 matrix containing the nine direction cosine elements. It is the product of three rotation matrices each of which specifies a rotation about a certain aircraft coordinate system axis. Thus,

$$[T]=[\phi][\theta][\psi] \tag{2}$$

where: $[\phi]$ represents a rotation about the fuselage axis (roll); $[\theta]$ represents a rotation about the wing axis (pitch); and $[\psi]$ represents a rotation about the vertical axis (heading).

The transformation matrix may now be expressed in terms of the three rotations (also known as the Euler angles):

$$[T]=\begin{pmatrix} \cos\psi\cos\theta & -\sin\psi\cos\phi+\cos\psi\sin\theta\sin\phi & \sin\psi\sin\phi+\cos\psi\sin\theta\cos\phi \\ \sin\psi\cos\theta & \cos\psi\cos\phi+\sin\psi\sin\theta\sin\phi & -\cos\psi\sin\phi+\sin\psi\sin\theta\cos\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{pmatrix} \tag{3}$$

Determination of the instantaneous values of the Euler angles proceeds as follows. First, the matrix containing the antenna locations in aircraft coordinates is formed prior to the start of attitude determination. Secondly, realtime GPS interferometric measurements are used to form the matrix containing the antenna locations in the space coordinate system. Solving for the transformation matrix involves postmultiplying the space coordinate matrix by the generalized inverse of the aircraft coordinate matrix:

$$[T]=[R]_s[R]_a^T\left([R]_a[R]_a^T\right)^{-1} \tag{4}$$

Finally, the Euler angles can be calculated as follows:

$$\theta=\sin^{-1}(-T_{31});\ \phi=\sin^{-1}\left(\frac{T_{32}}{\cos\theta}\right);\ \psi=\sin^{-1}\left(\frac{T_{21}}{\cos\theta}\right) \tag{5}$$

The calculations resulting in this Euler angle triad have been presented to provide insight into the attitude determination process. However, although this triad is probably the best understood means of attitude and heading expression, it is far from robust and arguably should not be considered for use in a realtime integrated navigation system (ref. 9). This lack of robustness can be seen most easily by examining equations (3) and (5) for the case of $\theta = \pm 90°$. Note that although θ can be recovered, ϕ and ψ are lost. As a result, the Euler angle triad is not an all-attitude parameterization. Although many parameterizations of [T] are possible, quaternions offer conciseness, efficiency, and stability not possible with other parameterizations (refs. 8-9).

Further practical considerations involve the aircraft structure. The derivation given above rests on the assumption that the antenna locations are fixed in the aircraft coordinate system. Error mechanisms are thereby introduced by the non-rigid body behavior of the aircraft. The fuselage by itself is a good rigid body approximation, but inclusion of the wings (with their normal flexing) make the overall aircraft far from being so. An additional consideration is more subtle. The derivation assumes that the location of the aircraft center of gravity is known relative to the antenna positions. Although a nominal location can be used, the true location will change over the course of the flight.

GPS INTERFEROMETRY

Consider a simplified GPS interferometer, as depicted in figure 1, consisting of two GPS antennas and the necessary electronics to determine the instantaneous phase of a received GPS carrier signal. Usually, the distance between the antennas and a GPS satellite is much

larger than the baseline length between the two antennas, such that the GPS carrier signal can be treated as a plane wave. Under the assumption that the satellite clock and the receiver clock are synchronized, the phase difference $\Delta\phi$ between simultaneous carrier phase measurements at the two antennas can be written as

$$\Delta\phi = \phi_1 - \phi_2 = \underline{b} \cdot \underline{\hat{e}} + N\lambda \qquad (6)$$

where ϕ_1 and ϕ_2 are the measured carrier phases in wavelengths for antennas 1 and 2, respectively; $\underline{b}$ is the three-dimensional baseline vector pointing from antenna 1 to antenna 2; $\underline{\hat{e}}$ is the unit vector pointing from the interferometer to the satellite; and N is an integer number accounting for the ambiguity due to the limited length of the carrier wavelength λ. The phase difference between antennas is also referred to as a Single Difference (SD) (ref. 10). The determination of the unit vector to the satellite requires the approximate coordinates of the interferometer and the satellite coordinates. The latter are calculated from the Navigation Data transmitted by the satellites. A standard GPS navigation solution is used to determine the interferometer coordinates. The baseline solution is not very sensitive to the interferometer coordinates, as was shown by Jurgens (ref. 11), because of the large distance to the satellite with respect to the relatively short baseline length of the interferometer.

If the satellite clock and the receiver clock are not synchronized, each of the phase measurements, ϕ_1 and ϕ_2, contain an unknown satellite clock phase offset and a receiver clock phase offset. If the phase measurements are made simultaneously, then the phase difference $\Delta\phi$ does not contain the unknown satellite clock phase offset. Actually, since the time of transmission for each of the phases transmitted by the satellite will be slightly different by an amount $\Delta t=(\Delta\phi\ \lambda)/c$, a small offset will be present in the phase difference due to the change in the satellite clock offset during the interval Δt. This offset can be corrected using the satellite clock coefficients transmitted by the satellite in the Navigation Data.

The unknown receiver clock offset can be eliminated by taking the difference between two independent Single Differences, also referred to as a Double Difference (DD)

$$DD = \Delta\phi_1 - \Delta\phi_2 = \underline{b} \cdot (\underline{\hat{e}}_1 - \underline{\hat{e}}_2) + (N_1 - N_2)\lambda \qquad (7)$$

Thus a Double Difference is a linear combination of four phase measurements obtained from two satellites at two antennas, as shown in figure 2. This can be extended to eliminate the unknown integer ambiguity number by differencing two Double Differences at two epochs to obtain a Triple Difference. The Triple Difference is useful for static positioning but only provides the rate of change of the Double Difference for dynamic applications. As such it does not provide an accurate instantaneous solution for the baseline vector.

Assuming that the integer ambiguities are known, then using equation (6), it follows that three independent Double Differences are required to solve for the baseline vector $\underline{b} = (b_1, b_2, b_3)^T$:

$$\begin{pmatrix} DD_1 \\ DD_2 \\ DD_3 \end{pmatrix} = \begin{pmatrix} (\underline{e}_1 - \underline{e}_2)^T \\ (\underline{e}_1 - \underline{e}_3)^T \\ (\underline{e}_1 - \underline{e}_4)^T \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{8}$$

Furthermore, if the norm of the baseline vector, $|\underline{b}|$, is known *a priori*, then only two Double Differences are needed:

$$\begin{pmatrix} DD_1 \\ DD_2 \\ |\underline{b}| \end{pmatrix} = \begin{pmatrix} (\underline{e}_1 - \underline{e}_2)^T \\ (\underline{e}_1 - \underline{e}_3)^T \\ b_1 \quad b_2 \quad b_3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{9}$$

Several techniques have been proposed to quickly determine the integer ambiguity for each of the Double Differences (refs. 12-13). In general, search methods are very effective to quickly resolve the integer ambiguities whenever redundant measurements are available. The basic idea is to evaluate the measurement residuals for potential baseline vector solutions. The techniques used are very similar to those applied to GPS fault detection (see ref. 14 for a summary of these techniques). If, for instance, signals from five satellites are received by one GPS interferometer, four independent Double Differences can be formed. Equation (8) is rewritten as:

$$\underline{DD} = H\underline{b} \tag{10}$$

where $\underline{DD}$ is a 4-by-1 Double Difference measurement vector and H is a 4-by-3 data matrix containing the satellite unit vector differences. The residuals are the differences between the actual Double Difference measurements and those predicted based on the least squares solution for $\underline{b}$, the approximate interferometer coordinates, and the known satellite coordinates. The least squares residual vector, $\underline{e}_r$, is obtained by projecting $\underline{DD}$ onto a space perpendicular to the estimation space spanned by the columns of H. Thus, $\underline{e}_r$ is orthogonal to the estimation space. An efficient method to obtain $\underline{e}_r$ is by decomposing the H matrix into the product of a real orthonormal matrix Q and an upper triangular matrix R using the QR factorization (refs. 15-16). Replacing H in equation (10) by the product QR and pre-multiplying both sides by Q^T yields ($Q^TQ = I$)

$$R\ \underline{b} = Q^T\ \underline{DD} \tag{11}$$

The rank of R is equal to the rank of H; therefore, the bottom row of R consists of zeros only. Equation (11) can be divided into two equations:

$$U\,\underline{b} = Q_1\,\underline{DD} \quad or \quad \underline{b} = U^{-1}\,Q_1\,\underline{DD}$$

$$\underline{0} = \underline{q}\,\underline{DD} \tag{12}$$

where U consists of the first three rows of R, Q_1 consists of the first three rows of Q^T, and $\underline{q}$ is the transpose of the fourth row of Q^T. The top expression of equation (12) relates the measurements to the baseline coordinates, forming the least squares solution. The vector $\underline{q}$ and the Double Difference vector $\underline{DD}$ are othogonal. The latter property is used to obtain the norm of the residual vector if measurement errors are present on the Double Differences:

$$|\underline{e}_r| = (q_1\ q_2\ q_3\ q_4)\begin{pmatrix} DD_1 \\ DD_2 \\ DD_3 \\ DD_4 \end{pmatrix} \tag{13}$$

The vector $\underline{q}$ has to be calculated only once for a search at one particular epoch. Thus, to calculate the residual for a potential solution, only four multiplications and three additions are required.

Some prior knowledge is required about the interferometer in order to reduce the potential number of solutions for $\underline{b}$. Several methods can be used, including:

- Limit the baseline length of the interferometer to, for instance, one meter such that the integer ambiguity number must lie in the interval [-5,5](one GPS wavelength, λ, is approximately 0.19 meters);
- Use *a priori* information about the orientation of the interferometer such that the search can be limited to correspond to the approximate baseline orientation. Approximate attitude and heading information could for instance be obtained from a low cost three-axis accelerometer and magnetometer;
- Use Double Differences calculated from the GPS pseudorange measurements (code phase Double Differences) to obtain approximate values for the carrier phase Double Differences.

If the Double Difference uncertainty can be limited to $\pm 5\lambda$, then even an exhaustive search would only require the calculation of $(11)^4 = 14{,}641$ residuals, which would require seven times that many or 102,487 floating point calculations. This can easily be accomplished within one second using a microcomputer. In this paper, we estimate the initial DD ambiguities by filtering the relatively noisy code phase Double Differences with the carrier phase Double Differences using a complementary Kalman filter (see also ref. 17). In the complementary filter, the carrier phase DDs are used to provide a reference for the code phase DDs. The performance of the complementary filter is very similar to the Hatch filter

(ref. 18), but has much lower computational and storage requirements. Although only one DD is estimated, the complementary filter model can easily be extended to include the rate of change of the DD as well. The equations for the (scalar) complementary Kalman filter are summarized below

$$\hat{x}_k^- = \hat{x}_{k-1}^+ + (DD_k - DD_{k-1})$$

$$p_k^- = p_{k-1}^+ + q$$

$$k_k = p_k^- (p_k^- + r)^{-1} \qquad (14)$$

$$\hat{x}_k^+ = \hat{x}_k^- + k_k(z_k - \hat{x}_k^-)$$

$$p_k^+ = (1 - k_k)p_k^-$$

where $\hat{x}$ denotes the estimated code phase DD with variance q; r is the variance of the code phase DD measurement z; and p is the variance of the estimation error. The first two lines in equation (14) propagate the estimated code phase DD and the estimation error variance from time (k-1) to time k, followed by the calculation of the Kalman gain (line 3), and the update of the estimated code phase DD and the estimation error variance (lines 4 and 5). Figure 3 shows an example of both code phase DDs and carrier phase DDs for a 20-minute section of a flight test. The DDs were obtained from a GPS interferometer on the fuselage of a DC-3 research aircraft during roll maneuvers. The next section provides a detailed description of the aircraft test bed. Note that the code phase DDs and the carrier phase DDs are separated by an unknown integer wavelength ambiguity. Figure 4 shows both the carrier phase DDs and the filtered code phase DDs. As expected, the filtered code phase DDs follow the dynamics very well. Finally, the difference between the carrier phase DDs and the filtered code phase DDs are shown in figure 5. After the filter is initialized with the first code phase DD, it takes typically less than 20 seconds for the carrier phase DD to be within $\pm 5\lambda$ with respect to the filtered code phase DD. The remaining differences between the filtered code phase DDs and the carrier phase DDs are due to the code phase multipath from the aircraft structure.

The filtered code phase DD is used to calculate an initial estimate for the absolute carrier phase DD. Next, an exhaustive search is performed to find those combinations of integer ambiguities that will result in low residuals. Depending on the carrier phase measurement noise and the satellite measurement geometry, more than one potential set of ambiguities may be found. Further eliminations of potential sets are achieved by applying known baseline length constraints. For an interferometer mounted on a semi-rigid platform, the baseline length can be determined to within a few millimeters. The knowledge of the norm of the baseline vector is limited by differential antenna phase center migrations, differential multipath and structural flexing of the platform. Next, additional redundant measurement data can be used to resolve the integer ambiguities, again by examining the residuals. Although statistically unlikely, it would still be possible to select the wrong set of

ambiguities; however, by continuing to examine all measurement residuals throughout the experiment, the resulting inconsistencies in the data would be quickly detected.

Finally, after the integer ambiguities have been determined for each of the Double Differences, they can be used to maintain aircraft attitude and heading for long periods of time until loss-of-lock occurs. It is also important to add that once the integer ambiguities are resolved, redundant measurements are no longer necessary to maintain attitude and heading. Assuming known baseline lengths, a rigid platform would only require continuous lock on the signals from three satellites at three antennas. In this case, to find the approximate interferometer coordinates, another sensor input, such as the baro-altimeter or a stable frequency reference source, is required. If the baselines are subjected to structural flexing, continuous lock on four satellites at three antennas is required.

DESCRIPTION OF THE TESTBED

The realtime attitude and heading determination testbed consists of three main components: 1) Aircraft, antennas, and cabling; 2) GPS receiver; 3) Data storage, computation, and display units, as depicted in figure 6. Four microstrip antennas have been installed on Ohio University's Douglas DC-3 research aircraft, N7AP. One antenna is located on each wingtip and two antennas are located on top of the fuselage with one near the nose and the other towards the tail (fig. 7). RG-400 cable is used because of its high phase stability and small physical diameter. Phase stability is essential since each of the four cables is approximately 25 meters in length. The cables feed the signal from the antenna preamplifiers to the GPS receiver located in a standard equipment rack installed in the cabin midway between the nose and tail of the aircraft.

The GPS receiver being used is the Ashtech Three-dimensional Direction Finding (3DF) 24-channel receiver (ref. 6). The 3DF is configured as four 6-channel sections with each section receiving signals from a separate antenna. The receiver is packaged in a ruggedized 3-5/8" X 8-1/2" X 8" housing and requires a 12 watt power source. Realtime measurement data at a selectable rate of up to once per second (both raw and processed) are available over two serial ports at data rates up to 115,200 baud.

The storage, computation and display units consist of two microcomputers. A 25 MHz 80386-based computer receives realtime data from the 3DF, performs storage, computation and sends realtime attitude information to a second computer. The second computer is a 12 MHz 80286-based PC. It generates an artificial horizon and heading indicator and drives it with the attitude information supplied by the 80386 machine.

Baselines between the antennas were determined using standard GPS static surveying techniques. The lengths of these baselines are given in Table 1. These measurements have also been verified using a steel tape measure and were found to agree to within 5 millimeters.

Baseline	Length in meters
ant 1 - ant 2	7.781
ant 1 - ant 3	14.468
ant 1 - ant 4	14.466
ant 2 - ant 3	14.941
ant 2 - ant 4	14.924
ant 3 - ant 4	28.194

Table 1. Antenna baseline lengths for Douglas DC-3 test bed.

GROUND and FLIGHT TESTING

Ground and flight testing of the GPS attitude and heading determination system started in April of 1991. To date, three demonstration flights have been conducted (April 18, May 8 and May 22). Each flight test lasted for approximately 1.5 to 2 hours, and was preceded by a data collection session on the ramp at Ohio University's airport (UNI) in Albany, Ohio. During the first two flights only raw measurement data were collected which were used for the continued development of realtime attitude and heading determination algorithms. During the third flight, long periods of realtime attitude and heading determination were achieved. Moreover, the 3DF GPS receiver maintained continuous lock on at least 3 satellites at 3 antennas during all maneuvers. This is an important result, because it allows for continuous attitude and heading determination. We also found that whenever the receiver would loose lock on a particular satellite due to temporary shielding by the airframe, the receiver would typically re-acquire the signal within a few seconds.

Aircraft flexing was found to be the main reason for discontinuities in the attitude and heading data. The carrier phase DD residuals would increase which caused the algorithm to re-initiate the integer ambiguity search. This could be avoided by allowing larger DD residuals in combination with the use of a higher baseline length uncertainty.
A representative section of the ramp test results collected from the 3DF receiver on May 22, 1991, is shown in figures 8 through 10. It should be noted that the aircraft was not static during this data collection session; small changes in pitch, roll and heading angles were introduced by wind gusts and by the flight crew moving around inside the aircraft. In addition, some multipath effects can be observed in the data. During this session, five healthy GPS satellites were visible (satellites 2, 6, 13, 14 and 15). The Position Dilution of Precision (PDOP) was approximately equal to 3 throughout the test. The noise level on the roll and heading angles is on the order of 0.1 mrad (0.006 degrees). Pitch angle noise is approximately three time larger because of the shorter baseline used for the pitch angle determination. Absolute accuracy results are not available due to the lack of a reference

for the attitude data which must be aligned with the phase centers of the GPS antennas. However, previous comparisons between the 3DF and kinematic GPS results were found to agree to within 0.1 mrad (ref. 6).

Realtime data collected from the 3DF receiver during two sections of the third flight test conducted in the evening of May 22, 1991 are presented next. The noise levels on the pitch, roll and heading angles are difficult to analyze directly, because of the many attitude and heading variations with relatively large magnitudes due to normal flight. Instead, the emphasis of the results presented in this section is on the capability of GPS interferometry to provide realtime, in flight attitude and heading information. Figure 11 shows the ground track of a 27-minute section of the third flight which was conducted in the vicinity of Ohio University. Figures 12 through 14 show the aircraft heading, roll and pitch angles as a function of time. Note that the noise-like behaviour of the angles are all due to actual aircraft dynamics (the period of the angle variations is on the order of 10 to 20 seconds). Approximately 12 minutes after the start of the run, attitude and heading information was lost for approximately 30 seconds. As discussed earlier, this was caused by poor DD residuals which caused the attitude and heading algorithm to re-initiate the integer ambiguity search. Shown in figure 15 is the lock history of the 3DF receiver; even in the presence of signal shielding due to airplane maneuvers, the receiver maintained lock on a minimum of four satellites at 3 antennas.

Figures 16 through 20 show the results for a 13-minute section of the third flight test during which a series of 10 and 30-degree roll maneuvers were performed. Continuous attitude and heading information was available using four or five GPS satellites. During the 30-degree rolls, only one of the five visible satellites was not observed continually by at least three antennas.

SUMMARY, DEVELOPMENTS AND APPLICATIONS

The feasibility of using realtime GPS interferometric attitude and heading determination for dynamic semi-rigid platforms has been demonstrated. During the course of this research, several issues have been identified which require further study. Among these are continued development of robust fault detection and isolation algorithms for the detection and correction of cycle slips, characterization and mitigation of multipath, and integration with an inertial measurement unit.

GPS interferometry has numerous applications including: reliable aircraft/spacecraft state estimation; measurement of structural flexing of large space structures; synthetic aperture radar (ref. 5); aerial photogrammetry; pointing systems; and rapid alignment of inertial navigation systems for submarines and aircraft while maneuvering.

ACKNOWLEDGEMENTS

This work was supported by the FAA/NASA Joint University Program for Air Transportation Research (Grant NGR-36-009-17). The 3DF GPS receiver was provided by Ashtech, Inc. The authors are indebted to Dr. Javad Ashjaee and Mr. Mark Kuhl for their support, including many technical discussions on the topic of GPS interferometry, and their help in processing the raw measurement data. The authors would also like to thank the DC-3 flight crew, Dr. Richard H. McFarland, Mr. Dennis J. Atkeson and Mr. Richard Zoulek, for their flexibility in dealing with the GPS satellite visibility windows.

REFERENCES

1. Greenspan, R. L., Ng, A. Y., Przyjemski, J. M., and Veale, J. D.: Accuracy of Relative Positioning by Interferometry with Reconstructed Carrier GPS: Experimental Results. 3rd International Geodetic Symposium on Satellite Doppler Positioning, Las Cruces, NM, Feb. 8-12, 1982.

2. Burgett, W. S., Roemerman, S. D., and Ward, P. W.: The Development and Applications of GPS-Determined Attitude. Presented to the National Telesystems Conference, November 1983.

3. Remondi, B. W.: Performing Centimeter-Level Surveys in Seconds with GPS Carrier Phase: Initial Results. Navigation: Journal of the Institute of Navigation, Vol. 32, No. 4, Winter 1985-86.

4. Kruczynski, L. R., Li, P. C., Evans, A. G., and Hermann, B. R.: Using GPS to Determine Vehicle Attitude: USS Yorktown Test Results. Proceedings of the 2nd International Technical Meeting of the Satellite Division of the Institute of Navigation, Colorado Springs, CO, Sep. 27-29, 1989.

5. Purcell, Jr., G. H., Srinivasan, J. M., Young, L. E., DiNardo, S. J., Hushbeck, Jr., E. L., Meehan, T. K., Munson, T. N. and Yunck, T. P.: Measurement of Aircraft Position, Velocity, and Attitude using Rogue GPS Receivers. Proceedings of the 5th International Geodetic Symposium on Satellite Positioning, Las Cruces, NM, March 1989.

6. Ferguson, K., Kosmalska, J., Kuhl, M., Eichner, J. M., Kepski, K, and Abtahi, R.: Three-Dimensional Attitude Determination with the Ashtech 3DF 24-Channel GPS Measurement System. Proceedings of the National Technical Meeting of the Institute of Navigation, Phoenix, AZ, Jan. 1991.

7. Goldstein, H.: Classical Mechanics, Second edition. Addison-Wesley, Reading, Massachusetts, 1980.

8. Farrell, J. L.: Integrated Aircraft Navigation. Academic Press, San Diego, 1976.

9. Farrell, J. L. and Van Graas, F.: That All-Important Interface. Proceedings of the Third International Technical Meeting of the Satellite Division of the Institute of Navigation, Colorado Springs, CO, September 1990.

10. Remondi, B. W.: Using the Global Positioning System (GPS) phase observable for relative geodesy: modeling, processing, and results. Ph.D. Dissertation, The University of Texas at Austin, Austin, Texas, 1984.

11. Jurgens, R. D.: Realtime GPS Azimuth Determining System. Proceedings of the National Technical Meeting of The Institute of Navigation, San Diego, CA, Jan. 23-25, 1990.

12. Hatch, R.: Ambiguity Resolution in the Fast Lane. Proceedings of the 2nd International Technical Meeting of the Satellite Division of The Institute of Navigation, Colorado Springs, CO, Sep. 27-29, 1989.

13. Hwang, P. Y. C.: Kinematic GPS: Resolving Integer Ambiguities on the Fly. Proceedings of the IEEE Position, Location, and Navigation Symposium, Las Vegas, NV, March 21-23, 1990.

14. Brown, R. G., Chin, G. Y., and Kraemer, J. H.: RAIM: Will it Meet the RTCA GPS Minimum Operational Performance Standards? Proceedings of the National Technical Meeting of The Institute of Navigation, Phoenix, AZ, Jan. 22-24, 1991.

15. Brenner, M.: Implementation of a RAIM Monitor in a GPS Receiver and an Integrated GPS/IRS. Proceedings of the 3rd International Technical Meeting of the Satellite Division of The Institute of Navigation, Colorado Springs, CO, Sept. 1990.

16. Golub, G. H. and Van Loan, C. F.: Matrix Computations, Second Edition. The Johns Hopkins University Press, Baltimore, MD, 1989.

17. Hwang, P. Y. C. and Brown, R. G.: GPS Navigation: Combining Pseudorange with Continuous Carrier Phase using a Kalman Filter. Proceedings of the 2nd International Technical Meeting of the Satellite Division of The Institute of Navigation, Colorado Springs, CO, Sep. 27-29, 1989.

18. Hatch, R.: The Synergism of GPS Code and Carrier Measurements. Proceedings of the 3rd International Geodetic Symposium on Satellite Doppler Positioning, Las Cruces, NM, Feb. 1982.

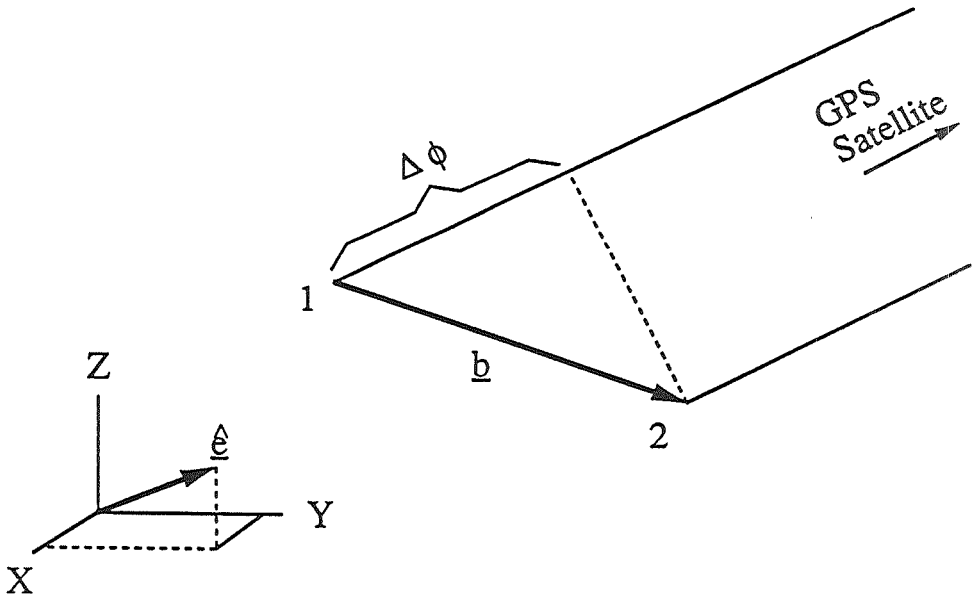


Figure 1. Simplified GPS Interferometer.

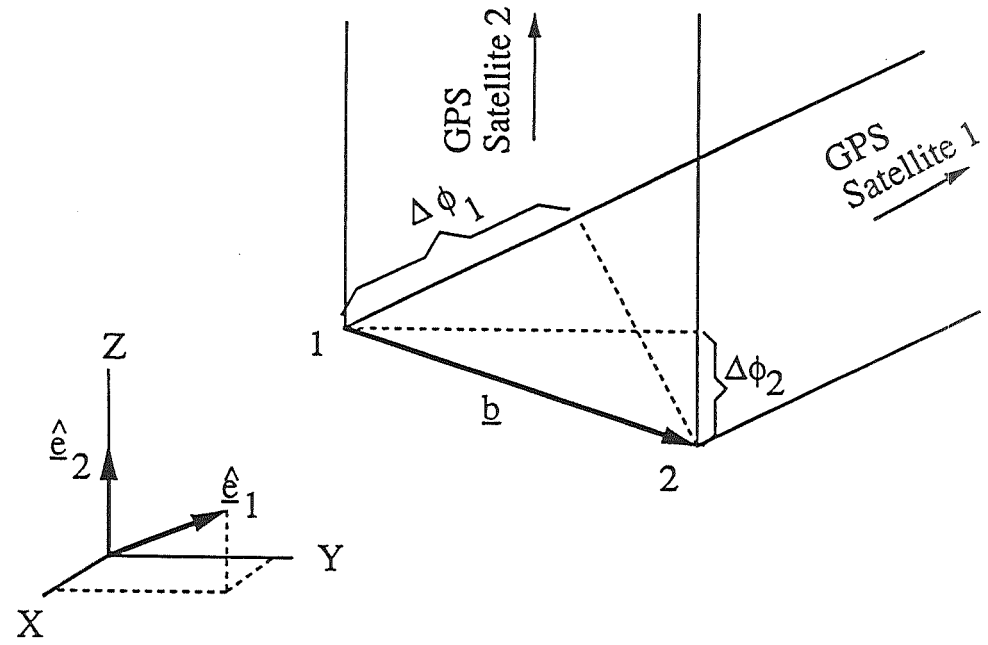


Figure 2. The Double Difference Observation.

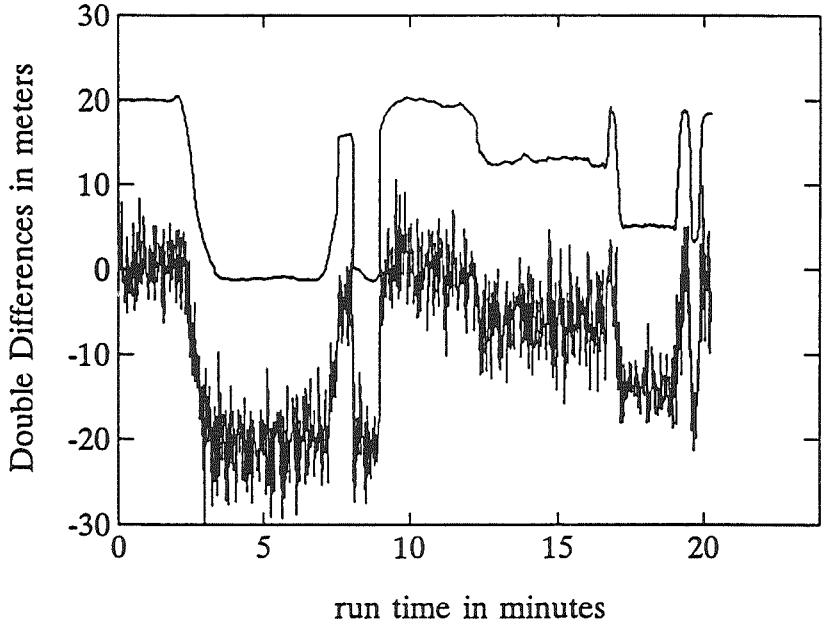


Figure 3. Code phase (bottom) and carrier phase (top) Double Differences.

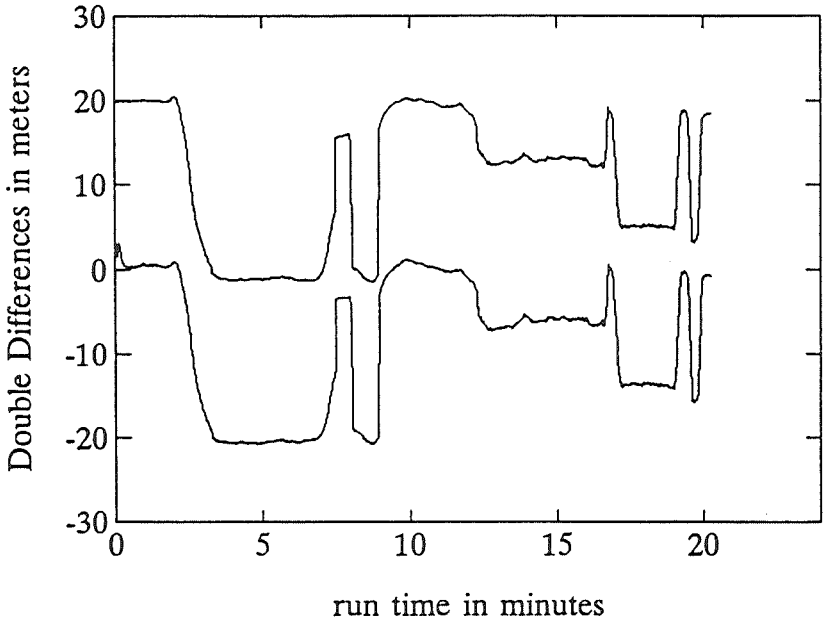


Figure 4. Carrier phase Double Differences (top) and filtered code phase Double Differences (bottom).

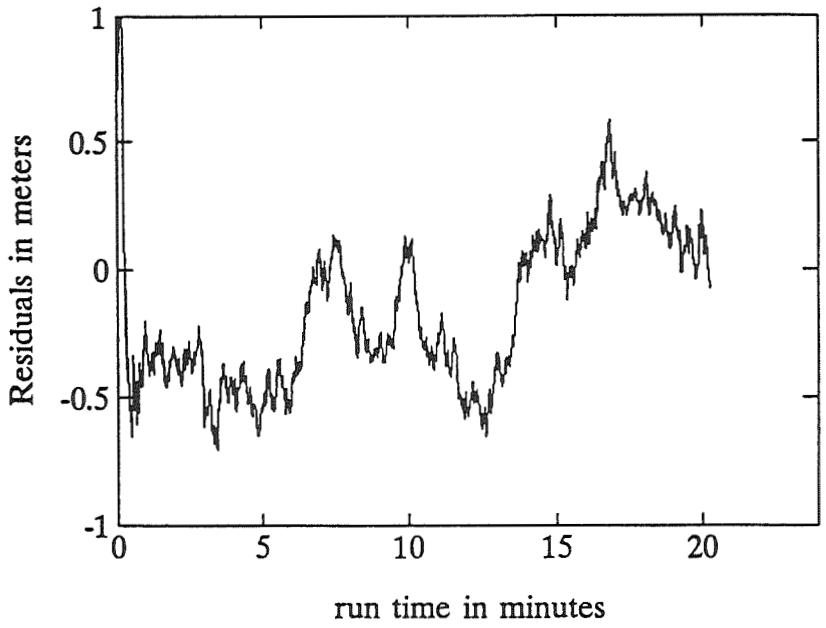


Figure 5. Difference between carrier phase Double Differences and filtered code phase Double Differences.

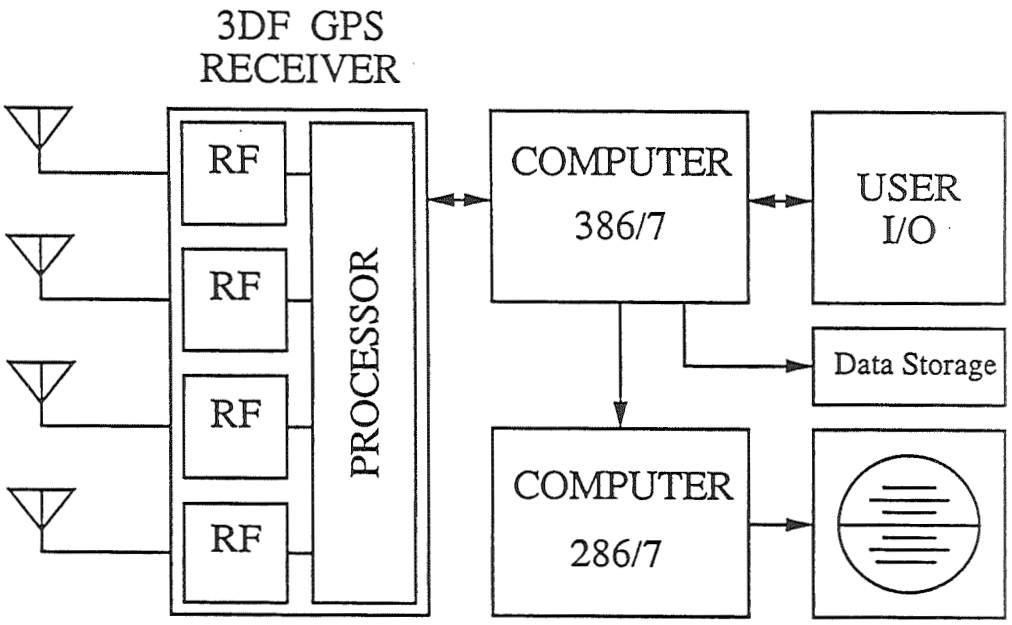


Figure 6. Block diagram of realtime attitude and heading determination test bed.

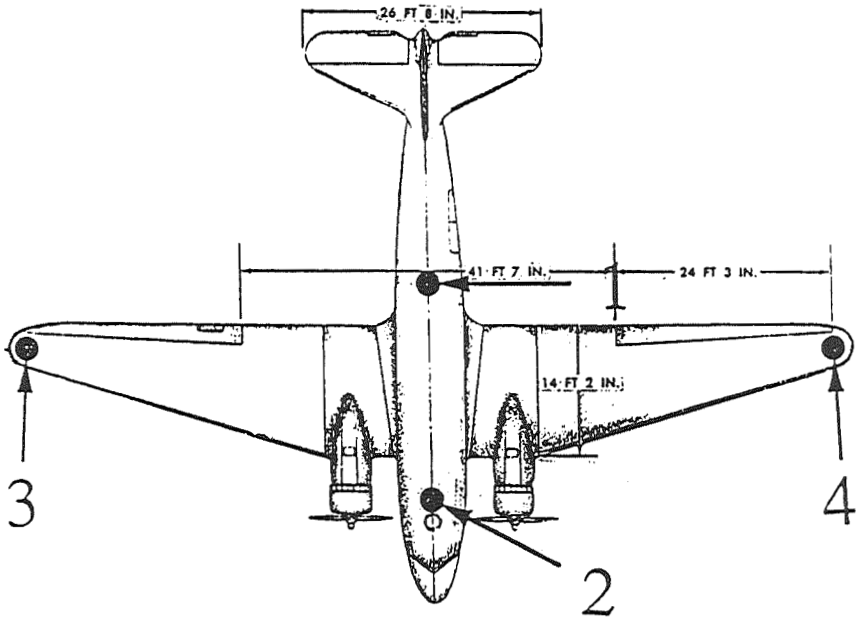


Figure 7. Douglas DC-3 GPS antenna locations.

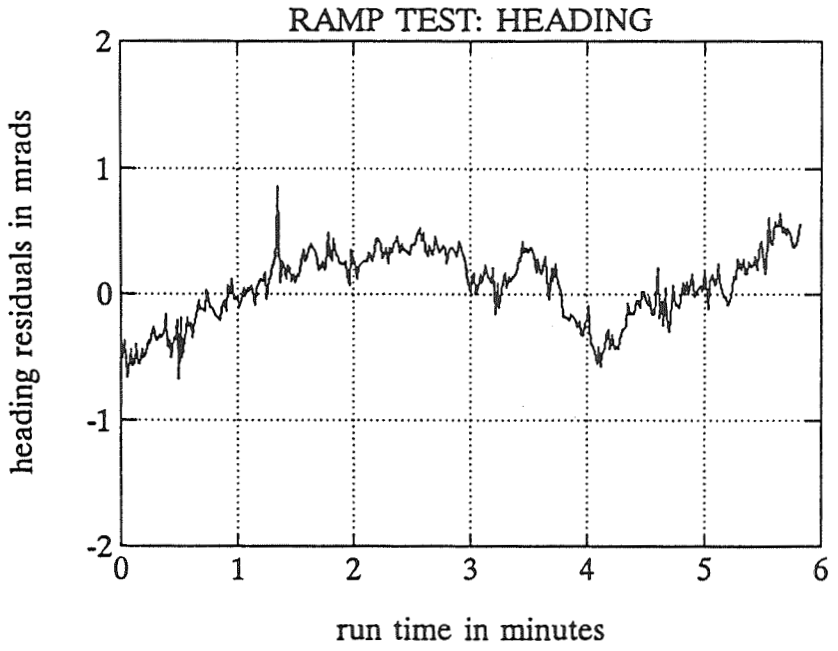


Figure 8. Ramp test heading residuals.

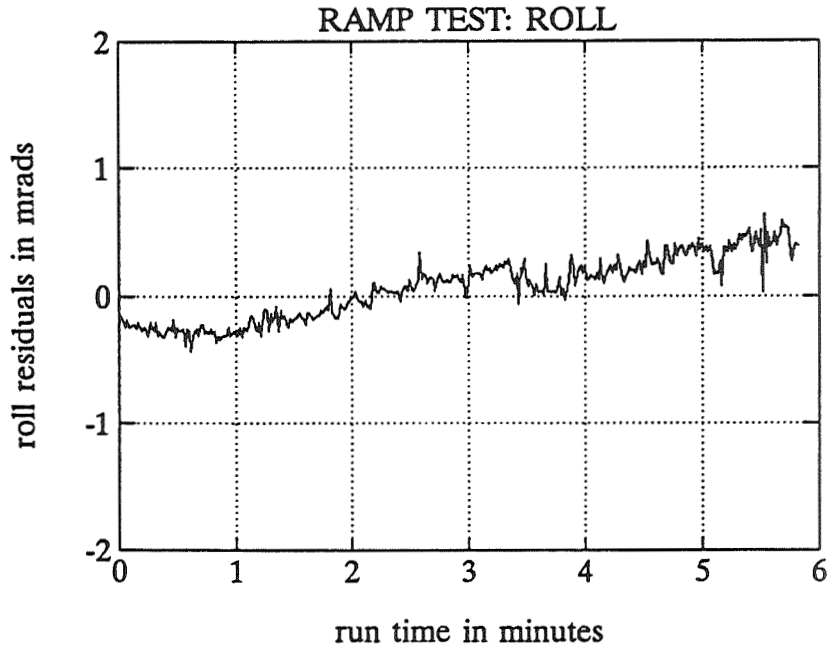


Figure 9. Ramp test roll residuals.

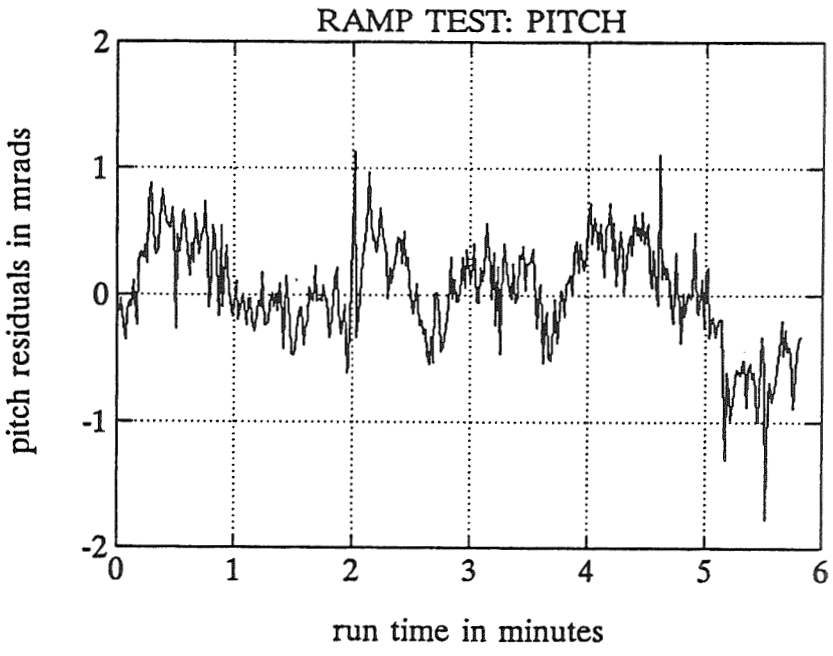


Figure 10. Ramp test pitch residuals.

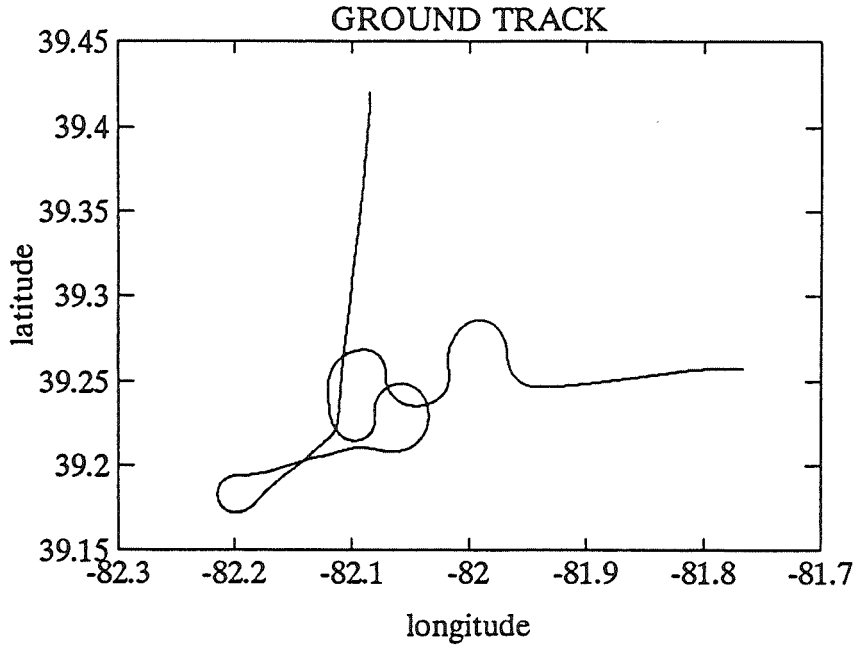


Figure 11. Ground track for a 27-minute section of the third flight test.

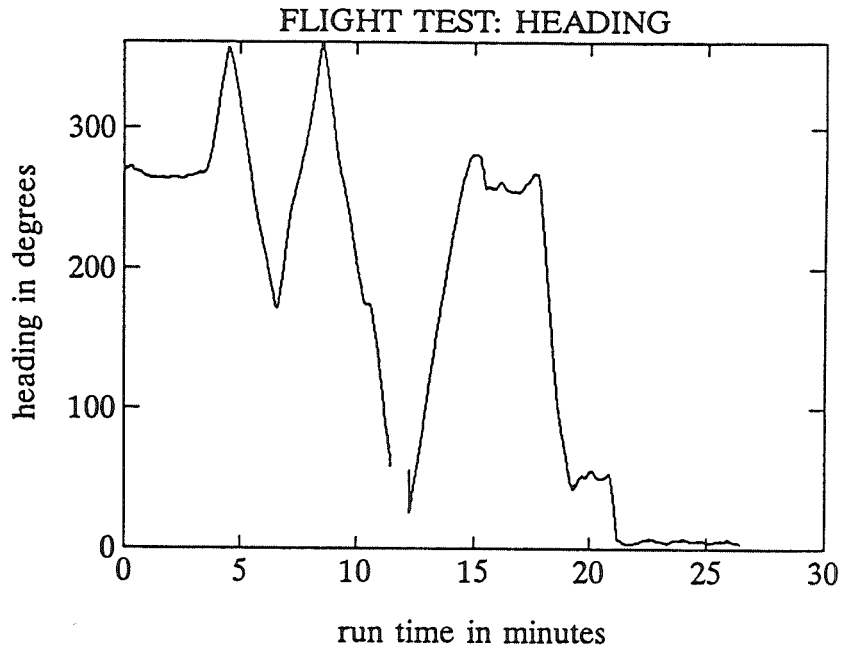


Figure 12. Flight test heading as a function of time.

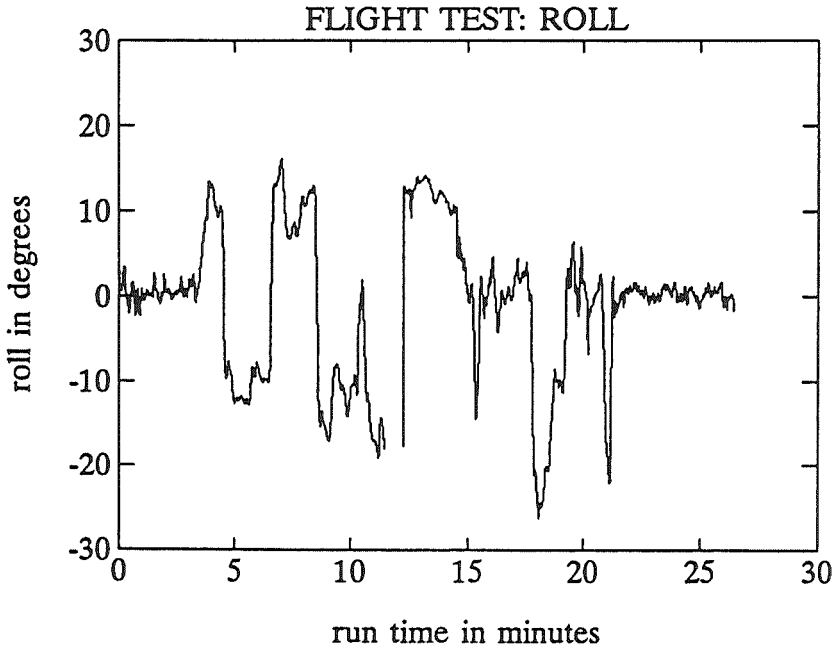


Figure 13. Flight test roll as a function of time.

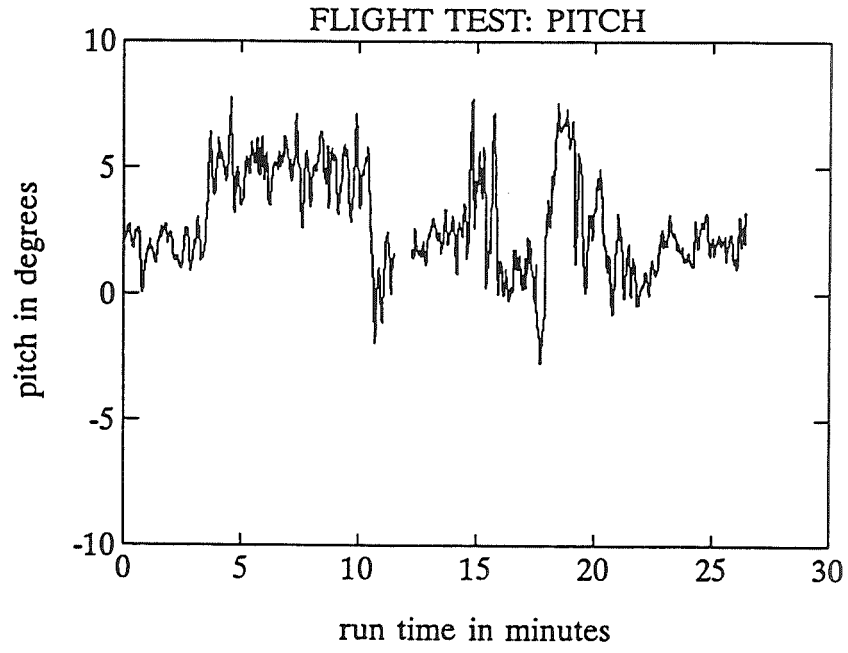


Figure 14. Flight test pitch as a function of time.

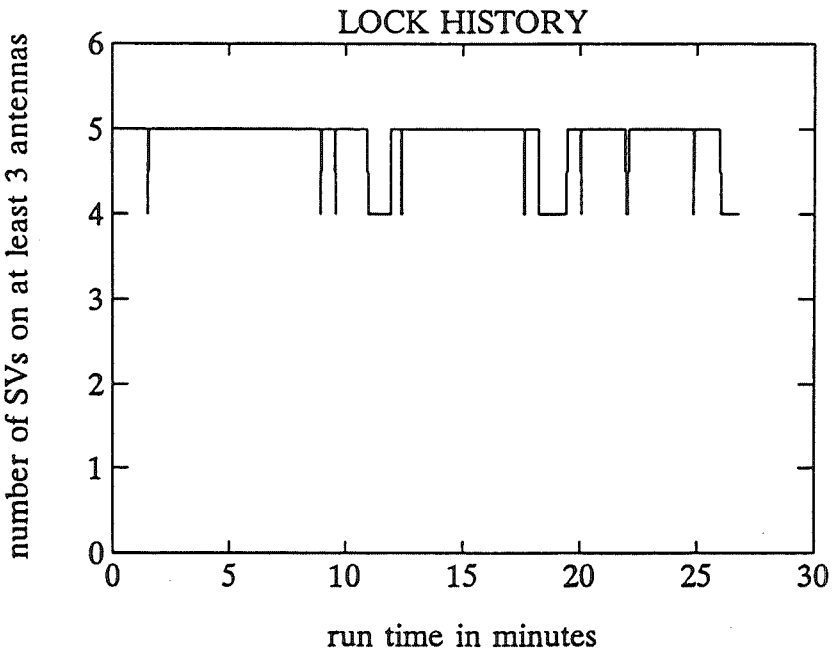


Figure 15. Flight test lock history.

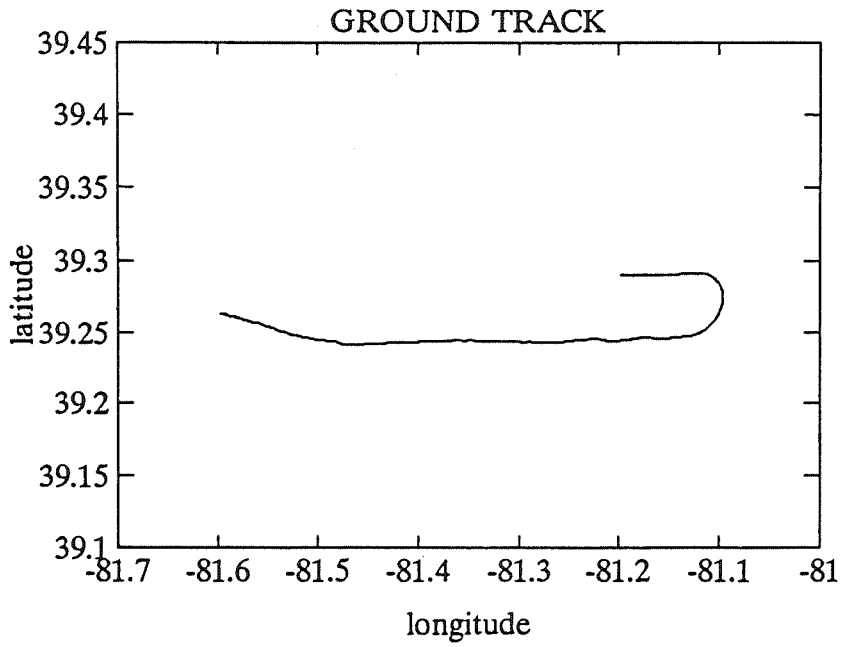


Figure 16. Ground track for a 13-minute section of the third flight test.

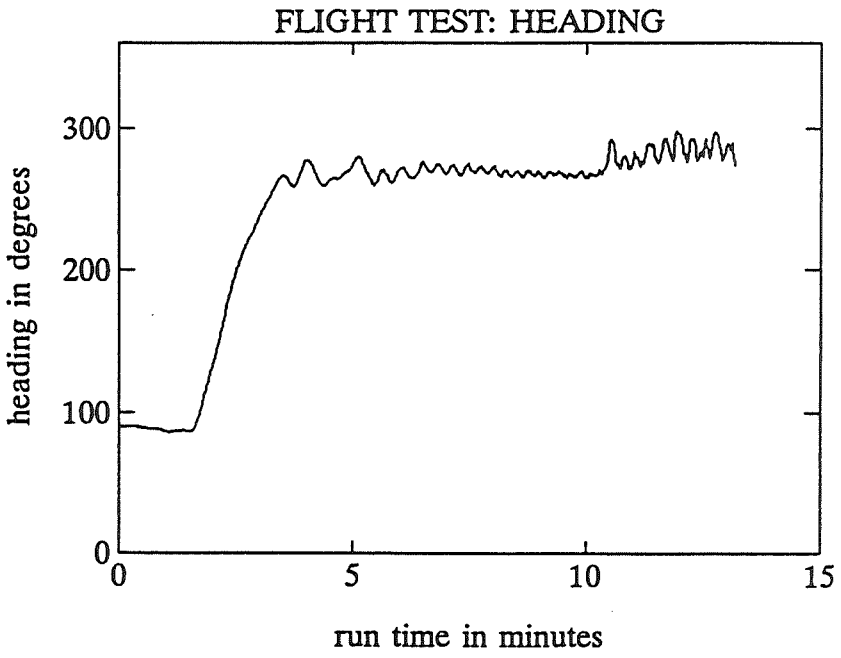


Figure 17. Flight test heading as a function of time.

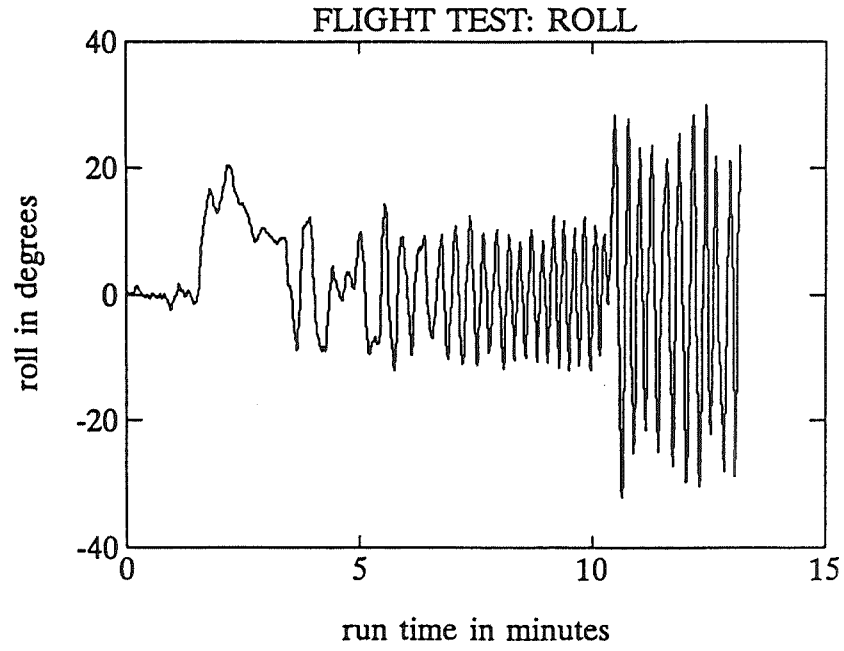


Figure 18. Flight test roll as a function of time.

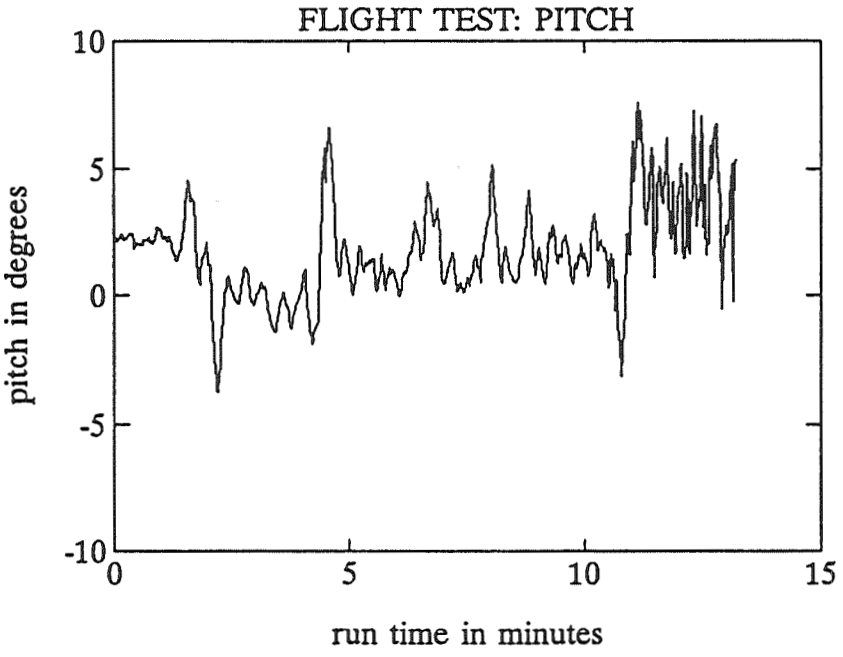


Figure 19. Flight test pitch as a function of time.

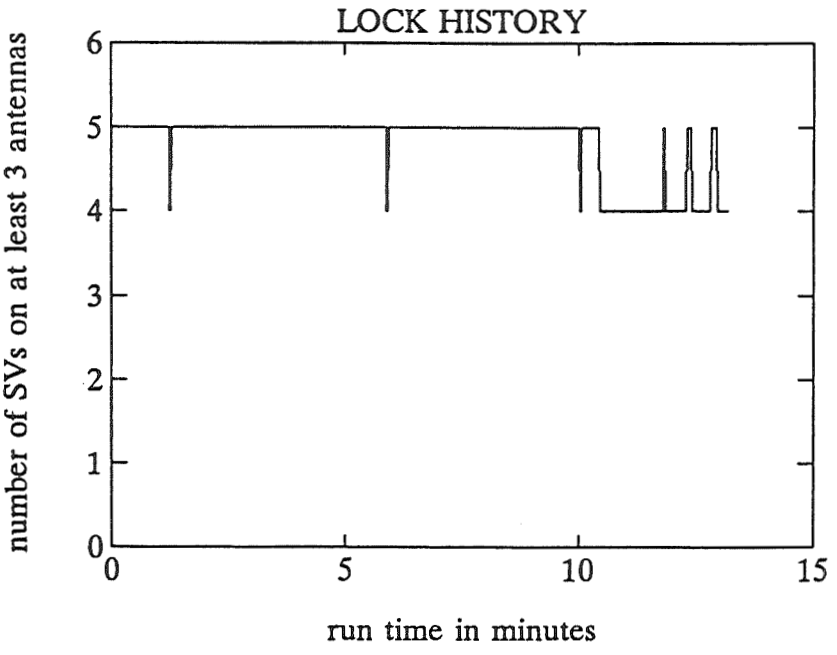


Figure 20. Flight test lock history.

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1990-91

Robert F. Stengel
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey 08544

SUMMARY OF RESEARCH

The Air Transportation Technology Program at Princeton University, a program emphasizing graduate and undergraduate student research, proceeded along six avenues during the past year:

- Microburst Hazards to Aircraft
- Intelligent Failure-Tolerant Control
- Computer-Aided Heuristics for Piloted Flight
- Stochastic Robustness of Flight Control Systems
- Neural Networks for Flight Control
- Computer-Aided Control System Design

This research has resulted in a number of publications, including a thesis, archival papers, and conference papers. An annotated bibliography of publications that appeared between June, 1990 and June, 1991 appears at the end of this report. The research that these papers describe was supported in whole or in part by the Joint University Program, including work that was completed prior to the reporting period.

Severe downdrafts and resulting high velocity outflows caused by microbursts present a significant hazard to aircraft on takeoff and final approach. *Microbursts*, which are often associated with thunderstorm activity, also can occur in the vicinity of dissipating convective clouds that produce no rainfall at ground level. Microburst encounter is a rare but extremely dangerous phenomenon that accounts for one or two air carrier accidents and numerous general aviation accidents each year (on average). Conditions are such that an aircraft's performance envelope may be inadequate for safe penetration unless optimal control strategies are known and applied.

Our current wind shear research focuses on avoiding wind shear during transport aircraft operations, as well as on developing cockpit strategies for wind shear recovery. Graduate student Alex Stratton is developing an expert system for wind shear avoidance that extends the FAA Microburst Windshear Guidelines to account for temporal and spatial variations in the evidence that wind shear is present [1,2]. The approach being taken is to develop a Bayesian Belief Network that relates information gathered from many sources to determine the probability of encountering a microburst on the intended flight path. Our principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., abort, go-around, normal climbout, or glide slope), and for using the aircraft's full potential to combat wind shear. This study requires the definition of deterministic and statistical techniques for fusing internal and external information, for making "go/no-go" decisions, and for generating commands to the aircraft's autopilot and flight directors in automatic and manually controlled flight.

A number of graduate students have developed a fixed-base cockpit simulator for microburst studies, and graduate student Sandeep Mulgund has begun to use the simulator to determine the feasibility of target-pitch-angle (TPA) guidance for propellor-driven commuter-type aircraft. The simulation incorporates a cockpit station with manual input devices, a graphical display of instruments, and an out-the-window view. Currently the simulator is programmed to simulate a twin-jet transport aircraft and a twin-engine general aviation airplane. Preliminary results compare TPA guidance with an exact optimal control history, and they show the significant differences between best target pitch angles for head-tailwind shear and downdraft encounters. (Although the best single angle of attack for wind shear encounter is essentially the same for equivalent horizontal shears and vertical downdrafts, the corresponding pitch angles are decidedly different.) Graduate student Darin Spilman has used the simulator to conduct a preliminary analysis of an encounter of the twin-jet transport with an intense wind "rotor" capable of rolling the aircraft to inverted attitude, and he will examine effects of unsteady aerodynamics and heavy rain in his future research. Prior research on optimal flight paths conducted by Mark Psiaki is documented in [3].

Undetected system failures and/or inadequately defined recovery procedures have contributed to numerous air carrier incidents and accidents. The infamous DC-10 accident at Chicago's O'Hare Airport, in which loss of an engine pod, subsequent loss of subsystems, and asymmetric wing stall led to disaster, provides a prototype for the kind of tragedy that could be averted

by intelligent, failure-tolerant flight control systems. A survey of related considerations and control design methods is contained in [4].

Helping a pilot make quick decisions under high workload conditions is important for aircraft missions of all types. In research principally supported by an Army/Navy grant but reported at numerous quarterly reviews of the Joint University Program, Brenda Belkin developed an expert system of expert systems called AUTOCREW. In her M.S.E. thesis, Ms. Belkin used the paradigm of a hypothetical aircraft crew to facilitate the assignment of tasks, rules, and data within parallel knowledge bases. Ms. Belkin was the recipient of the 1990 William E. Jackson Award of the Radio Technical Commission for Aeronautics for her thesis, and the research is further presented in [5-7].

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model. Graduate student Laura Ryan Ray completed her Ph.D. thesis on this topic [8] and co-authored a number of related papers [9-11]. Chris Marrison currently is applying *Stochastic Robustness Analysis* to ten controllers designed in response to the 1990 American Control Conference Benchmark Control Problem challenge [12]. His results show the marked disparity in control systems designed by various methods, not only in their nominal performance but in the likelihood that they will achieve design goals.

There is growing interest in the use of neural networks for computational decision-making and control, brought about by the advent of small, fast, inexpensive computers. The neural network paradigm offers a potentially attractive structure for flight control systems that adapt to changing flight conditions and system failures, but much is to be learned about the practicality of such an approach. Graduate student Dennis Linse has begun to examine this potential. Current research focuses on the application of the feed-forward back-propagation network to identifying and modeling the nonlinear aerodynamics of a twin-jet transport aircraft [13].

Graduate student Subrata Sircar has begun to examine concepts for the next generation of computer-aided flight control system design through development of a comprehensive computer program called *FlightCAD*. The

program contains a variety of modeling, synthesis, simulation, and evaluation alternatives, and it will be applied to the design of flight control logic for the *1991 AIAA Controls Design Challenge*. It is organized around a desktop metaphor that takes advantage of unique capabilities of the NeXT Computer. A direct digital synthesis technique is employed; it will produce a proportional-integral-filter controller with scheduled linear-quadratic-Gaussian gains. Tight following of pilot commands will be assured by a forward-loop command generator tracker, and the controller will be sufficiently robust to account for specified levels of parameter uncertainty. A principal feature of the control design package is the enhanced ability to iterate and search during the modeling, design, and analysis process.

ANNOTATED BIBLIOGRAPHY OF 1990-91 PUBLICATIONS

1. Stratton, D.A., and Stengel, R., Probabilistic Reasoning for Intelligent Wind Shear Avoidance, *Proceedings of the 1990 AIAA Guidance, Navigation & Control Conference*, Portland, OR, Aug 1990, pp. 1099-1107.*

Avoiding severe wind shear challenges the ability of flight crews, as it involves assessing risk from uncertain evidence. A computerized intelligent cockpit aid can increase flight crew awareness of wind shear, improving avoidance decisions. A primary task in the development of such a cockpit aid is providing a means of assessing risk from evidence of wind shear from sources with varying reliability. The Federal Aviation Administration's Windshear Training Aid provides guidelines for assessing the risk of wind shear encounter from meteorological evidence. Use of these guidelines in the cockpit is complicated by uncertainty surrounding meteorological knowledge of wind shear. Bayesian network representation is discussed as a means of modeling this uncertain knowledge in a computer. A probabilistic model of the Windshear Training Aid guidelines using Bayesian network representation is presented. This model combines evidence from sources of varying reliability and incorporates results from meteorological studies of wind shear. The probabilistic model can provide flight crews with meaningful estimates of risk to aid their decisions, using evidence from a variety of sources and a base of meteorological knowledge.

2. Stratton, D.A., and Stengel, R., Stochastic Prediction Techniques for Wind Shear Hazard Assessment, *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, Dec 1990, pp. 702-707.*

The threat of low-altitude wind shear has prompted development of aircraft-based sensors that measure winds directly on the aircraft's intended flight path. Measurements from these devices are subject to turbulence inputs and measurement error, as well as to the underlying wind profile. In this paper, stochastic estimators are developed to process on-board doppler sensor measurements, producing optimal estimates of the winds along the path. A stochastic prediction technique is described to predict the hazard to the aircraft from the estimates as well as the level of uncertainty of the hazard prediction. The stochastic prediction technique is demonstrated in a simulated microburst wind shear environment. Use of the technique in a decision-making process is discussed.

---

* Conference Paper

3. Psiaki, M., and Stengel, R., Optimal Aircraft Performance During Microburst Encounter, *J. Guidance, Control, and Dynamics*, Vol. 14, No. 2, Mar-Apr 1991, pp. 440-446.**

The effects of microburst characteristics on the optimal penetration performance of jet transport and general aviation aircraft are presented. The purpose is to determine the best possible performance that can be achieved in a broad range of microbursts. A secondary goal is to illustrate good strategies for dealing with a range of microbursts during takeoff and landing. Over 1100 optimal trajectories were computed for two aircraft types flying through idealized microbursts using a Successive Quadratic Programs trajectory optimization algorithm. Contours of safety metrics are plotted as functions of the length scales, magnitudes, and locations of horizontal wind shears and vertical downdrafts. These performance contours show three length-scale regimes for optimal microburst penetration. At short length scales, hazards usually associated with gustiness predominate. At intermediate length scales, a degraded ability to maintain flight path and/or vertical velocity poses the most serious threat. At very long microburst length scales, excessive touchdown velocities may result. The ability to transit a microburst successfully also varies strongly with microburst location. The results show that both aircraft types could penetrate some very severe microbursts if optimal control histories were followed. Nevertheless, these control strategies assume perfect prior knowledge of the wind, and practical limits to successful encounter with real-time control capabilities would be lower. The optimally controlled jet transport can successfully penetrate higher intensity microbursts than can the general aviation aircraft.

4. Stengel,. R., Intelligent Failure-Tolerant Control, *Proceedings of the 5th IEEE International Symposium on Intelligent Control*, Philadelphia, Sept 1990, pp. 548-557. (To appear in the *IEEE Control Systems Magazine.*)*

An overview of failure-tolerant control is presented, beginning with robust control, progressing through parallel and analytical redundancy, and ending with rule-based systems and artificial neural networks. By design or implementation, failure-tolerant control systems <u>are</u> "intelligent" systems. All failure-tolerant systems require some degree of robustness to protect against catastrophic failure; failure tolerance often can be improved by adaptivity in decision-making and control, as well as by redundancy in measurement and actuation. Reliability, maintainability, and survivability can be

** Archival Paper

enhanced by failure tolerance, although each objective poses different goals for control system design. Artificial intelligence concepts are helpful for integrating and codifying failure-tolerant control systems, not as alternatives but as adjuncts to conventional design methods.

5. Belkin, B., and Stengel, R., Quantitative Knowledge Acquisition for Expert Systems, *Engineering Applications of Artificial Intelligence*, Vol. 3, No. 4, Dec 1990, pp. 271-281.**

A common problem in the design of expert systems is the definition of rules from empirical data obtained in system operation or simulation. While it is relatively easy to collect numerical data and to log the comments of human operators engaged in experiments, generalizing such information to a set of rules has not previously been a straightforward task. This paper presents a statistical method for generating the needed rule base from numerical data, motivated by an example based on vehicle navigation with multiple sensors. The specific objective is to design an expert system that selects a satisfactory suite of measurements from a dissimilar, redundant set, given an arbitrary navigation geometry and possible sensor failures.

6. Belkin, B., and Stengel, R., Knowledge Acquisition for Expert Systems Using Statistical Methods, *Knowledge Based System Applications for Guidance and Control*, AGARD CP-474, Madrid, Sept 1990, pp. 25-1 to 25-17.**

This paper expands on the previous paper, describing the systematic development of a Navigation Sensor Management (NSM) Expert System from Kalman Filter covariance data. The development method consists of the two statistical techniques: *Analysis of Variance (ANOVA)* and the *ID3 algorithm*. The ANOVA technique indicates whether variation of a problem parameter gives *statistically* different covariance results, and the ID3 algorithm identifies the *relationships* between the problem parameters using probabilistic knowledge extracted from a simulation example set. ANOVA results show that statistically different position accuracies are obtained when different navaids are used, the number of navigation aids is changed, the trajectory is varied, or the performance history is altered. By indicating that these four factors significantly affect the decision metric, an appropriate parameter framework was designed, and a simulation example base was created. The example base contained over 900 training examples from nearly 300 simulations. The ID3 algorithm was used to determine the NSM Expert's classification "rules" in the form of *decision trees*. The performance of these decision trees was assessed on two arbitrary trajectories, and

the performance results are presented using a predictive metric. The test trajectories used to evaluate the system's performance show that the NSM Expert adapts to new situations and provides reasonable estimates of the expected hybrid performance. The results also show how the NSM Expert chooses optimal or next-best navigation strategies when limited computational resources are available; in simple cases, its solutions are commensurate with the designer's intuition.

7. Belkin, B., and Stengel, R., Systematic Methods for Knowledge Acquisition and Expert System Development, *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, Dec 1990, pp. 2191-2197 (To appear in *IEEE Systems Magazine*).*

Nine interacting rule-based systems collectively called AUTOCREW were designed to automate functions and decisions associated with a combat aircraft's subsystems. The organization of tasks within each system is described; performance metrics were developed to evaluate the workload of each rule base and to assess the cooperation between rule bases. Each AUTOCREW subsystem is composed of several expert systems that perform specific tasks. The NAVIGATOR was analyzed in detail to understand the difficulties involved in designing the system and to identify the tools and methodologies that ease development.

8. Ray, L.R., Stochastic Robustness of Linear Multivariable Control Systems: Towards Comprehensive Robustness Analysis, Ph.D. Thesis, Princeton University, Princeton, January, 1991.+

Stochastic robustness, a simple numerical procedure for estimating the stability and performance effects of parameter uncertainty in multivariable, linear, time-invariant control systems, is presented. Based on Monte Carlo evaluation of the system's closed-loop eigenvalues (given parameter statistics), this analysis approach introduces the probability of instability as a scalar robustness measure. The related stochastic root locus, a portrayal of the root probability density, provides insight into modal variations of the system. Parameter uncertainties may be Gaussian or non-Gaussian, bounded or not. Confidence intervals for the scalar probability of instability address computational issues inherent in Monte Carlo simulation. Appropriate confidence intervals are presented, and aids in choosing the required number of Monte Carlo evaluations are developed.

+ Ph.D. Thesis

Stochastic performance robustness measures are based on classical design criteria, as well as criteria specific to a particular application. Robust control system synthesis concepts ensue from demonstrating the ability of stochastic robustness analysis to tradeoffs in design parameters, and optimization methods suited to maximizing stochastic robustness are suggested. Four examples demonstrate aspects of the analysis, the use of confidence intervals, stability-performance tradeoffs, and synthesis through optimization. It is concluded that analysis of stochastic robustness offers a good alternative to existing robustness metrics that is inherently intuitive and precise, that is easily implemented, and that has direct bearing on engineering objectives.

9. Ray, L.R., and Stengel, R., Stochastic Performance Robustness of Aircraft Control Systems, *Proceedings of the 1990 AIAA Guidance, Navigation & Control Conference*, Portland, OR, Aug 1990, pp. 863-873.*

Stochastic robustness, a simple technique used to estimate the robustness of linear, time-invariant systems, is applied to a twin-jet transport aircraft control system. Concepts behind stochastic *stability* robustness are extended to stochastic *performance* robustness. Stochastic performance robustness measures based on classical design specifications and measures specific to aircraft handling qualities are introduced. Confidence intervals for both individual stochastic robustness measures and for comparing two measures are presented. The application of stochastic performance robustness, the use of confidence intervals, and tradeoffs between performance objectives are demonstrated by means of the twin-jet aircraft example.

10. Ray, L.R., and Stengel, R., Computer-Aided Analysis of Linear Control System Robustness, *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, Dec 1990, pp. 3468-3469 (To appear in *Mechatronics*).

Stochastic robustness is a simple technique used to estimate the stability and performance robustness of linear, time-invariant systems. The use of high-speed graphics workstations and control system design software in stochastic robustness analysis is discussed and demonstrated.

11. Ray, L.R., and Stengel, R., Stochastic Robustness of Linear-Time-Invariant Control Systems, *IEEE Trans. Automatic Control*, Vol. 36, No. 1, Jan 1991, pp. 82-87.**

A simple numerical procedure for estimating the stochastic robustness of a linear, time-invariant system is described. Monte Carlo evaluation of the systems' eigenvalues allows the probability of instability and the related stochastic root locus to be estimated. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Confidence intervals for the scalar probability of instability address computational issues inherent in Monte Carlo simulation. Trivial extensions of the procedure admit consideration of alternate discriminants; thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions also can be estimated. Results are amenable to graphical presentation.

12. Stengel, R., and Marrison, C., Robustness of Solutions to a Benchmark Control Problem, *Proceedings of the 1991 American Control Conference*, Boston, June 1991.*

The stochastic robustness of solutions to a benchmark control design problem presented at the *1990 American Control Conference* has been analyzed. The analysis quantifies the controllers' stability and performance robustness with structured uncertainties in up to six system parameters. The analysis provides insights about system response that are not readily derived from other robustness criteria, providing a common ground for judging controllers produced by alternative methods.

13. Linse, D., and Stengel, R., A System Identification Model for Adaptive Nonlinear Control, *Proceedings of the 1991 American Control Conference*, Boston, June 1991.*

A neural network model for generalized-spline function approximation in nonlinear control is described. The control system contains three elements: a nonlinear-inverse-dynamic control law that depends on a comprehensive model of the plant, a state estimator whose outputs drive the control law, and a function approximation scheme that models the system dynamics. An extended Kalman filter provides data for continuous training of the neural network during normal operation. The results of an application of the identification techniques to a nonlinear transport aircraft model are presented.

Target Pitch Angle for the Microburst Escape Maneuver

Sandeep S. Mulgund
Department of Mechanical and Aerospace Engineering
Princeton University

Background

- FAA's Windshear Training recommends that upon recognition of encounter with a microburst, pilot should command full thrust and rotate aircraft to an initial target pitch attitude of 15°
- Advantages of the 15° target:
  - Easily recalled in emergency situations
  - Prominently displayed on attitude indicators
- It should be possible to postulate a target pitch angle for any airplane
- Objective of the present work is to better define the optimum target pitch angle, and develop a methodology for identifying the target pitch for any aircraft

The objective of this study was to investigate the constant pitch attitude strategy as a possible non-precision maneuver for recovery from inadvertent wind shear encounters. The Wind Shear Training Aid published by the FAA recommends that upon encountering a severe wind shear, the pilot should apply maximum thrust and rotate the aircraft to an initial pitch target angle of 15°. The 15° target was identified through rigorous analyses using six-degree-of-freedom flight simulators and microburst models representative of actual accident cases. It was found that 15° was an effective target for a wide range of shears, and was generally applicable to most jet transports.

This work was undertaken to examine the issue of recovery performance in wind shear of other classes of aircraft - notably turboprop commuters and propeller-driven general aviation planes. It should be possible to postulate a target pitch angle (TPA) for such aircraft, as well.

Effect of Windshear on Aircraft Performance

- The impact of a microburst on an aircraft's climb performance is described by the "F-Factor":

$$F = \frac{\dot{w}_x}{g} - \frac{w_h}{V}$$

- F is a function of the aircraft's trajectory through the wind shear, and relates directly to its constant speed climb gradient capability

- Maximum trim rate of climb under an imposed F-Factor was computed for a Cessna 402, under two different assumptions:

$$\text{(a) } F = \frac{\dot{w}_x}{g}$$

$$\text{(b) } F = -\frac{w_h}{V}$$

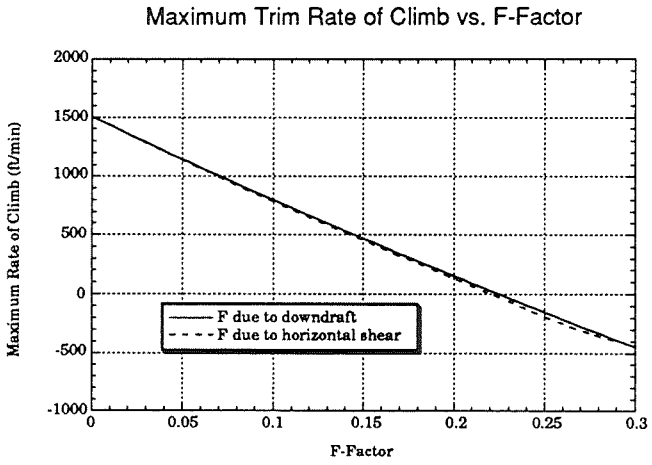
The effect of a wind shear on the climb performance of an airplane is well described the F-Factor, a hazard index developed at NASA Langley Research Center. The F-Factor depends on the time-rate-of-change of the horizontal wind component $\dot{w}_x$, the vertical wind component w_h, the airplane's true airspeed V, and the acceleration due to gravity g.

An F-Factor of about 0.15 effectively cancels out the climb capability of most transport category aircraft. In an analysis of five wind shear-related accidents, it was found that the aircraft experienced average F-Factors of 0.2 to 0.25.

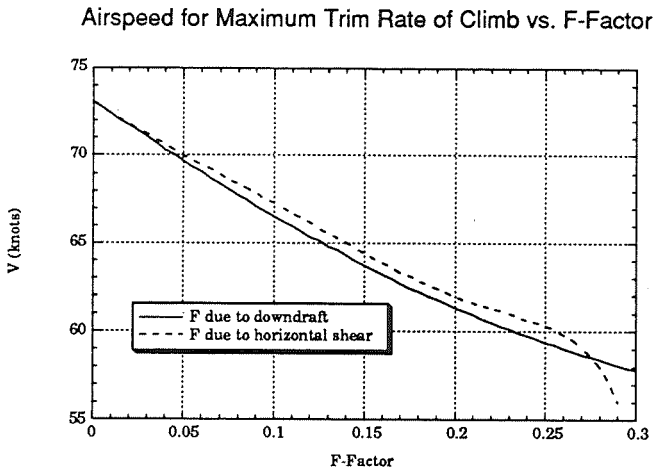
To gain insight into the effect of wind shear on climb performance, the maximum constant-speed rate of climb of a general aviation airplane was calculated as a function of an imposed F-Factor. Two different assumptions were made about the source of the F-Factor: first, that it was due entirely to the horizontal shear term, and second, that it was due entirely to the downdraft term.

Aircraft Model

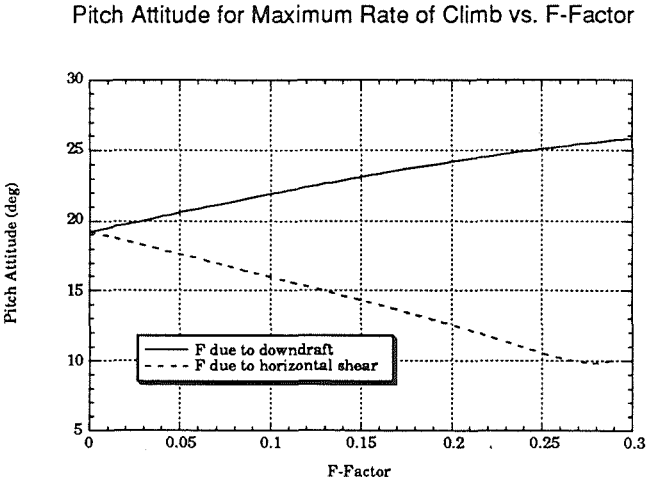
- Aerodynamic data taken from simulation model of the Cessna 402B developed at NASA Ames Research Center
- Point mass model of aircraft dynamics
- State variables: V, h, and γ
- Controls are throttle setting and pitch rate
- Climb performance in a wind shear was studied with the aircraft in initial approach configuration: 45° flaps, gear retracted



It was found that the maximum trim rate of climb of the airplane decreased almost linearly with the magnitude of F. It made little difference whether the F-Factor was due to a downdraft or a horizontal shear. It can be seen from this plot that if F exceeds 0.22, the airplane is incapable of maintaining altitude at any constant airspeed.



This plot indicates that in a wind shear environment, maximum rates of climb are achieved at airspeeds considerably lower than the "speed for best climb", which is about 73 knots for this airplane model in the approach configuration. This result may be non-intuitive to many pilots, who associate best rate of climb with a specific airspeed. The trend in airspeed for best climb vs. F is the same whether the F-Factor is due to a downdraft or horizontal shear, except at extremely high values of F.



It is here that the difference is seen between the effects of a downdraft and a horizontal shear on the climb performance of the airplane. In the case of a horizontal shear, the greater the F-Factor, the lower the pitch attitude for maximum rate of climb (or minimum rate of descent if F exceeds 0.22). Conversely, in a pure downdraft, the greater the F-Factor, the higher the pitch attitude for best rate of climb. This suggests that an effective recovery guidance strategy in a wind shear depends on identifying the source of the threat to the aircraft. In an actual wind shear encounter, an airplane will simultaneously experience horizontal shears and downdrafts.

The objective of this study is to develop a methodology for identifying a single target pitch angle for an airplane which is effective for a reasonable spectrum of possible wind shear encounters. This analysis of climb performance vs. F-Factor suggests that any such target pitch angle will be a compromise between two considerations - In downdrafts, best climb performance is achieved at relatively high pitch attitudes, while in a pure horizontal shear resulting in the same F-Factor hit, best climb performance is achieved at much lower attitudes.

Validation of the Target Pitch Angle

- Target pitch angle will be identified for two aircraft:

	MTOW (LB)	Powerplant
Cessna 402	6,300	2 x 300 HP
B737-100	93,500	2 x 12,000 LB

- Recovery performance at the target pitch angle will be evaluated by simulating the response of each aircraft to a wind shear generated by an analytic downburst model

The TPA will be identified for two aircraft, representing drastically different weight and powerplant categories. A six-degree of freedom simulation model of the Boeing 737 is currently in use in Princeton's fixed-base simulator. The recovery performance of each aircraft will be simulated using an analytic model of a microburst wind shear.

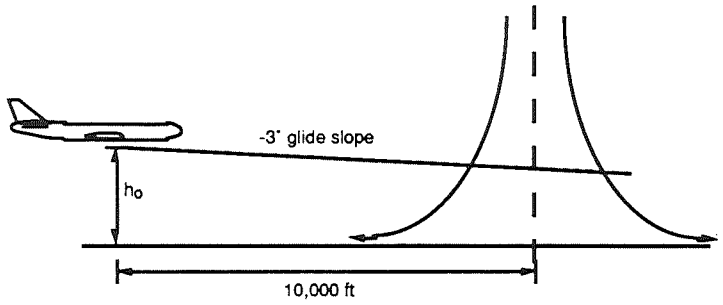
Analytic Downburst Model

- Developed at NASA LaRC by Rosa Oseguera and Roland Bowles.
- Represents an axisymmetric stagnation point flow, based on velocity profiles from the Terminal Area Simulation System Model.
- Time invariant microburst
- Permits simulation of different shears through the specification of three parameters:

  1) Radius of the downdraft column
  2) Maximum horizontal wind speed
  3) Altitude of maximum outflow

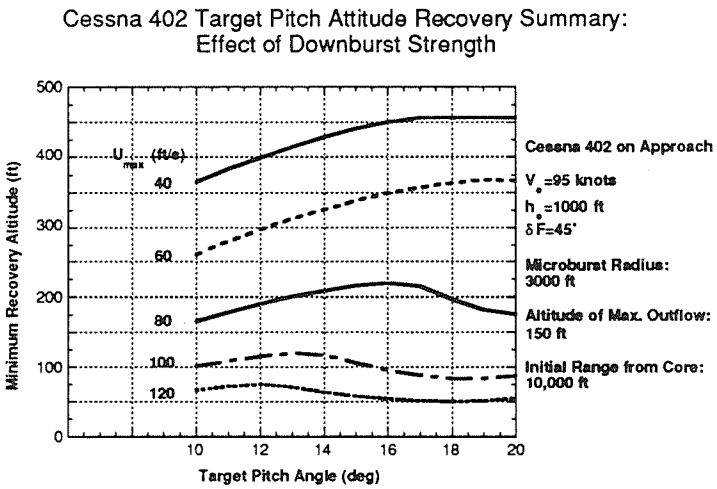
- Wind components in x,y, and z directions are easily obtained from the velocity equations for a given aircraft position relative to downburst core.

Simulation of Microburst Encounters



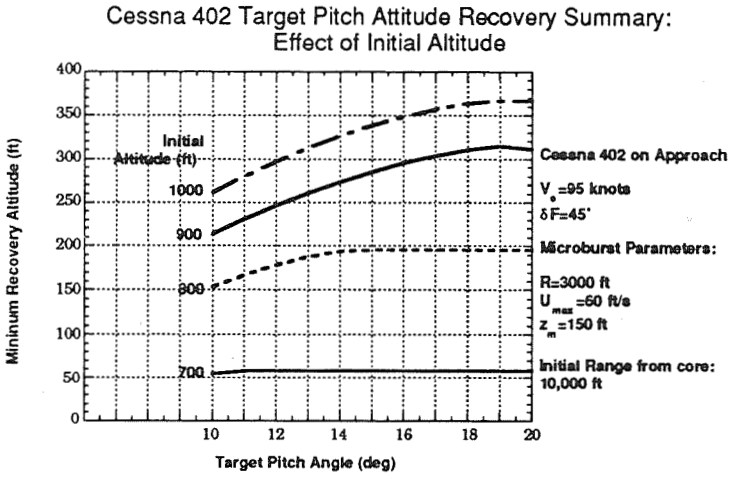
- Aircraft configured on approach along nominal glide slope
- Microburst core placed directly on flight path

Batch simulations of the recovery performance of the Cessna 402 were conducted by initializing the airplane on the glideslope 10,000 ft away from the microburst core. The initial altitude h_o was varied during the simulation runs, as was the strength of the microburst. The control logic was set up to track the glide slope until the F-Factor exceeded 0.1, at which point a recovery was initiated by applying full thrust and rotating the aircraft to the TPA. "Stick Shaker" was respected by limiting the angle of attack to 12°.



The impact of shear strength was assessed by conducting a series of batch simulations in which the radius of the microburst and the initial location of the airplane relative to its core were kept fixed, and the maximum horizontal wind speed was varied.

The trends in the results suggest that the TPA for maximizing the minimum recovery altitude in a wind shear encounter during approach depends on the strength of the shear. As the maximum horizontal wind speed was increased, it was found that the pitch attitude for maximum recovery altitude decreased. If the target pitch angle was set to one greater than that which maximized recovery altitude, the aircraft was exposed to prolonged periods of stall warning and reduced recovery performance.



The impact of the altitude at which a wind shear encounter occurs was assessed by fixing the size and strength of the microburst, and varying the initial altitude of the airplane. It was found that as the initial altitude was decreased, so did the target pitch angle that maximized minimum recovery altitude. This result is consistent with those from the analysis of climb performance vs. F-Factor. It was found earlier that in horizontal shears, pitch attitude for best rate of climb decreased with the magnitude of the corresponding F-Factor. Conversely, in downdrafts, pitch attitude for best rate of climb increased with the magnitude of F. The simulation results can be interpreted in this context: In low altitude encounters, the aircraft generally encounters "more" horizontal wind than downdraft, simply because vertical winds are smaller nearer to the ground. Consequently, it might be expected that in very low-altitude microburst encounters, best recovery performance is biased toward relatively low target pitch angles.

Future Work

- Develop a concise definition of the target pitch angle.
- Evaluate wind shear recovery performance in batch simulations for the two aircraft under consideration.
- Evaluate impact of including pilot in the loop using piloted simulations of wind shear encounters on the LCA Fixed-Base Flight Simulator.

The simulation results of the Cessna 402 model showed what pitch attitudes might be needed for good recovery performance in a wind shear. It was found that there was not one single pitch attitude which produced the best recovery performance in all cases. The target pitch attitude for best recovery depended on the strength of the shear and the altitude at which the encounter took place. It remains to determine what classes of wind shear encounters are most suitable for the definition of the target pitch angle.

There are significant implementation issues associated with the definition of the TPA. For example, can a pilot really distinguish 13.5° from 15°? A typical pitch attitude indicator has markings at 10, 12.5, 15, 17.5, and 20°. As a practical matter it may be necessary to limit any postulated TPA to one of these values. Also, there is the issue of how well a pilot can maintain any particular pitch attitude in the dynamic environment of a wind shear. In addition to trying to maintain some pitch attitude, a pilot will also be using the ailerons to maintain a near wings-level attitude. These issues will be explored by conducting piloted six-degree-of-freedom simulations in the fixed-base flight simulator in the Laboratory for Control and Automation. Presently , this facility is configured to simulate a Boeing 737-100 transport.

N92-17997

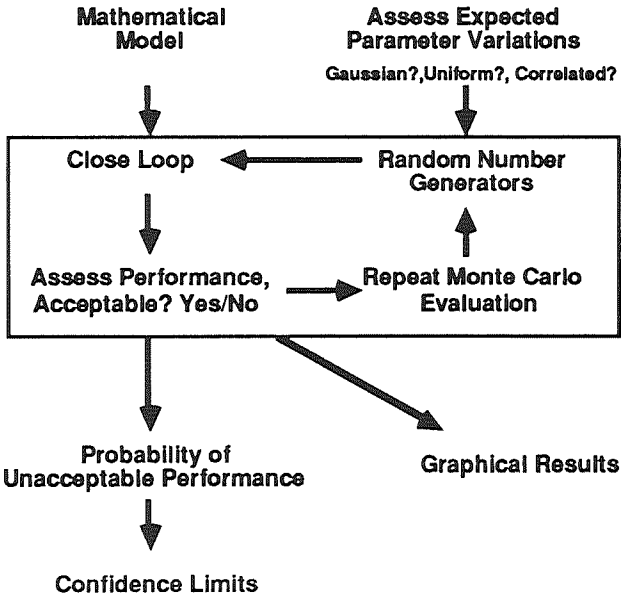
Stochastic Robustness

C. Marrison

Joint University Program

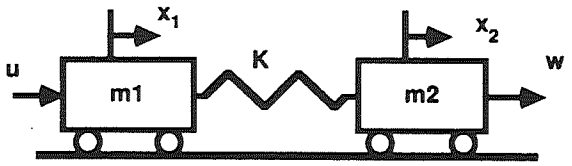
Stochastic robustness analysis (SRA) gives a direct, scalar measure of a conrol system's robustness by assessing the probability that the actual system will have acceptable performance.

Stochastic Robustness Analysis



To carry out stochastic robustness analysis, an expected probability distribution is assigned to each uncertain parameter in the system. The Monte Carlo analysis proceeds by repeatedly assigning shaped random values to each plant parameter, evaluating the stability or performance metric, and performing the binary classification (stable/unstable, etc.). If the system is stable, the state response to a unit disturbance impulse can be propagated to establish whether the response would violate settling time envelopes and whether peak actuator use would violate predetermined maximums. The final estimates of the probability of each form of unacceptable behavior are found by dividing the number of cases in which the overall system had that form of unacceptability by the number of cases run. Stability robustness can be portrayed graphically using the stochastic root locus and by using histograms of parameter values found in the unacceptable cases.

Benchmark Problem



$$y = x2 + v$$
$$z = x2$$

$$T_{wz} = \frac{(1/m_2)(s^2 + k/m_1)}{s^2[s^2 + k(m_1 + m_2)/m_1 m_2]}$$

$$k_o = 1, \quad m_1 = 1, \quad m_2 = 1$$

Nominal Roots at 0, 0, ±1.41

This benchmark problem was presented at the 1990 American Controls Conference.

The benchmark plant consists of a dual-mass/single-spring system with non-colocated sensor and actuator, as shown in the Fig where x1 and x2 are the positions of the two masses, x3 and x4 are their velocities, and u is a control force on m1. The plant is subject to the disturbance w on m2, and the measurement of x2 is corrupted by noise v.

The baseline plant is undamped, with eigenvalues at ±j(k(m1 + m2)/m1m2), 0, and 0. By the problem specification, a single-input/single-output feedback controller must close its loop around T_{wz}.

Benchmark Design Task

Hard Requirements

1) **The closed loop system should be Stable for $0.5 < k < 2$.**

2) **The Settling Time after an impulsive disturbance on the nominal system should be less than 15 secs.**

Soft Requirements

3) **The system should be robust against variations in m_1 and m_2.**

4) **There should be minimal control usage.**

5) **The rejection of noise should be good.**

Optional

6) **The system should compensate for a 0.5 rad/s disturbance input.**

Three design problems are posed. Problem 1 requires a) 15-sec settling time for unit disturbance impulse and nominal mass-spring values (m1 = m2 = k = 1), and b) closed-loop stability for fixed values of mass and $0.5 < k < 2$. It is further directed that "reasonable" robustness should be achieved and that controller effort and controller complexity should be minimized. An optional problem replaces the unit disturbance by a sinusoidal disturbance with 0.5-rad frequency. Asymptotic rejection of this signal should be achieved with a 20-sec settling time for the nominal system.

Parameter Variations for Testing

All Probability Distributions Uniform

1) $0.5 < k < 2$

2) $0.5 < k < 2$
$0.5 < m_1 < 1.5$
$0.5 < m_2 < 1.5$

3) $0.5 < k < 2$
$0.5 < m_1 < 1.5$
$0.5 < m_2 < 1.5$
$0. < c < 0.1$ (internal damping)
$0.9 < f < 1.1$ (loop gain variation)
$0.001 < \tau < 0.4$ sec. (actuator lag)

Closed Loop Transfer Function

$$Twz = \frac{(s^2 m_1 + sc + k)}{(s^4 m_1 m_2 + s^3 c(m_1 + m_2) + s^2 k(m_1 + m_2)) - f(s\tau + 1)(sc + k)(C)}$$

where C is the controller transfer function.

These are the parameter variations used for the Monte Carlo Analysis.

Performances Assessed

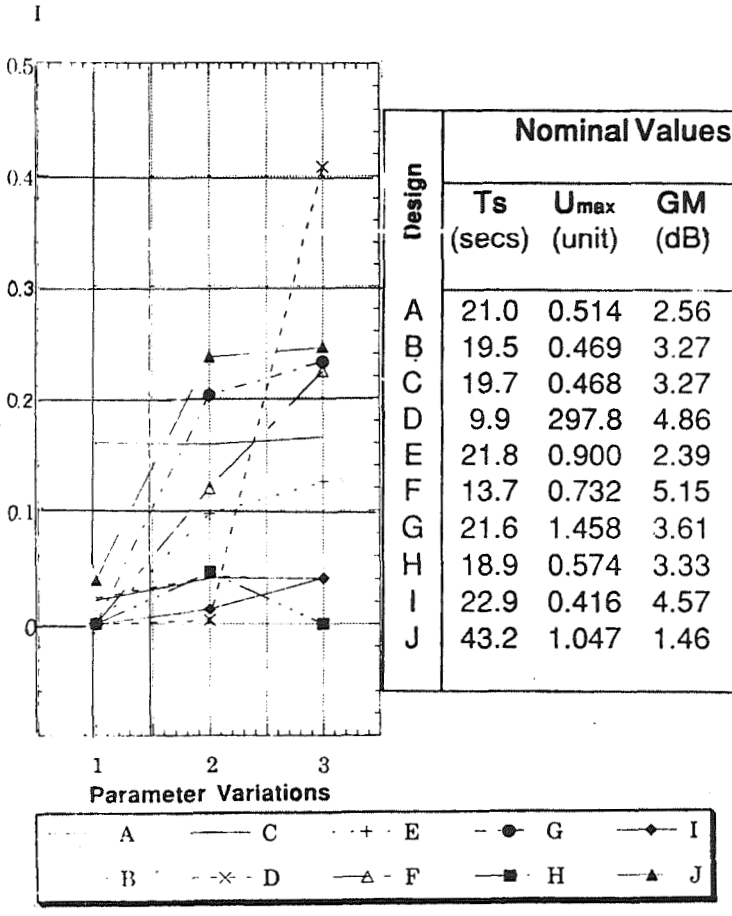
$\mathbb{P}_I$, ***probability of instability.*** **Tested by eigenvalue calculation.**

$\mathbb{P}_{Ts0.1}$, ***probability of settling time exceedance.*** **Response to unit w impulse falling outside a ±0.1 - unit envelope >15 secs after impulse**

$\mathbb{P}_{u1}$, ***probability of control-limit exceedance.*** **Using same time histories testing if peak actuator use >1 in response to unit disturbance impulse**

$\mathbb{P}_f$, ***probability of unsatisfactory sinusoidal disturbance rejection.*** **Steady state frequency response at w = 0.5 rad/sec tested for Mag > 0dB.**

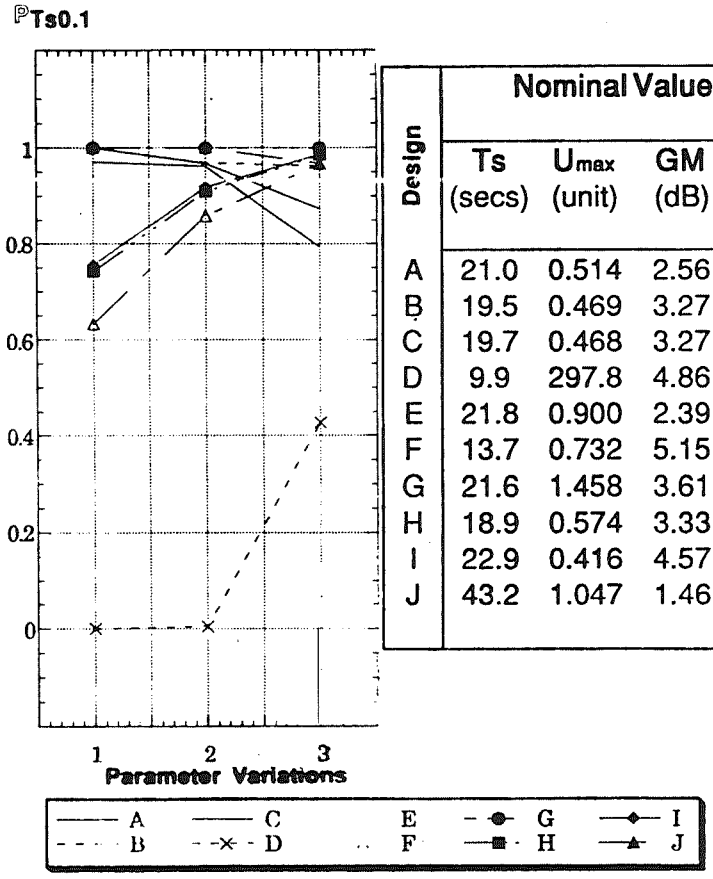
Results of Stability Assessment



Design	Nominal Values			
	Ts (secs)	U_{max} (unit)	GM (dB)	PM (deg)
A	21.0	0.514	2.56	26.7
B	19.5	0.469	3.27	26.8
C	19.7	0.468	3.27	26.5
D	9.9	297.8	4.86	58.7
E	21.8	0.900	2.39	22.0
F	13.7	0.732	5.15	23.7
G	21.6	1.458	3.61	25.4
H	18.9	0.574	3.33	24.5
I	22.9	0.416	4.57	27.5
J	43.2	1.047	1.46	63.0

The results of the stability analysis show several interesting features. For instance design D has a large gain and phase margin and a very good nominal settling time; and with parameter variations 1 and 2 the probability of instability is the least of all the designs but when the slight time delay is added in Parameter variation 3 its probability of instability is the worst. Looking at Design I we see a very similar gain and phase margin yet its probability of instability is one of the best under parameter variation 3. This shows that gain and phase margin can be very poor indicators of relative stability.

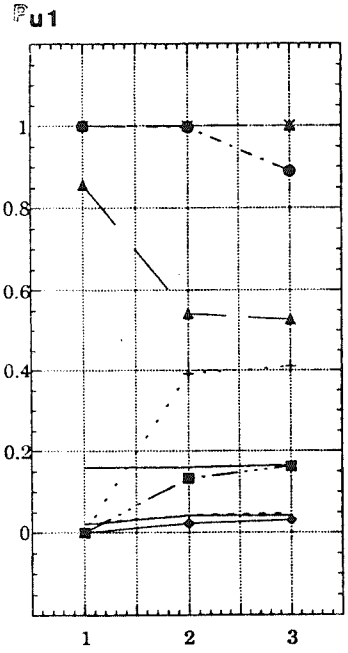
Results of Settling Time Assessment



Design	Nominal Values			
	Ts (secs)	Umax (unit)	GM (dB)	PM (deg)
A	21.0	0.514	2.56	26.7
B	19.5	0.469	3.27	26.8
C	19.7	0.468	3.27	26.5
D	9.9	297.8	4.86	58.7
E	21.8	0.900	2.39	22.0
F	13.7	0.732	5.15	23.7
G	21.6	1.458	3.61	25.4
H	18.9	0.574	3.33	24.5
I	22.9	0.416	4.57	27.5
J	43.2	1.047	1.46	63.0

The results of the probability of settling time violation again show us that the nominal times can be misleading. Design F has a nominal settling time of 13.7 secs but under parameter variation 3 there is a higher probability of it violating the 15 sec settling time than Design A which has a nominal settling time of 21 secs.

Results of Control Saturation Assessment

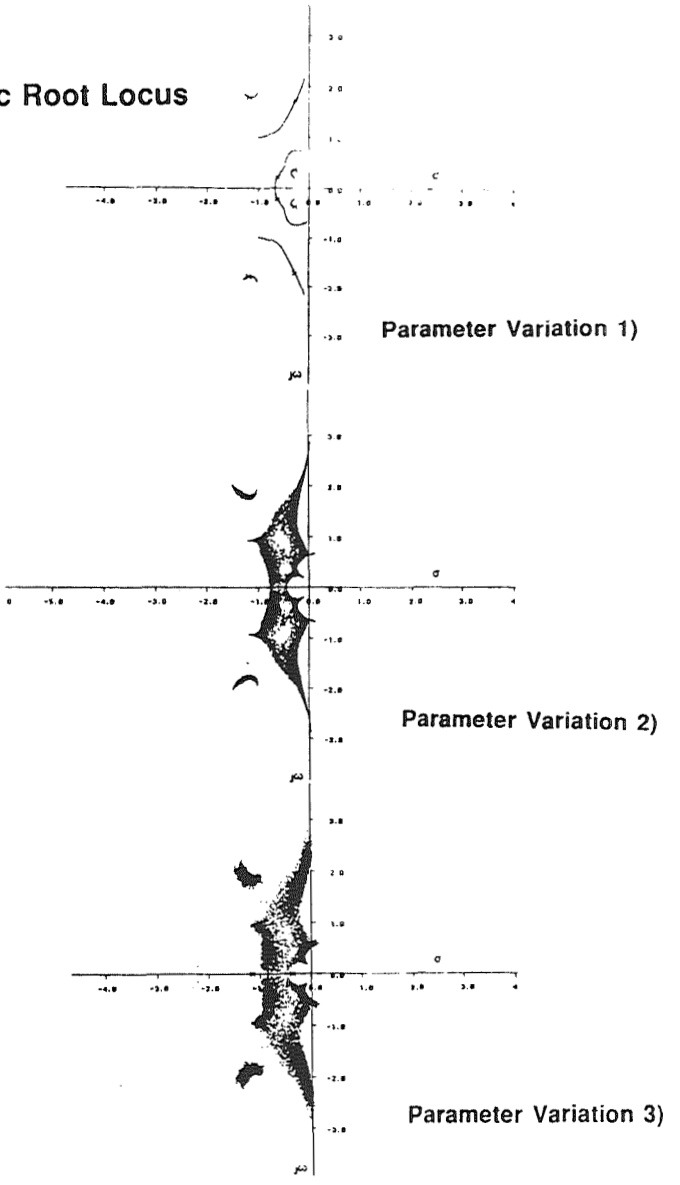


Design	Nominal Values			
	Ts (secs)	U_{max} (unit)	GM (dB)	PM (deg)
A	21.0	0.514	2.56	26.7
B	19.5	0.469	3.27	26.8
C	19.7	0.468	3.27	26.5
D	9.9	297.8	4.86	58.7
E	21.8	0.900	2.39	22.0
F	13.7	0.732	5.15	23.7
G	21.6	1.458	3.61	25.4
H	18.9	0.574	3.33	24.5
I	22.9	0.416	4.57	27.5
J	43.2	1.047	1.46	63.0

The probabilities of control saturation are reflected by the nominal values of control use.

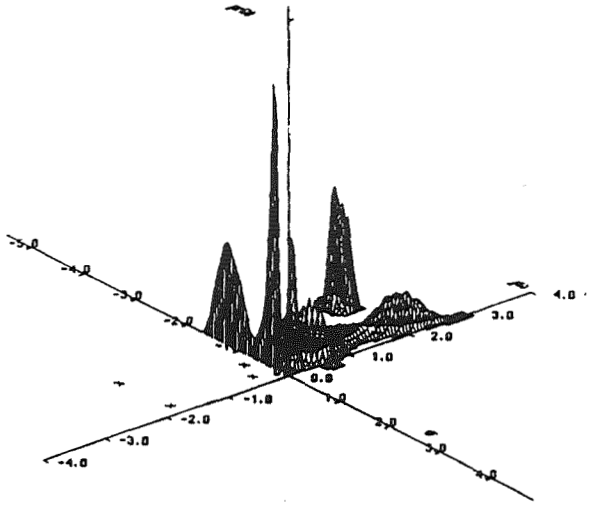
Stochastic Root Locus

Design H



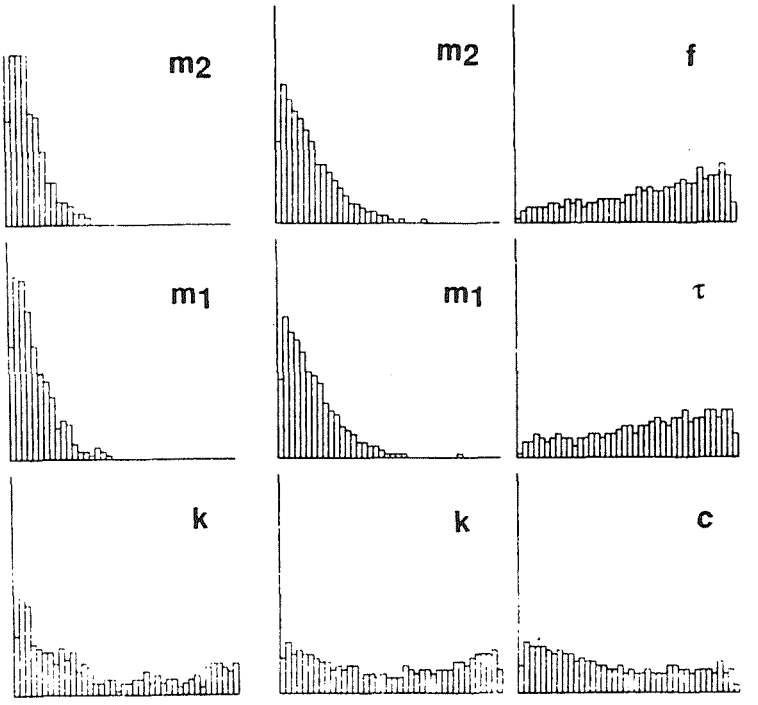
We can obtain graphical data regarding the effects of the parameter variations by looking at the stochastic root locus. For this design we find where the roots will most probably lie, how close the system is to instability and at what frequency unstable roots will have; here we can see two sets of unstable roots, one low frequency, the other high.

Three Dimensional Stochastic Root Locus



The root density can be more clearly seen by plotting in a third dimension.

Design H, Parameter Histograms



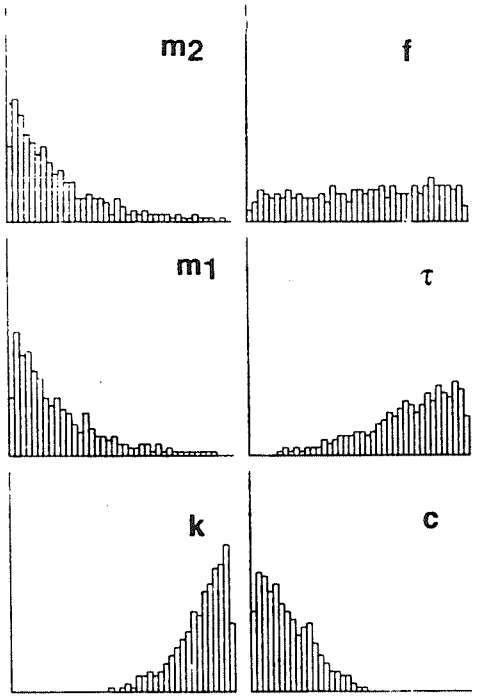
Parameter Variation 2)

Parameter Variation 3)

By storing the value of the parameters each time the closed loop is found to be unstable we can see which parameters are causing instability. Here we see that m1 and m2 have the strongest effect. From this graph we could suggest that lower nominal values of m1 and m2 should be used in the synthesis.

Design H, Parameter Histograms

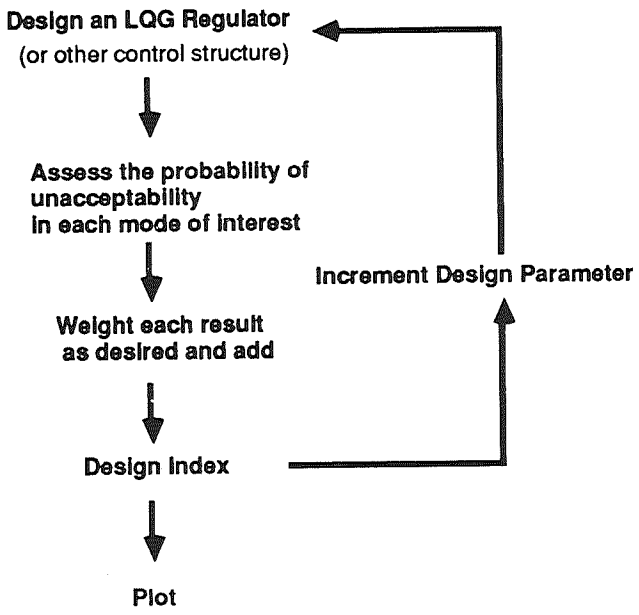
Parameter Variation 3)



High Frequency Roots

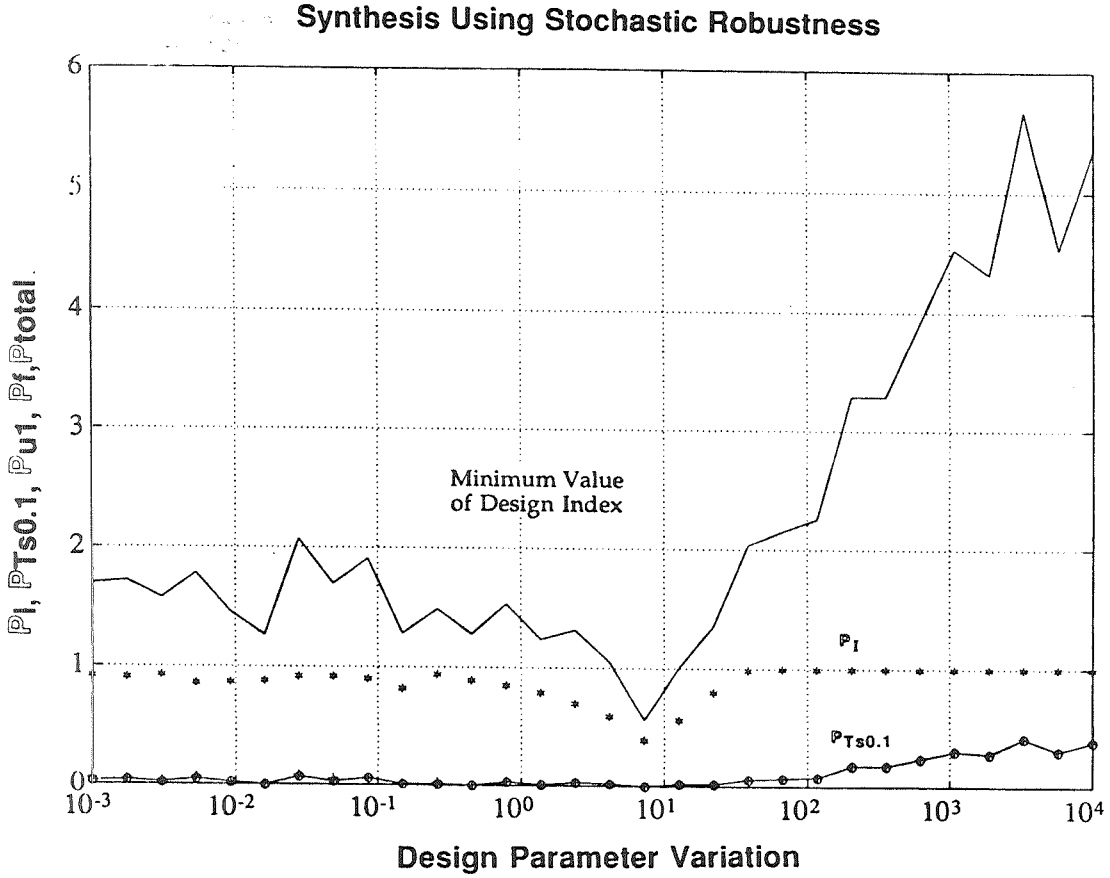
By storing parameters of the high and low frequency roots separately we can determine which parameters cause which instability. For instance here we see that low values of k do not cause high frequency instability.

Synthesis using Stochastic Robustness



Using the fact that Stochastic Robustness Analysis allows us to rank control systems in terms of their overall probability of performing satisfactorily we can carry out synthesis. We first create a series of similar controllers by adjusting a design parameter eg gain, then carry out a Monte Carlo analysis on each one.

Depending on which performances are considered important the results are weighted and added together eg 10*PI+3*Pu+1*PTs. The result is a curve showing the weighted probability of satisfactory performance against the value of the design parameter. There will typically be a minimum in this curve showing the most robust design.



In this rudimentary attempt at synthesis we can see a distinct minimum in the curve of the weighted design index. Setting the design parameter to this value ($10^{0.8}$) we will achieve a good combination of stability and performance robustness. This design seems to be better than the others synthesized for the benchmark problem but further fine tuning and analysis needs to be done.

Conclusions

We can use Stochastic Robustness techniques very flexibly

We can obtain information which is not obvious from other sources

We can rank control systems and suggest changes ie aid synthesis

Neural Networks for Aircraft System Identification

Dennis J. Linse
Department of Mechanical and Aerospace Engineering
Princeton University

Artificial neural networks offer some interesting possibilities for use in control. Our current research is on the use of neural networks as generalized splines for identification of aircraft aerodynamics. A system identification model is outlined which can be used to train neural networks on an aircraft model. The model can then be used in a nonlinear control scheme. The effectiveness of network training is demonstrated.

Nonlinear Control Law for Aircraft

The nonlinear dynamic equations of motion of an aircraft can be separated into two parts. The **f** term is a function of the state vector, **x**, only. It represents the well known dynamics and kinematics of the aircraft and is a function of the mass and inertias of the vehicle. The **g** term is a function of both the state and the controls, **u**. This term represents the aerodynamic and thrust interactions of the aircraft. These effects are less well known and are not as accurately modeled.

There are several methods for developing nonlinear control laws for systems. A Nonlinear-Inverse-Dynamic control law takes on the given form. The development of this control law requires that the **f** and **g** terms be differentiable with respect to the state and control vectors. This imposes limitations on the methods of approximation of the **g** vector.

- Nonlinear dynamic equations of an aircraft

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}, \mathbf{u}) + \mathbf{w}$$

f – dynamics and kinematics
g – aerodynamics and thrust

- Nonlinear Inverse Dynamics control law

$$\mathbf{u} = \mathbf{c}(\mathbf{x}) + \mathbf{C}(\mathbf{x})\mathbf{v}$$

- Requires differentiability and invertibility

System Identification Task

The system identification problem that we are investigating has additional constraints imposed on it due to the nonlinear control schemes selected. The identifier must make a functional approximation of the unknown portions of the model. It must also be a sufficiently smooth approximation so that the desired number of differentiations (typically 2 or 3 for aircraft applications) can be made in the nonlinear-inverse-dynamic type control law. We would also like to be able to update the approximation concurrent to operation, so an on-line model is desirable.

- Functional Approximation
- Smooth Fit
- On-line updates

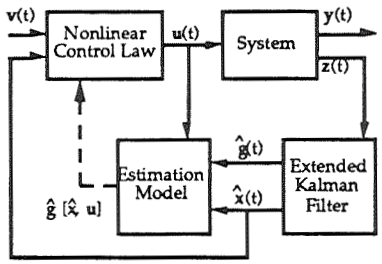
On-line Model Estimation

Based on the Estimation-Before-Modeling paradigm previously investigated at Princeton, an on-line estimation structure is developed. An Extended Kalman Filter is implemented which estimates an augmented state vector. This augmented vector includes both the original states and new states which represent the unknown portion of the original dynamic system, **g**. (Through non-dimensionalization, **g** is recognized as the six basic aerodynamic force and moment coefficients of an aircraft.) These unknown functions are modeled as random walk (Gauss-Markov) processes. The estimated values of **x** and **g**, along with **u**, are then given to an Estimation Model. The estimation model then has sufficient information to generate a functional form for **g**. As a side benefit, the Extended Kalman Filter also generates the estimate of the original state vector which is needed by most nonlinear control schemes.

Our interests are in artificial neural networks and they will be used as the estimation model in what is to follow.

Estimate augmented state vector

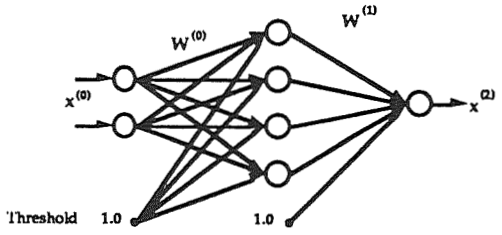
$$\mathbf{x}_a(t) = \{\mathbf{x}(t)^T \; \mathbf{g}[\mathbf{x}(t), \mathbf{u}(t)]^T\}^T$$



Feedforward Neural Networks

One of the most common artificial neural networks is the feedforward network. It uses a weighted interconnection of nonlinear nodes often separated into distinct layers. The nonlinear function of the network nodes is usually a squashing function like a sigmoid. When all of the weights and connections are determined, the network defines a (possibly) multi-input multi-output function.

The power of neural networks is the ability to learn from examples. Using the on-line estimate model previously described, a neural network can be trained to approximate any continuous function. Although many training algorithms are possible, the standard Back-Propagation algorithm was used for this research.



$$\mathbf{x}^{(k)} = \mathbf{s}^{(k)}\left[\mathbf{W}^{(k-1)}\,\mathbf{x}^{(k-1)}\right]$$

$$\mathbf{x}^{(N)} = \hat{\mathbf{g}}\left[\mathbf{x}^{(0)}\right]$$

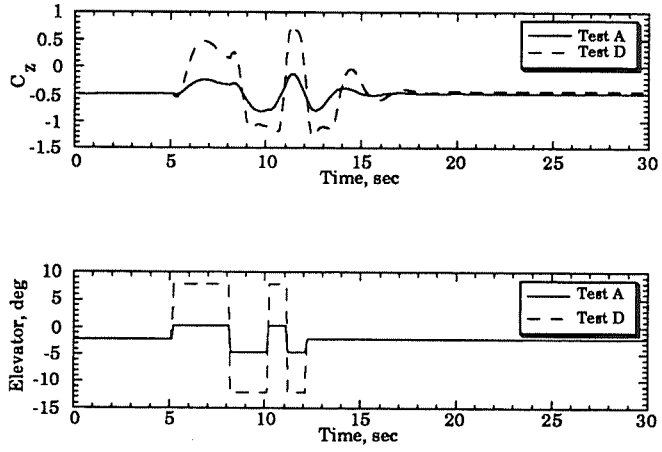
Feedforward Network Properties

Feedforward networks have many properties which are very useful in function approximation. The networks, as defined, perform generalized spline interpolation. The network nodes are the basis functions for this interpolation. Feedforward networks easily handle multivariate function which is a major difficulty for standard interpolation schemes. The networks have good localization properties, i.e. nearby inputs cause nearby outputs. This localization also has implications in training the networks. Finally, since the feedforward networks are weighted interconnections of nonlinear elements, the differentiability of the entire network depends on the differentiability of the individual nodes. Using the standard sigmoid as the nonlinear function in the nodes results in a network that is infinitely differentiable.

- Generalized spline interpolation
- Multivariate
- Input localization
- Differentiable

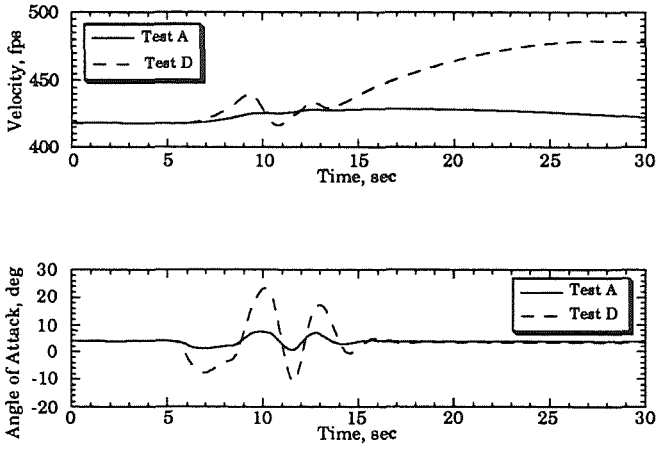
Network Training Input

To test the effectiveness of a neural network in learning the types of functions desired, an example is developed. First, a"3211" elevator input is put into a nonlinear simulation of a 737 aircraft. The amplitude of the input is varied from +/- 2.5 degrees to +/- 10 degrees in Tests A through D. The measurements of the aircraft are processed by an Extended Kalman Filter as detailed in the on-line estimation model. Included in the output is an estimate of the aircraft normal force coefficient, Cz. Since the normal force coefficient is essentially the negative of the lift coefficient, the stall of the aircraft can be seen in the Test D output at a normal force coefficient of approximately -1.2.



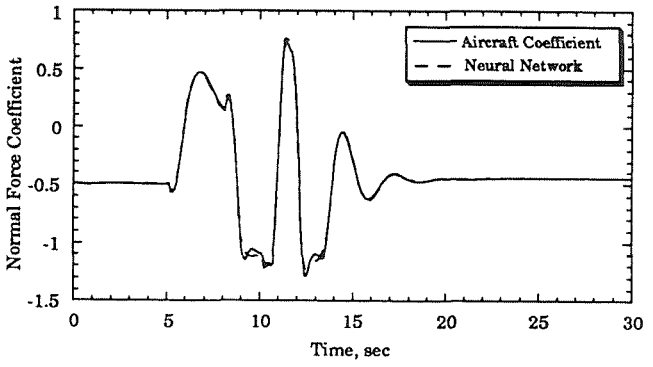
Network Training Input (cont.)

The velocity and angle of attack encountered during the training input is given here. The severest case, Test D, causes a 60 feet/second speed change and angles of attack ranging from -10 degrees to over 20 degrees. Since stall is in the 15 degree range for this aircraft, this test puts the aircraft through extensive maneuvers.



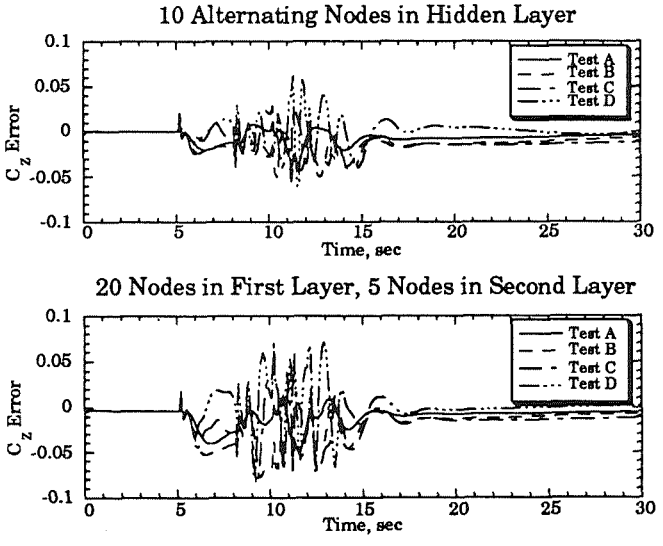
Neural Network Output

Using a single hidden layer network with 10 nodes, the network can accurately model the time evolution of the normal force coefficient of the aircraft. The expanded vertical scale allows the effects of stall in the 10 and 13 second periods to be viewed. About 800 presentations of test data A through D were used to train the network.



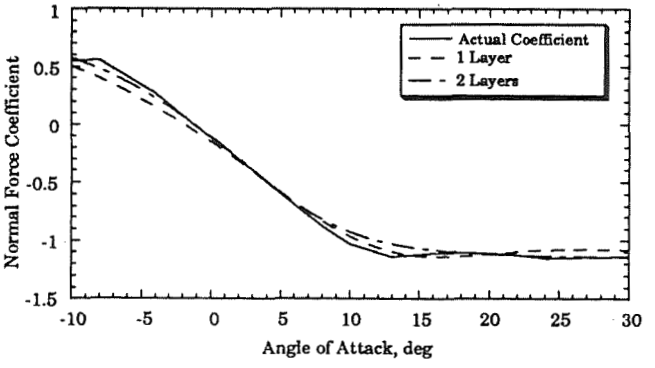
Network Coefficient Error

To better demonstrate how well the networks learn, the difference between the neural network output and the actual normal force coefficient are plotted for two different neural networks. While very busy, the bounds of the errors indicate that the networks have learned the coefficients to within about 5% during the maneuvers and much better at other times.



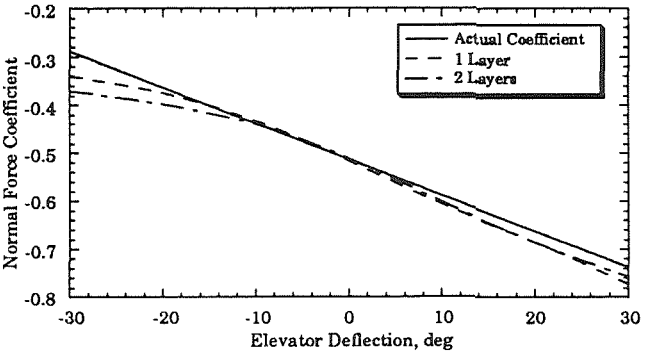
Network Aerodynamic Derivatives

For control of an aircraft, the derivatives of the aerodynamic coefficients are of possibly more importance than the value of the coefficient themselves. This slide shows the variation of the normal force coefficient with respect to the angle of attack of the aircraft. Once again, since the normal force coefficient is related to the lift coefficient, this curve is essentially the negative of the lift curve of the aircraft except at high angles of attack. The fit of the two different neural networks is very good especially near the trim angle of attack at about 4 degrees. The fit is reasonable at very high and low angles of attack too.



Network Aerodynamic Derivatives (cont.)

Looking at the variation of the normal force coefficient with the elevator deflection is also informative. Once again, the neural networks are able to reasonably accurately fit the desired curve, especially in the region of the trim position of about -2 degrees. This figure has an expanded abscissa. The network is only trained with elevator data between +/- 10 degrees, but does a reasonable job in the region outside of this. Care must be taken when the generalization ability of the networks are used like this.



Future Considerations

Future investigations of neural networks for system identification will take several directions. The networks here were trained starting from random initial weights. This increases the burden of learning by requiring the network both to find the form of the function and also to accurately match the function. By using a pre-training algorithm, possibly based on a least-squares fit, initial weights may be chosen which give a reasonable initial fit for the function and may indicate the number of nodes and layers that are needed for such fits. A second consideration is the conflicting demands of aircraft system identification and neural network training. The system identification needs inputs which will excite the aircraft while the network needs input information which will cover the input space sufficiently to allow good function fits everywhere. Since the aircraft is reasonably well understood, extra information, such as aircraft derivatives, may be beneficially used in the training process. Also, there are many other network training algorithms that may be advantageous to use.

- Initial Weight Selection
- Input Space Coverage
- Extra Training Information
- Network Training Algorithms

Conclusions

The system identification model method is an effective method to provide the information necessary to train neural networks on-line. Feedforward networks can be used to model the aircraft aerodynamic coefficients although there is room for improvement in both the fit of the functions and the methods used to train them.

- System identification model based on Estimation-Before-Modeling is novel and effective
- Feedforward neural networks can model aerodynamic coefficients
- Room for improvement in training and modeling

Optimal Aircraft Performance During Microburst Encounter

Mark L. Psiaki*
Cornell University, Ithaca, New York 14850
and
Robert F. Stengel†
Princeton University, Princeton, New Jersey 08540

The effects of microburst characteristics on the optimal penetration performance of jet transport and general aviation aircraft are presented. The purpose is to determine the best possible performance that can be achieved in a broad range of microbursts. A secondary goal is to illustrate good strategies for dealing with a range of microbursts during takeoff and landing. Over 1100 optimal trajectories were computed for two aircraft types flying through idealized microbursts using a Successive Quadratic Programs trajectory optimization algorithm. Contours of safety metrics are plotted as functions of the length scales, magnitudes, and locations of horizontal wind shears and vertical downdrafts. These performance contours show three length-scale regimes for optimal microburst penetration. At short length scales, hazards usually associated with gustiness predominate (e.g., high normal load factor, rotational upset). At intermediate length scales, a degraded ability to maintain flight path and/or vertical velocity poses the most serious threat. At very long microburst length scales, excessive touchdown velocities may result. The ability to transit a microburst successfully also varies strongly with microburst location. The results show that both aircraft types could penetrate some very severe microbursts if optimal control histories were followed. Nevertheless, these control strategies assume perfect prior knowledge of the wind, and practical limits to successful encounter with real-time control capabilities would be lower. The optimally controlled jet transport can successfully penetrate higher intensity microbursts than can the general aviation aircraft.

I. Introduction

ON August 2, 1985, a Delta Airlines Lockheed L-1011 crashed while attempting to land at the Dallas/Ft. Worth Airport because it encountered a severe microburst wind shear (like that in Fig. 1). The flight data records[1] provide insight into the combination of wind conditions and control strategy that caused this accident. During the headwind portion of the microburst, the aircraft was pitched up, apparently in an attempt to reduce airspeed. Because of the accompanying downdraft, it failed to experience much altitude gain, but it did experience a significant groundspeed loss. After regaining the initial approach airspeed, the aircraft kept losing airspeed rapidly because it entered the tailwind portion of the microburst. When the airspeed got within 10-15 kt of the stall speed, it was pitched down, apparently in an attempt to regain airspeed. It did regain airspeed, but in the process, it crashed into a highway 1.1 nm short of the runway. This is a classic example of how the conventional piloting strategy of maintaining airspeed with the elevator can spell disaster in a microburst encounter.

With this experience (and others like it) in mind, a number of efforts have attacked the microburst encounter problem from the perspective of improving control strategy.[2-16] Each control strategy realized significant performance improvements over existing systems in test cases. They did this by using pitch (and, in some cases, thrust) inputs to counteract the microburst's disturbing effects on such quantities as altitude and airspeed. Methodologies ranging from classical feedback design[3,11] to deterministic trajectory optimization[12,15,16] have been used. In fact, many of the control schemes were able to safely penetrate wind profiles that had caused serious accidents. This improved performance is encouraging information, but in most cases, the actual number of simulated wind shear encounters was small for each control law, so more extensive testing is clearly in order. Effects of variations in microburst intensity were considered in Refs. 9, 12, 15, and 16, whereas the effects of differing aircraft performance characteristics were treated in Refs. 12 and 16.

The present work is part of an extended effort to address the problem or to improve piloting strategies for microburst encounter.[4,11,12,17,18] The effort began with a classic control design study[11] and progressed to deterministic optimization of encounters with Doppler-radar-derived microburst wind fields.[12] The present paper characterizes the dependence of deterministically optimized penetration trajectories on microburst properties, including length scale, intensity, location, and relative magnitude of horizontal vs vertical winds. This dependence shows how different microbursts pose different types and levels of threat to an optimally controlled aircraft. The basic method of determining this dependence is to compute many deterministic optimal trajectories through a parametric family of microbursts. Results are presented for two different aircraft types, a small general aviation aircraft (GA)

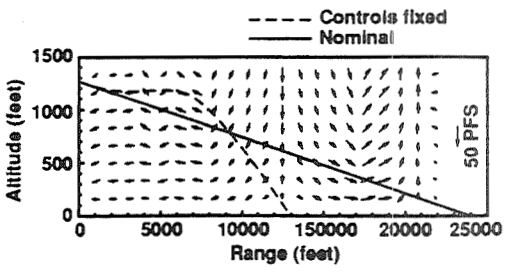


Fig. 1 Two JT approach trajectories through the most severe Joint Airport Weather Studies headwind tailwind shear.

Presented as Paper 88-4367 at the AIAA 15th Atmospheric Flight Mechanics Conference, Minneapolis, MN, Aug. 15-17, 1988; received Jan. 9, 1989; revision received Feb. 5, 1990.

*Assistant Professor, Mechanical and Aerospace Engineering. Member AIAA.

†Professor, Mechanical and Aerospace Engineering. Associate Fellow AIAA.

and a 150,000-lb jet transport (JT). The takeoff and landing-approach flight phases are both considered. Reference 17 contains further results in this area.

As a side benefit, the numerous trajectory optimizations of this study point to possible strategy improvements that could be realized by practical control systems, as noted. Reference 18 follows up on this idea, presenting simplified, realizable implementations of these strategies. It is similar to efforts of Miele et al., who have also used optimization results to deduce practical strategies.[7,8,13]

II. Analysis Highlights

Microburst Encounter Modeling

The longitudinal motions of an aircraft were modeled by seven coupled, scalar, nonlinear, ordinary differential equations that solve for inertial speed V_i, flight-path angle γ_i, pitch rate q_i, angle of attack α_i, altitude h, range r, and thrust or power T or P

$$\dot{V}_i = \frac{\{-\bar{q}S[C_D\cos(\alpha_i-\alpha_a)+C_L\sin(\alpha_i-\alpha_a)]+T\cos(\alpha_i+i_T)\}}{m} - g\sin\gamma_i \tag{1a}$$

$$\dot{\gamma}_i = \frac{\left[\dfrac{\bar{q}S[C_L\cos(\alpha_i-\alpha_a)-C_D\sin(\alpha_i-\alpha_a)]+T\sin(\alpha_i+i_T)}{m}\right] - g\cos\gamma_i}{V_i} \tag{1b}$$

$$\dot{q}_i = \frac{\bar{q}S\bar{c}C_M}{I_{yy}} \tag{1c}$$

$$\dot{\alpha}_i = q_i - \dot{\gamma}_i \tag{1d}$$

$$\dot{h} = V_i\sin\gamma_i \tag{1e}$$

$$\dot{r} = V_i\cos\gamma_i \tag{1f}$$

and

$$\dot{T} = \frac{(\delta T - T)}{\tau_e} \qquad \text{(JT aircraft)} \tag{1g}$$

or

$$\dot{P} = \frac{(\delta T - P)}{\tau_e} \quad \text{with} \quad T = \frac{E(V_a)P}{V_a} \qquad \text{(GA aircraft)} \tag{1h}$$

Here, the reference area is S, g represents gravitational acceleration, m and I_{yy} are aircraft mass and pitching moment of inertia, i_T is the thrust incidence angle, and δT is the throttle setting. The time constant for thrust (or power) lag τ_e is 4 s for the jet transport and 1 s for the general aviation aircraft; $E(V_a)$ represents the airspeed-dependent propeller efficiency of the general aviation aircraft. The wind shear effects on the aircraft's motions enter through the air-relative angle of attack α_a and the airspeed V_a,

$$\alpha_a = \alpha_i + \gamma_i - \tan^{-1}\left(\frac{V_i\sin\gamma_i + w_v}{V_i\cos\gamma_i + w_h}\right) \tag{2a}$$

$$V_a = \sqrt{V_i^2 + w_v^2 + w_h^2 + 2V_i(w_v\sin\gamma_i + w_h\cos\gamma_i)} \tag{2b}$$

and the dynamic pressure is computed as $\bar{q} = \rho V_a^2/2$.

The aerodynamic models for pitching moment, lift, and drag coefficients C_M, C_L, and C_D are quasisteady models that are reasonably accurate at low Mach number and below the stall angle of attack

$$C_M = C_{M_0} + C_{M_\alpha}\alpha_a + C_{M_{\delta E}}\delta E + C_{M_{\hat{q}}}\hat{q}_i + C_{M_{\dot{\hat{\alpha}}}}\dot{\hat{\alpha}}_a \tag{3a}$$

$$C_L = C_{L_0} + C_{L_\alpha}\alpha_a + C_{L_{\delta E}}\delta E + C_{L_{\hat{q}}}\hat{q}_i + C_{L_{\dot{\hat{\alpha}}}}\dot{\hat{\alpha}}_a \tag{3b}$$

$$C_D = C_{D_0} + \epsilon C_L^2 \tag{3c}$$

where δE is the elevator setting, $\hat{q} = q\bar{c}/2V_a$, $\dot{\hat{\alpha}} = \dot{\alpha}\bar{c}/2V_a$, and $\bar{c}$ is the mean aerodynamic chord. Table 1 gives the coefficient values used in this study. Much of the JT aeordynamic model was taken from Ref. 19. The GA coefficients are based on data in Ref. 20. Note that the available $C_{L_{max}}$ of each model can be derived from the C_{L_0}, C_{L_α}, and α_{max} data in Table 1; $C_{L_{max}} = 2.50$ for the JT; $C_{L_{max}} = 1.77$ for the GA. Trajectory optimization inequality constraints enforce these $C_{L_{max}}$ values by enforcing the inequality constraint $\alpha_a \le \alpha_{max}(<\alpha_{stall})$. Enforcing an α_{max} slightly less than α_{stall} (ulimate maximum of C_L) decreases the likelihood of departure from controlled flight. Note also that the zero lift line occurs at negative angle of attack in both cases, representing flaps-down configurations. Zero lift of the JT configuration occurs at $\alpha = -15.5$ deg because of the high value of C_{L_0} taken from Ref. 19; hence, JT negative angle-of-attack time histories appear anomalous (as in Fig. 4), but they correspond to positive C_L.

The $C_{L_{max}}$ is the significant quantity that limits an aircraft's ability to safely negotiate a microburst wind shear, and it takes on reasonable values for both configurations.

The seven scalar differential equations can be put in the form $\dot{x} = f(x,u,w)$ by defining the state vector, $x = (V_i, \gamma_i, q_i, \alpha_i, h, r, T \text{ or } P)^T$, the control vector, $u = (\delta E, \delta T)^T$, and the disturbance vector, $w = (-w_h, -w_v)^T$. Except for the w dependence, this form is a special case of the form needed for trajectory optimization.

For the present study, the microburst wind field was modeled as a function of range along the intended flight corridor; our prior optimization study considered altitude-dependent profiles measured in the Joint Airport Weather Studies (JAWS) program as well.[12] Five parameters characterize the engineering approximation model: the wave length and intensity of a sinusoidal headwind/tailwind, the wave length and intensity of a (1-cosine) downdraft, and the location of the microburst core (Fig. 2). This model has been found to produce aircraft trajectories like those through the JAWS microburst profiles.[12,17] Headwind/tailwind and downdraft effects were considered separately so that parametric trends could be identified.

Trajectory Optimization

The trajectory optimization problem is of the continuous-time, fixed-end time, free-end-point, Bolza type.

Find

$$u(t) \qquad \text{for} \qquad t \in [t_0, t_f] \tag{4a}$$

Table 1 Aerodynamic properties of the Jet Transport (JT) and General Aviation (GA) Models

Aircraft	C_{L0}	$C_{L\alpha}$	$C_{L\delta E}$	C_{Lq}	$C_{L\dot{\alpha}}$	α_{max}, deg
JT	1.36	5.04	0.40	9.3	6.6	13
GA	0.61	4.73	0.48	15.9	8.0	14

Aircraft	C_{M0}	$C_{M\alpha}$	$C_{M\delta E}$	C_{Mq}	$C_{M\dot{\alpha}}$	C_{D0}	ϵ
JT	0.0	−1.21	−1.43	−29.5	−17.7	0.064	0.067
GA	−0.09	−0.93	−1.35	−24.6	−12.3	0.066	0.055

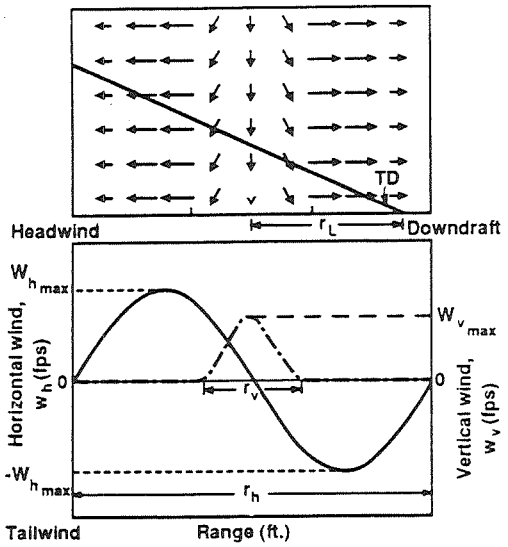


Fig. 2 An engineering-approximation wind shear model: a) vertical cross-section; b) variation of horizontal and vertical winds.

to minimize

$$J = \int_{t_0}^{t_f} L[x(t),u(t),t]\,dt + V[x(t_f),t_f] \tag{4b}$$

subject to

$$\dot{x} = f(x,u,t) \tag{4c}$$

$$x(t_0) = x_0 \tag{4d}$$

$$c[x(t),u(t),t] \le 0 \text{ for all } t\epsilon[t_0,t_f] \tag{4e}$$

where the typical $L = \{2[(h - h_0)\ \cos\gamma_{i0} - (r - r_0)\ \sin\gamma_{i0}]^2 + 0.04[\delta E - \delta E_0]^2 + 0.004[\delta T - \delta T_0]^2)\}$ and $V = \frac{1}{2}\Delta x_f^T P_N \Delta x_f$; they penalize deviations from the intended flight path. In these cost function expressions distances are expressed in feet, elevator angles in degrees, and throttle settings in percent of maximum. The quantities h_0 and r_0 are initial values; γ_{i0}, δE_0, and δT_0 are nominal values. The vector Δx_f is the final state perturbation from the nominal flight condition and flight path, and the matrix P_N is the solution of a discrete-time algebraic Riccati equation involving the sampled-data linearization of Eq. (1) about the nominal flight condition and involving the cost function integrand L. The function f describes the nonlinear longitudinal motions of an aircraft in a wind shear, c describes state and control inequalities that must be enforced for the optimal solution to make physical and practical sense (e.g., throttle saturation and stall angle of attack), x_0 gives the initial position and flight condition, and t_0 and t_f ensure that the trajectory traverses the entire microburst.

The function f contains global information about the wind shear because this is a deterministic formulation of the trajectory optimization problem; therefore, the optimal solution "knows" ahead of time what the microburst is going to do. This foreknowledge of the wind field makes the selection of x_0 important. If the initial range location is farther ahead of the center of the microburst, the optimization will have more time to prepare for the encounter. For example, the optimization might call for the airplane to put the throttle forward to store excess airspeed in preparation for coming tailwinds even before the airplane actually enters the tailwind section.

In this study, the initial range location is just at the onset of any winds, at the beginning of the headwind build-up section if $w_{h_{max}} \neq 0$, at the beginning of the downdraft build-up section otherwise. The initial airspeed is the nominal airspeed during takeoff and landing. By comparison, Miele et al.[9,10,16] and Zhao and Bryson[15] assume that the initial range location is at the maximum headwind and that the initial airspeed is at the nominal. In their optimal trajectories, the aircraft has less time to prepare for the microburst tailwinds, and it starts with a deficit in inertial speed. Consequently, for given peak-to-peak winds, their results are more conservative.

A second difference is in the cost functions optimized here and in the referenced studies. Whereas we have minimized a quadratic path-following error, Miele et al.[9,10,16] and Zhao and Bryson[15] minimized the maximum altitude or altitude rate deviations that occurred on the trajectories. The minimum-maximum and quadratic altitude deviation cost functions yield similar optimal trajectories. Zhao and Bryson also investigated a cost function based on final air-relative total energy, which produces very different optimal trajectories.

Solutions to the trajectory optimization problem in Eq. (4) have been obtained by application of the method of Successive Quadratic Programs (SQP) to a discrete-time approximation of the original problem.[17] The discretization is like that of a zero-order hold: it breaks the interval from t_0 to t_f into N equal subintervals, t_k to t_{k+1} for $k = 0...N - 1$; $t_N = t_f$. The controls are held fixed over each subinterval, and the continuous-time inequality constraint is enforced only at the sampling times t_k. The optimization used 70 sampling intervals per microburst encounter.

Parametric Study of Effects of Microburst Variations

Optimal trajectories were computed for various values of the microburst parameters, r_h, $w_{h_{max}}$, r_v, $w_{v_{max}}$, and r_L, for two aircraft types, and for takeoff and approach flight paths. Each parameter combination leads to a different optimal microburst encounter trajectory. Several quantities that characterize the safety of the encounter were extracted from each trajectory and were associated with a particular point in microburst parameter space. Thus, altitude deviation, sink rate, minimum airspeed, and other measures of the trajectories can be portrayed as functions of the microburst parameters, and acceptable limiting values can be expressed in terms of these parameters. The resulting plots present standards against which practical sub-optimal controllers can be judged.

Altitude deviation on landing approach leads to touchdown-point deviation. Vertical velocity deviations can be hazardous on both climb-out and approach. On approach, too high a sink rate at touchdown can cause the aircraft structure to fail. If the ascent rate is too small or negative on climb-out, eventual obstacle impact becomes likely. Airspeed-deviation hazards, as considered here, are those associated with the loss of rotational stability and control through the loss of aerodynamic moments; it could become the critical safety factor in very intense, very short length-scale microbursts. The 1-g stall limit on airspeed is not considered as a safety factor because stall has been directly factored into the trajectory optimizations as a state inequality constraint on angle of attack. Inertial velocity deviations during landing approach are considered because of the necessity to stop the airplane in the given runway length after landing.[21]

This study examines headwind/tailwind (H/T) effects and downdraft (DD) effects separately. The separation reduces the scope of our parametric study to a feasible level. An actual microburst has both horizontal and vertical wind components. For a predominantly H/T microburst encounter (as happens very near the ground), our study of pure H/T encounter gives a good prediction of the trends. Also, for a predominantly DD microburst encounter (as happens further aloft), our study of pure DD encounter gives a good prediction of the trends. For a mixed H/T-DD microburst encounter, superposition of actual trajectory response applies up to a certain level of wind activity, but superposition makes no sense in reference to maximal performance contours. Nevertheless, the performance study of the separate cases gives insight into the quali-

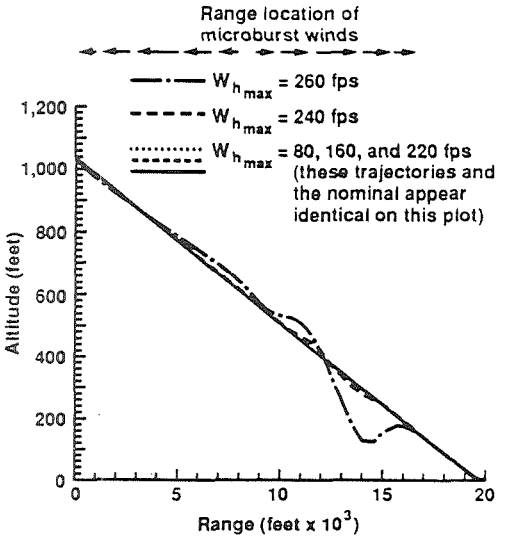


Fig. 3 Optimal JT landing-approach trajectories through five headwind tailwind microbursts of varying intensity (r_h = 16,000 ft).

tative trends in strategy and performance that can be expected in the mixed case.

The dependence on microburst location, r_L, was suppressed for the H/T and DD results. Instead of fixing the value of r_L, r_L takes its worst-case value with respect to any given safety-related quantity. Suppose q_i is the quantity in question and that higher values of the quantity indicate a deterioration of safety, then the value of the quantity used in the H/T cross section would be

$$q_{i_{max}}(r_h, w_{h_{max}}) = \max_{r_L} q_i(r_h, w_{h_{max}}, r_v = 0, w_{v_{max}} = 0, r_L) \quad (5)$$

Variations of r_L do not affect the optimal control actions in the analysis of this paper because flare and pre-takeoff runway roll have not been considered. Microburst location variations only affect the location of the penetration trajectory with respect to the ground. Therefore, maximizations as in Eq. (5) can be carried out over a single optimal trajectory, i.e., by computing a single trajectory for a given r_h and $w_{h_{max}}$ and maximizing q_i over that trajectory. Effects of microburst core location, r_L, also were considered separately.

III. Trajectory Optimization Results

Over a thousand optimal microburst encounters have been computed. The optimal trajectories fall broadly into eight distinct categories, each distinct combination of aircraft type (JT or GA), wind feature (DD or H/T), and flight phase (landing approach or takeoff). In each of these eight categories, 60–210 distinct optimal trajectories were computed at grid points of microburst intensity vs length scale. The microburst length scales range from shorter than the distance traveled during one short period to longer than the distance traveled during one phugoid period. The microburst wind intensities range up to hundreds of ft/s; details for all of these trajectories, can be found in Ref. 17.

The solution of a trajectory optimization problem is a state-space trajectory and its associated control time history. Figures 3 and 4 present examples of aircraft-in-a-microburst optimal trajectories. The JT aircraft is used in these examples, the microbursts are H/T microbursts, the primary cost function component is mean-square altitude deviation, and the flight phase is landing approach. On the altitude vs range plots (Fig. 3), three of the optimal trajectories are indistinguishable from the nominal trajectory. The plots of associated time histories (Fig. 4) show significant deviations from nominal in all cases. These solutions provide good examples of the information contained in an optimal trajectory.

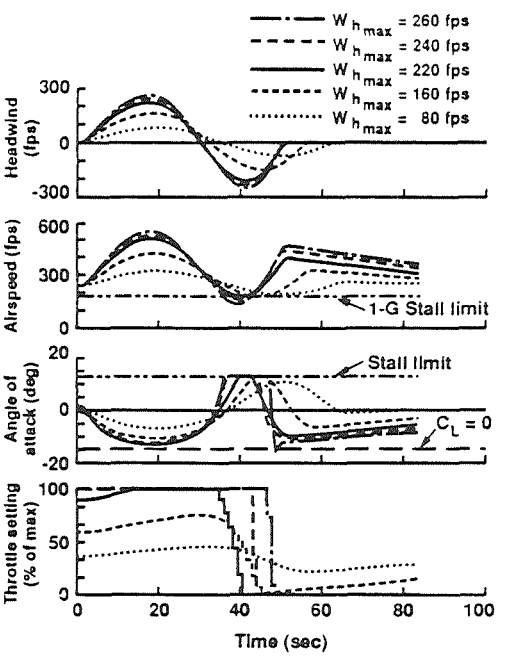


Fig. 4 Time histories associated with five optimal JT landing approaches through headwind/tailwind microbursts of varying intensity (r_h = 16,000 ft).

Insight into the characteristics of good control in a microburst has been a side benefit of this parametric study. In Figs. 3 and 4, unconventional pitch steering maintains lift, thereby maintaining glide path—note the 180 deg phase shift between the airspeed and angle-of-attack time histories on Fig. 4. As the microburst intensity increases, this strategy eventually leads to saturation of the angle of attack at the stall limit. The negative angles of attack indicated on Fig. 4 still correspond to positive lift; note the line indicating the zero-lift angle of attack. Throttle activity increases to avoid airspeed excursions below the 1-g stall limit, and microburst transit occurs more quickly because of the resultant higher inertial velocities. These combined strategies do a good job of maintaining the intended glide path up to a microburst intensity of 220 ft/s (corresponding to a peak-to-peak variation of 440 fps).

The throttle setting increases from the nominal value in the headwind zone in all cases of Figs. 3 and 4 as a consequence of the global knowledge assumed during optimization. In other words, the optimization algorithm benefits from predictive information that would require "look ahead" measurements in the aircraft or "uplink" of ground-based measurements of the wind field. Reference 18 presents a related real-time throttle strategy that does not require predictive capability.

The control strategies used by all of the optimizations of this study have the following common characteristics. They use pitch control in an unconventional manner to maintain lift while using throttle to keep from running out of airspeed—tailwinds steal airspeed directly and downdrafts steal energy, which is equivalent to airspeed. When an extreme microburst causes the airspeed to fall below the 1-g stall limit, despite counteracting throttle activity, the unconventional strategy is continued: the angle of attack is held at the stall limit until the microburst subsides. The net altitude loss due to this maneuver is less than what would be caused by diving to regain airspeed. The current results suggest that α be raised to the stick-shaker limit, where C_L is very high but not yet at its ultimate maximum and where the likelihood of lateral-directional control problems still is quite low. By bleeding off excess airspeed in favor of altitude down to the stick-shaker limit, an aircraft may ride out a severe microburst transient without deviating appreciably from its intended flight path.

Performance as a Function of Microburst Parameters

Landing Approach Performance Through Headwind/ Tailwind Microbursts

Figures 5 and 6 present limiting H/T intensities as functions of length scale for the JT and GA aircraft executing landing approach. The figures also present an approximate curve for the altitude deviation performance. The symbols on the plots refer to points of the contours that have been computed via interpolation with respect to microburst intensity. Extrapolations beyond the symbols indicate that some data exist at the adjacent length scale, but not at high enough microburst intensity to determine the next interpolation point. The microburst length scales at which optimal trajectories have been computed range from 250 to 16,000 ft for the GA aircraft and from 1000 to 64,000 ft for the JT aircraft on a logarithmic grid centered near the respective aircraft's phugoid length scale. The different intensity scales of the two plots reflect the lower ability of the GA aircraft to handle severe microburst winds.

The performance contours for the two aircraft display several similarities. The primary safe-performance measure, the maximum altitude deviation, has contours that are concave upward on both plots for higher values of the maximum deviation. The minimum intensity for a given contour occurs at a microburst length scale between that of the short period mode (GA: 230 ft, JT: 1800 ft) and that of the phugoid mode (GA: 2510 ft, JT: 10,470 ft). The maximum-altitude-deviation contours are close together at long microburst length scales but diverge at short microburst length scales.

At short microburst length scales, the minimum-airspeed limit becomes important. The minimum-airspeed contours reach high microburst intensities for the longer length-scale microbursts. At short microburst length scales, however, the microburst intensities of the minimum-airspeed contours decrease while the intensities of other contours increase, thus making the minimum airspeed limit a critical factor in this regime. As a point of reference, the 1-g stall airspeeds are 180 and 85 ft/s for the JT and GA aircraft, respectively.

At longer microburst length scales, the upper limit on inertial velocity may become a crucial limit. These contours do just the reverse of the minimum-airspeed contours: they are at high microburst intensity for short microburst length scale, but their intensities decrease with increasing microburst length scale. Recalling that a limit to inertial velocity arises from braking-distance/runway limitations, one can compute reasonable values for this limit based on flight manuals such as in Ref. 21.

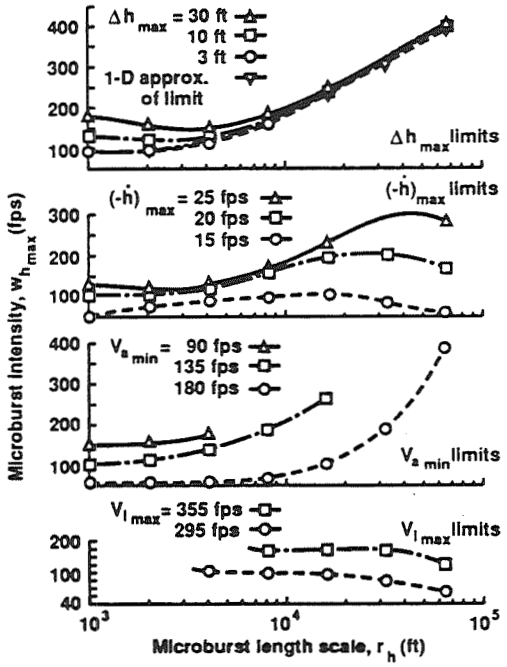


Fig. 5 Performance safety contours for optimal JT landing approach in headwind/tailwind microbursts.

In contrast to the other contours, the maximum-descent-rate contours for the two aircraft (the second graph from the top on each of Figs. 5 and 6) are dissimilar. As a point of reference, the nominal descent rates are 12 and 6 ft/s for the JT and GA aircraft. Contours corresponding to a maximum descent rate that is double the nominal descent rate fall in regions of microburst parameter space where other safety factors would allow a landing approach, e.g., $w_{h_{max}} = 120$ fps, $r_h = 3000$ ft for JT penetration, and $w_{h_{max}} = 60$ fps, $r_h = 1000$ ft for the GA penetration. Descent-rate limits must be considered to determine the safety of landing approach in a microburst.

Some of the trends of these performance safety contours have simple explanations. A dynamic, one-dimensional analysis agrees well with the long length-scale altitude-deviation performance contours. References 17 and 18 explain this analysis, which is a dynamic version of the analysis used to generate an aircraft's power curve from its lift, drag, and thrust characteristics. Stall also plays a role in this analysis, which estimates microburst-induced airspeed variations that occur while tracking the glide path. The airspeed variations are estimated by integrating a differential equation for dV/dr that is derived under the assumptions of 100% throttle setting and perfect glide-path-maintaining pitch control. On the maximum-altitude-deviation contour graphs (top graphs) of Figs. 5 and 6, the curves marked "1-D Approx. of Limit" were generated from this analysis. For long length-scale H/T microbursts, these curves agree well with the actual maximum-altitude-deviation contours.

The maximum inertial-velocity contours can be explained in terms of the general strategy that has been described above. The trajectory optimizations try to avoid excursions of the airspeed below the 1-g stall limit by accelerating the aircraft

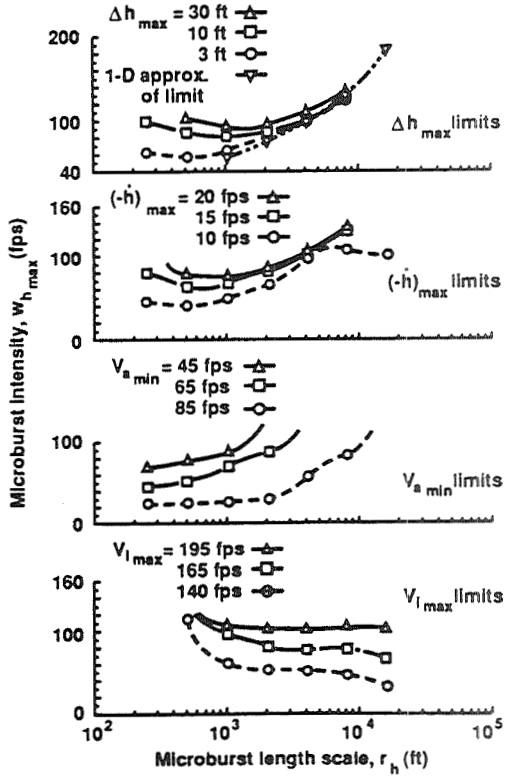


Fig. 6 Performance safety contours for optimal GA landing approach in headwind/tailwind microbursts.

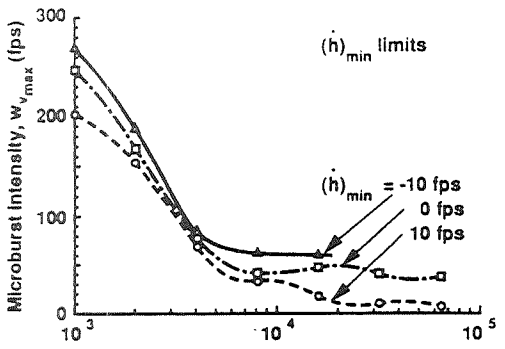


Fig. 7 Performance safety contours for optimal JT climb-out in downdraft microbursts.

with thrust. This strategy becomes more and more effective for longer length-scale microbursts because the thrust has more time to work. Increased inertial velocities in the longer length-scale microbursts go hand in hand with this increased ability to maintain flight path, thus the downward trend of the inertial-velocity contours at long microburst length scales: higher inertial-velocity maxima occur in less severe microbursts. At very long length scales, the throttle is fully capable of maintaining airspeed, and penetration safety is simply a function of the ability of the aircraft to land in a steady tailwind. At microburst length scales shorter than the phugoid length scale, on the other hand, the aircraft cannot change its inertial velocity much due to its large mass; therefore, these contours go to very high microburst intensities and become irrelevant to the question of safety.

For microburst length scales below the phugoid length scale, an aircraft's inertial-velocity cannot vary significantly, and the maximum airspeed loss equals the maximum tailwind. At these short length scales, the microburst intensity of each minimum-airspeed contour is approximately the difference between the nominal airspeed and the contour airspeed.

In summary, the performance safety contours for optimal aircraft landing approach in H/T microbursts display three regimes. In the shortest length-scale regime, shorter than the scale of the aircraft short-period mode, the hazard of upset associated with very low airspeed and loss of aerodynamic control torque is the most restrictive safety criterion. The ability to maintain aircraft altitude is one of the two most important factors in the length-scale regime between the short-period mode and the phugoid mode. The ability to maintain slow descent rate is the other important factor in this regime. The maximum allowable inertial velocity at touchdown can limit the ability to successfully negotiate H/T microbursts in the length-scale regime above the phugoid mode.

Takeoff Performance of a Jet Transport in Downdraft Microbursts

Figure 7 presents results for JT climb-out through downdraft microbursts. The meanings of symbols and the contour plotting methods have been carried over from the previous two figures. For comparison's sake, the same flaps-down aircraft configuration has been used for this study as for the landing approach study, even though this is not the standard takeoff configuration.

In this case, the minimum ascent rate is the critical performance safety factor over the entire length-scale regime that has been considered. At microburst length scales longer than the phugoid length scale, most of the minimum-ascent-rate contours level out and approach a microburst intensity equal to the maximum static ascent rate less the minimum ascent rate associated with the particular contour. At length scales below the phugoid length scale, the microburst intensities of the contours increase with decreasing length scale.

Reference 17 presents the other five sets of intensity/length-scale performance safety contours for optimal microburst encounter. They are similar in structure to the three sets presented here, but shed additional light on the questions of what factors limit safety in what regimes and why they limit it.

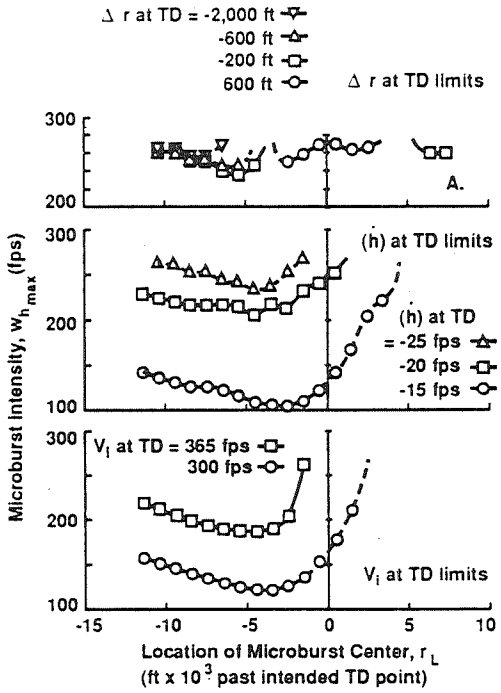


Fig. 8 Performance safety contours for optimal JT landing approach as functions of microburst core location and headwind/tailwind intensity (r_h = 16,000 ft, $w_{v_{max}}$ = 0).

Location-Dependent Safety Contours: A Landing-Approach Example

The gross effects of microburst location were evaluated by shifting the core of H/T profiles and examining changes in touchdown parameters, neglecting flare dynamics. Actual touchdown values would be affected by flare but would follow the trends shown here.

Figure 8 presents three safety-measure contours for the case of optimal JT landing approach through H/T microbursts of length scale r_h = 16,000 ft. The horizontal axes of the graphs give microburst location relative to the intended touchdown point, r_L. Positive location indicates that the microburst center occurs beyond the intended touchdown point (i.e., the aircraft is nominally on the runway). The vertical axes give microburst intensity, $w_{h_{max}}$ ($w_{v_{max}} = 0$). The top graph on the figure presents contours of constant range deviation at touchdown. The + 600 ft contour indicates cases that overshoot the intended touchdown point by 600 ft (land 600-ft long); the negative-valued contours correspond to cases that produce short landings.

The range-deviation-at-touchdown contours indicate that the worst-case microburst location for producing short landings has the center of the outflow occurring 5000 to 9000 ft before the aircraft reaches its intended touchdown point. Given the microburst length scale of 16,000 ft and its sinusoidal H/T variation, the maximum tailwinds occur 1000 to 5000 ft before the intended touchdown point. This makes sense; the tailwinds produce the hazard in H/T microbursts, and if they are to cause an aircraft to land significantly short, they must occur far enough before the runway threshold without being so far from the runway as to allow for recovery before ground impact occurs.

The descent-rate contours and the inertial-velocity contours also have the worst-case location for the microburst center occurring before the aircraft reaches its intended touchdown point. These worst-case locations are 2000 to 5000 ft before the intended touchdown point for the descent rate-at-touchdown criterion and 3000 to 5000 ft before it for the inertial-velocity-at-touchdown criterion. These criteria have the worst-

case microburst occurring closer to the airport than does the range-deviation-at-touchdown criterion.

IV. Conclusions

This paper has presented and discussed the results of a study of the dependence of optimal aircraft penetration trajectories on microburst wind shear characteristics. The method of study plots contours of safety metrics in a parameter space of microburst characteristics. Optimal performance has been studied for two aircraft, a jet transport and a general aviation aircraft, and for two flight phases, climb-out and landing approach.

For optimally controlled aircraft, performance safety on landing approach is limited in medium length-scale microbursts primarily by the aircraft's ability to track the nominal glide path or to maintain low descent rate at touchdown. In long-length-scale microbursts, a critical limit to safe performance on landing approach is the aircraft's maximum touchdown inertial velocity. Safe performance on climb-out is limited primarilty by the aircraft's ability to continue climbing in the microburst. For short length-scale microbursts with a large tailwind, a factor that may limit safe performance is the danger of losing aerodynamic control torque at very low airspeeds. Microburst location also has a strong effect on the safety of an encounter. On landing approach, the most threatening peak tailwind locations for an optimally controlled aircraft occur from zero to one mile before the runway threshold.

The reported results have been achieved by optimal strategies that tightly control altitude via elevator inputs and that use throttle inputs to keep airspeed above the 1g stall limit. These strategies track the intended flight path by taking advantage of transient phenomena and by trading airspeed for altitude down to the stick-shaker limit.

Acknowledgments

This research was supported in part by the Federal Aviation Administration and NASA under Grant NGL 31-001-252.

References

[1]Bach, R. E., Jr., and Wingrove, R. C., "The Analysis of Airline Flight Records for Winds and Performance with Application to the Delta 191 Accident," AIAA Paper 86-2227, 1986.

[2]Rynaski, E. G., and Govindaraj, K. W., "Control Concepts for the Alleviation of Windshears and Gusts," NASA CR-166022, July 1982.

[3]Lehman, J. M., Heffley, R. K., and Clement, W. F., "Simulation and Analysis of Wind Shear Hazard," FAA-RD-78-7, Dec. 1977.

[4]Stengle, R. F., "Optimal Control Laws for Microburst Encounter," *Proceedings of the 15th Congress of the ICAS*, ICAS-86-5.6.3, London, Sept. 1986.

[5]Belcastro, C. M., and Ostroff, A. J., "Total Energy-Rate Feedback for Automatic Glide-Slope Tracking During Wind-Shear Penetration," NASA TP-2412, May 1985.

[6]König, R., Krauspe, P., and Schanzer, G., "Procedures to Improve Flight Safety in Wind Shear Conditions," *Proceedings of the 12th Congress of the ICAS*, ICAS-80-22.3, Munich, Oct. 1980.

[7]Miele, A., Wang, T., and Melvin, W. W., "Optimization and Gamma/Theta Guidance of Flight Trajectories in a Wind Shear," *15th Congress of the ICAS*, ICAS-86-564, London, Sept. 1986.

[8]Miele, A., Wang, T., and Melvin, W. W., "Guidance Strategies for Near-Optimum Take-Off Performance in a Windshear," *Journal of Optimization Theory and Applications*, Vol. 50, No. 1, 1986, pp. 1–47.

[9]Miele, A., Wang, T., and Melvin, W. W., "Optimal Flight Trajectories in the Presence of Windshear, Part 1, Take-Off," AIAA Paper 85-1843, 1985.

[10]Miele, A., Wang, T., and Melvin, W. W., "Optimal Take-Off Trajectories in the Presence of Windshear," *Journal of Optimization Theory and Applications*, Vol. 49, No. 1, 1986, pp. 1–45.

[11]Psiaki, M. L., and Stengel, R. F., "Analysis of Aircraft Control Strategies for Microburst Encounter," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 5, 1985, pp. 553–559.

[12]Psiaki, M. L., and Stengel, R. F., "Optimal Flight Paths Through Microburst Wind Profiles," *Journal of Aircraft*, Vol. 23, No. 8, 1986, pp. 629–635.

[13]Miele, A., Wang, T., and Melvin, W. W., "Optimization and Guidance of Penetration Landing Trajectories in a Windshear," *Proceedings of the American Control Conference*, Inst. of Electrical and Electronics Engineers, Piscataway, NJ, 1988, pp. 1428–1439.

[14]Miele, A., Wang, T., and Melvin, W. W., "Optimization and Acceleration Guidance of Flight Trajectories in a Windshear," *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 4, 1987, pp. 368–377.

[15]Zhao, Y., and Bryson, A. E., "Optimal Paths Through Downbursts," *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Washington, DC, 1989, pp. 1139–1149.

[16]Miele, A., Wang, T., Wang, H., and Melvin, W. W., "Overview of Optimal Trajectories for Flight in a Windshear," AIAA Paper 89-0812, 1989.

[17]Psiaki, M. L., "Control of Flight Through Microburst Wind Shear Using Deterministic Trajectory Optimization," Ph.D. Thesis, Princeton Univ., Dept. of Mechanical and Aerospace Engineering, Princeton, NJ, Oct. 1987.

[18]Psiaki, M. L., and Park, K., "Thrust Laws for Microburst Wind Shear Penetration," *Proceeding of the AIAA Guidance, Navigation, and Control Conference*, AIAA, Washington, DC, 1989, pp. 1127–1138.

[19]Turkel, B. S., and Frost, W., "Pilot-Aircraft System Response to Wind Shear," NASA CR-3342, Nov. 1980.

[20]Shivers, J. P., Fink, M. P., and Ware, G. M., "Full-Scale Wind Tunnel Investigation of the Static Longitudinal and Lateral Characteristics of a Light, Single-Engine, Low-Wing Airplane," NASA TN-D-5857, June 1970.

[21]"Landing Field Length and Speed," Boeing 727-30 Operations Manual, Boeing, Seattle, WA, Nov. 1977.

Intelligent Failure-Tolerant Control

Robert F. Stengel*

Princeton University
Department of Mechanical and Aerospace Engineering
Princeton, NJ 08544

Abstract

An overview of failure-tolerant control is presented, beginning with robust control, progressing through parallel and analytical redundancy, and ending with rule-based systems and artificial neural networks. By design or implementation, failure-tolerant control systems are "intelligent" systems. All failure-tolerant systems require some degree of robustness to protect against catastrophic failure; failure tolerance often can be improved by adaptivity in decision-making and control, as well as by redundancy in measurement and actuation. Reliability, maintainability, and survivability can be enhanced by failure tolerance, although each objective poses different goals for control system design. Artificial intelligence concepts are helpful for integrating and codifying failure-tolerant control systems, not as alternatives but as adjuncts to conventional design methods.

Introduction

Many devices depend on automatic control for satisfactory operation, and while assuring stability and performance with all components functioning properly remains the primary design goal, there is increasing need for controlled systems to continue operating acceptably following failures in either the system to be controlled (the plant) or in the control system itself.[1] A distinction should be made between system *failures*, which occur when components break or misbehave, and system *faults*, which include improper design as well. Our attention is directed at the former, as improper design is a separate issue.

Failure-tolerant control systems can be characterized as robust, reconfigurable, or some combination of the two. A well-designed feedback controller typically reduces the plant's output sensitivity to measurement errors and disturbance inputs; if the plant is lightly damped or unstable, it provides closed-loop stability as well. It is designed assuming some nominal physical structure for the plant, expressed by a mathematical model and a set of parameters. A controlled system that retains satisfactory performance in the presence of variations from this model without changes in the control system's structure or parameters is said to be *robust*. The degree of failure that can be accommodated by a fixed control structure is more restricted than that of a variable control structure. If the structure or parameters can be altered following system failure, the control system is *reconfigurable*.

In the latter case, the control system detects, identifies, and isolates failures, and it modifies control laws to maintain acceptable performance. A system that is failure-tolerant through reconfiguration is both adaptive and redundant. It is *adaptive* in its ability to adjust to off-nominal behavior, as occurs from loss or degradation of sensors, actuators, and power supplies, damage to signal and power transmission channels, or unexpected alteration of the plant's characteristics. It is *redundant* in its ability to overcome lost capabilities with remaining resources. Redundancy can be provided by similar parallel channels for measurement and control, or it may result from flexible logic that synthesizes missing measurements or control forces using operable sensors and actuators, effectively invoking dissimilar parallel channels. A reconfigurable control system must be robust enough to preclude controlled system failure while adaptation is taking place.

While there is much debate as to what constitutes true "machine intelligence," it can be argued that adaptivity and redundancy are attributes of intelligence and, in the same light, that feedback control makes use of information in an intelligent fashion. The issue is not that adaptive, feedback controllers pass the seminal Turing test [1] or possess "consciousness" [2]. It is that they exhibit the "ability involved in calculating, reasoning, perceiving relationships and analogies, learning quickly, storing and retrieving information, ... classifying, generalizing, and adjusting to new situations," [3] at least in a symbolic or quantitative sense. To the extent that symbols and instructions reflect knowledge and decisions, a failure-tolerant, feedback control system can be called intelligent, and that context is adopted here.

Controlled Systems

Attention is focused on the control of continuous-time dynamic systems (or plants) whose motions can be represented by integrals of nonlinear ordinary differential equations,

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t),\mathbf{u}(t),\mathbf{w}(t),\mathbf{p}] \qquad (1$$

where $\mathbf{x}(t)$ is the n-dimensional state, $\mathbf{u}(t)$ is the m-dimensional control, $\mathbf{w}(t)$ is an s-dimensional disturbance, and $\mathbf{p}$ is an l-vector of parameters. The state is observed through the measurement r-vector,

$$\mathbf{z}(t) = \mathbf{h}[\mathbf{x}(t),\mathbf{u}(t),\mathbf{w}(t),\mathbf{n}(t),\mathbf{p}] \qquad (2$$

where $\mathbf{n}(t)$ is an r-dimensional measurement-error vector. Along a nominal trajectory specified by $\mathbf{x}_0(t)$, $\mathbf{u}_0(t)$, $\mathbf{w}_0(t)$, and $\mathbf{n}_0(t)$ for t in (t_0, t_f), perturbations of the state and observation vectors are governed approximately by linear, time-varying equations,

---

* Professor. Senior Member, IEEE.

[1] For the purposes of this paper, the *plant* is defined as a dynamic system containing components that impart distinctive physical properties like mass, inertia, elasticity, forces, and moments. The plant's motion (position and velocity) must be controlled for satisfactory operation. The *control system* is an assemblage of additional components -- motion sensors, force and moment actuators, and computers -- that provide this service. A *controlled system* is a plant plus its control system.

Presented at the *5th IEEE International Symposium on Intelligent Control*, Philadelphia, PA, Sept 1990.

$$\Delta\dot{x}(t) = F(t)\Delta x(t) + G(t)\Delta u(t) + L(t)\Delta w(t) \quad (3$$

$$\Delta z(t) = H_x(t)\Delta x(t) + H_u(t)\Delta u(t) + H_w(t)\Delta w(t) + n(t) \quad (4$$

F, G, L, H_x, H_u, and H_w are conformable Jacobian matrices expressing sensitivities to the perturbation variables. At discrete instants of time, t_k, t_{k+1}, and so on, the state and measurement perturbations can be approximated by

$$\Delta x_{k+1} = \Phi_k \Delta x_k + \Gamma_k \Delta u_k + \Lambda_k \Delta w_k \quad (5$$

$$\Delta z_{k+1} = H_{x_k} \Delta x_k + H_{u_k} \Delta u_k + H_{w_k} \Delta w_k + n_k \quad (6$$

where the subscript "k" indicates evaluation at t_k. Here, Φ, Γ, and Λ have the same dimensions as F, G, and L and are derived from the system's state transition properties (e.g., [4]). These models provide a foundation for the remaining discussion.

Control logic for the nonlinear plant (eq. 1 and 2) typically takes the form of a *dynamic compensator*,

$$\Delta u_k = - C_k \xi_k \quad (7$$

$$\xi_{k+1} = \Psi_k \xi_k + \Theta_k \Delta u_k + K_k[z_k - h(\hat{x}_k, u_k)] \quad (8$$

$$\xi_k \triangleq [\Delta\hat{x}_k^T \ \chi_k^T]^T \quad (9$$

$$u_k = u_{o_k} + \Delta u_k \quad (10$$

$$\hat{x}_k \triangleq x_{o_k} + \Delta\hat{x}_k \quad (11$$

This linear, time-varying structure exemplifies estimation and control functions for discussion purposes, but more complex structures -- particularly nonlinear ones -- may be employed. It is equivalent to a feedback control law (eq. 7) operating on the internal state estimate $\Delta\hat{x}$ contained in the $(n + k)$-dimensional ξ_k (eq. 8). χ_k is a k-vector of compensation components, such as integrals of state elements. The control and estimation gains, C_k and K_k, are selected to provide satisfactory nominal response and may vary in time. Ψ_k and Θ_k normally represent nominal values of Φ_k and Γ_k plus integrating (i.e., accumulating) or filtering operations associated with χ_k. The desired state and corresponding control for the nonlinear plant, x_{o_k} and u_{o_k}, enter as in eq. 10 and 11.

Figure 1 represents an idealized controlled system, with disturbance and noise inputs not shown. While the figure identifies the elements of nominal control system design, it provides little insight about control system components, all of which may fail. Tangible components are needed for measurement and actuation (Fig. 2), the control logic described by eq. 7 to 11 is executed in a computer, these components are enabled by a power supply, and the power supply also is subject to failure. An ancillary issue is that sensors and actuators -- themselves physical systems -- have scale factors, biases, and dynamic characteristics to be considered during failure detection and identification. The simplest means of doing this is to incorporate these characteristics in the plant model (eq. 1 and 2), with estimation and control logic modified accordingly.

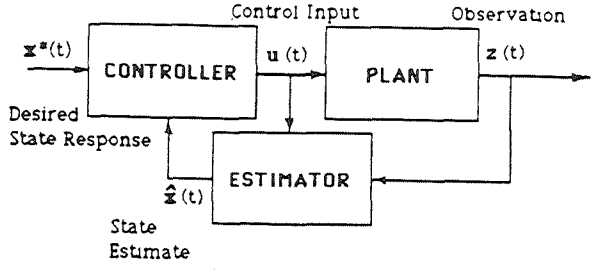


Figure 1. Idealization of a controlled system.

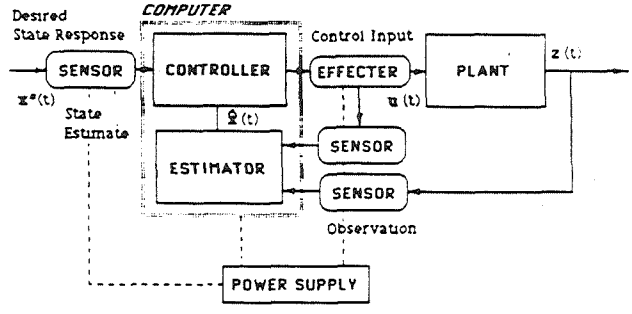


Figure 2. Components of a controlled system.

Objectives and Issues for Failure-Tolerant Control

Failure tolerance may be called upon to improve system reliability, maintainability, and survivability. The requirements for failure tolerance are different in these three cases. *Reliability* deals with the ability to complete a task satisfactorily and with the period of time over which that ability is retained. A control system that allows normal completion of tasks after component failure improves reliability. *Maintainability* concerns the need for repair and the ease with which repairs can be made, with no premium placed on performance. Failure tolerance could increase time between maintenance actions and allow the use of simpler repair procedures. *Survivability* relates to the likelihood of conducting an operation safely (without danger to human operators or the controlled system), whether or not the task is completed. Degraded performance following failure might be permitted, as long as the system can be brought to an acceptable state of rest.

Improving the reliability of individual components clearly helps in all three categories; however, it does not follow that what aids one objective aids another. For example, replacing a single string of control system components by three parallel strings of identical components (plus selection or averaging logic) may improve reliability, but it also increases the likelihood of component failures, degrading maintainability. Conversely, redundancy within line-replaceable units (LRUs) could improve maintainability if it allows LRUs to be changed less often. Adding a separate string of less-capable components may improve survivability without improving reliability while decreasing maintainability.

The principal categories of failure are plant alterations, actuator and sensor failures, computer failure, and power sup-

ply/transmission failure. Actuators, sensors, and other analog components are subject to many failure types, some of which may be subtle but nonetheless damaging: parameter variation, abrupt or random bias shift, abrupt or random scale factor shift, change in saturation limits, drift, open circuit, hardover (or stuck), and noise. Digital computer hardware failures have entirely different characteristics, but it can be argued that they are never subtle, as internal clock rates are high and the loss of coherent output is obvious [5]. Computer software does not fail *per se*, but it is susceptible to programming faults that may surface unexpectedly and that may be hard to detect. Multiple failures can occur, particularly as a consequence of physical damage, and they may be intermittent; hence, reconfiguration logic must do more than just accommodate isolated failures. While not strictly system failures, operator blunders and power transients may produce system states that require prompt response.

Many factors must be considered in designing failure-tolerant controls, including: allowable performance degradation in the failed state, criticality and likelihood of the failure, urgency of response to failure, tradeoffs between correctness and speed of response, normal range of system uncertainty, disturbance environment, component reliability vs. redundancy, maintenance goals (mean-time-between failures, mean-time-to-failure, mean-time-to-repair, maintenance-hours/operation-hours, etc.), size and cost of LRUs, system architecture, limits of manual intervention, and life-cycle costs. Assessing each of these factors requires detailed knowledge of the plant and its control objectives.

Robust Control

Controlled system *robustness* is the ability to maintain satisfactory stability and performance in the presence of parameter variations, which could be due to component failures in either the plant or the control system. All practical controlled systems must possess some degree of robustness against operational parameter variations. Maintaining stability with component failures is a particular challenge when the plant is open-loop-unstable, as control-system failure may mean that the system becomes partially "open-loop." Alternatively, a plant alteration (e.g., the breaking of a stabilizing spring or the loss of an aircraft's stabilizing surface) may force an ordinarily stable system to become unstable. In either case, reconfiguration may offer the only recourse for stable control. It also is possible for an open-loop-stable plant to be destabilized by a feedback controller with failed control loops [6]. This lack of robustness is most likely to occur in high-gain controllers, where open- and closed-loop dynamics are substantially different; robustness recovery typically requires lowering the control gains in systematic fashion [4,6,7]. The inherent stability margins of certain algebraic control laws (e.g., the linear-quadratic (LQ) regulator [4,8-10]) may become vanishingly small when dynamic compensation (e.g., the estimator in a linear-quadratic-Gaussian (LQG) regulator) is added [11]. Restoring the robustness to that of the LQ regulator typically requires increasing estimator gains using the loop-transfer-recovery method [4,12].

Subjective judgments have to be made in assessing the need for robustness and in establishing corresponding control system design criteria, as there is an inevitable tradeoff between robustness and nominal system performance [13]. The designer must know the normal operating ranges and distributions of parameter variations, as well as the specifications for system operability with failed components, else the final design may afford too little robustness for possible parameter variations or too much robustness for satisfactory nominal performance. Robustness traditionally has been assessed deterministically [14]; it is an inherent part of the classical design of single-input/single-output systems, and there are multi-input/multi-output equivalents based on singular-value analysis of various frequency-domain matrices [e.g., 4,10,12,15]. The most critical difficulty in applying these techniques is relating singular-value bounds on return-difference and inverse-return-difference matrices to real parameter variations in the controlled system.

There is increasing interest in statistical alternatives that make full use of knowledge about potential system variations and that work directly with real parameter variations. The *probability of instability* was introduced in [16] and is further described in [17,18]. This method determines the *stochastic robustness* of a linear, time-invariant system by the probability distributions of closed-loop eigenvalues, given the statistics of the variable parameters in the controlled system's dynamic model. The probability that any of these eigenvalues have positive real parts is the scalar measure of robustness, a figure of merit to be minimized by control system design. Extensions to the analysis of performance robustness and of nonlinear, time-varying systems are direct. This approach provides logical connections to reliability analysis of control systems, discussed below.

It is easy to pose unreachable or irrelevant goals for control robustness. Problems that must be addressed in robust control system design include: retaining controllability and observability following component failure, achieving satisfactory off-design performance (including steady-state and tracking response as well as stability), minimizing compromises to on-design performance, and relating robustness criteria to real component failures.

Parallel Redundancy

In principle, tolerance to control system failures can be improved if two or more strings of sensors, actuators, and computers, each separately capable of satisfactory control, are implemented in parallel (Fig. 3). A *voting* scheme is used for redundancy management, comparing control signals to detect and overcome failures. With two identical channels, a comparator can determine whether or not control signals are identical; hence, it can detect a failure but cannot identify which string has failed. Using three identical channels, the control signal with the middle value can be selected (or voted), assuring that a single failed channel never controls the plant. A 2-channel system is considered *fail-safe* because the presence of a failure can be determined, but it is left to additional in-line (or "built-in test") logic to select the unfailed channel for control. The 3-channel system is *fail-operational*, as the task can be completed following a single failure. Systems with four identical control channels are called "fail-op/fail-op" because they can tolerate two failures and still yield nominal performance. In any voting system, it remains for additional logic to declare unselected channels failed. Given the vectorial nature of control, this declaration may be equivocal, as middle values of control-vector elements can be drawn from different strings.

Of course, the voting logic itself has some probability of failure, and a *single-point failure* of a voting component could be catastrophic. Consequently, it may be preferable to let each channel remain independent through the application of control force, letting *force averaging* mediate failures. If control outputs are averaged, small variations among the parallel channels tend to cancel, and the net output is smooth; however, a runaway failure can bias the net signal away from its desired value. Voting and isolation of failed channels then can be carried out as an auxiliary process whose own failure would not disable the entire system. Once a failed channel has been disengaged, the total available

control force is reduced, changing the performance characteristics of the controlled system.

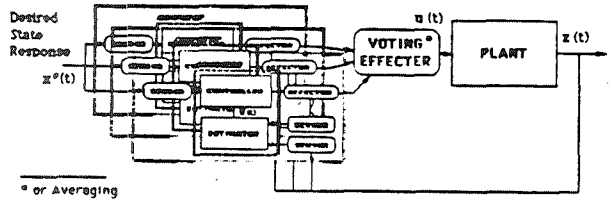


Figure 3. A triply redundant controlled system.

For perfect output voting of M identical parallel channels each with N serial components, the failure probability P_f of the overall control system is,

$$P_f = \prod_{j=1}^{M}\left[1 - \prod_{i=1}^{N} e^{-\lambda_i t}\right] \approx [(\lambda_s + \lambda_c + \lambda_a)(t_f - t_0)]^M \triangleq (1 - R)^M \qquad (12$$

Sensor, computer, and actuator failure rates[1] are λ_s, λ_c, and λ_a (assumed to be small and uncorrelated), $(t_f - t_0)$ is the mission duration, and R is the single-string reliability [19,20]. If the components can be *cross-strapped* perfectly (i.e., if a failed component from one string can be connected to an unfailed string), the overall probability of failure is reduced to

$$P_f = 1 - \prod_{j=1}^{M}\left[1 - \prod_{i=1}^{N}(1 - e^{-\lambda_i t})\right] \approx (\lambda_s{}^M + \lambda_c{}^M + \lambda_a{}^M)(t_f - t_0)^M \qquad (13$$

Unfortunately, failures cannot be detected perfectly, and cross-strapping itself is subject to failure. The probability of detecting, isolating, and recovering from a failure -- called *coverage* -- is a more meaningful measure than P_f. For a 3-channel control system with output voting alone, the coverage C [21], or net reliability, is

$$C = R^3 + 3R^2(1 - R)P_{r_1} + 3(1 - R)^2 R P_{r_2} \qquad (14$$

where P_{r_1} is the probability of recovering from the first failure and P_{r_2} is the probability of recovering from a second failure. These probabilities are not necessarily the same, as different processes may be used for failure detection: voting for the first failure, in-line detection for the second. Unless the recovery probabilities are very nearly one, the maximum benefits of redundancy will not be realized.

Problems encountered in implementing parallel redundancy include: selection logic, nuisance trips, generic failures, reliability of voting/selection units, control force contention, cross-strapping, increased cost and maintenance, number of operating channels required for dispatch, and connectors. Failure-detection logic must be sensitive to failures yet insensitive to small operational errors, including those due to non-colocation of sensors or actuators. Nuisance trips (false indications of failure) must be minimized to assure that useful resources are kept on-line and missions are not aborted prematurely. Redundancy does not preclude identical damage to parallel systems, especially when they are located in close proximity. Cross-strapping implies complex, "intelligent" interconnections; however, if it is not implemented, a single component failure brings down an entire control string. Voting can be done in all operating control computers, but arbitration is required when these computers disagree. For the ideal parallel system, the probability P_c that some component will fail is,

$$P_c \approx M[(\lambda_s + \lambda_c + \lambda_a)(t_f - t_0)] \qquad (15$$

so the likelihood of component failure is increased by redundancy. It is necessary to establish rules for dispatching the controlled system: if one control string is not operational but the others are, should the process be initiated? For a manufacturing system, the answer might be "yes," while for a transport aircraft, it might be "no." A non-trivial aspect of redundant control is the need for more electrical connectors, the components most likely to cause trouble!

One insidious problem associated with parallel redundancy is the lack of controllability of internal state components [22]. Consider the dual-redundant controlled system of Fig. 4, where the individual control outputs are averaged by $M_1 = M_2$, and $F_1 = F_2$, $G_1 = G_2$, and $N_1 = N_2$. The dynamic equations can be expressed as

$$\begin{bmatrix}\dot{x}_A\\ \dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}F_A & G_A M_1 & G_A M_1\\ G_1 N_1 & F_1 & 0\\ G_1 N_1 & 0 & F_1\end{bmatrix}\begin{bmatrix}x_A\\ x_1\\ x_2\end{bmatrix} + \begin{bmatrix}0\\ G_1\\ G_1\end{bmatrix} u_D \qquad (16$$

The controllability matrix C of this system is

$$\mathbb{C} = \begin{bmatrix}0 & 2G_A M_1 G_1 & 2(F_A G_A M_1 + G_A M_1 F_1)G_1 & \cdots\\ G_1 & F_1 G_1 & (2N_1 G_A M_1 + F_1{}^2)G_1 & \cdots\\ G_1 & F_1 G_1 & (2N_1 G_A M_1 + F_1{}^2)G_1 & \cdots\end{bmatrix} \qquad (17$$

Complete controllability requires that $\mathbb{C}$ be of maximal rank; however, that is not possible because the bottom two rows are repeated. In other words, the compensator state elements are not controllable. If the corresponding modes are stable, then small variations between the two controllers tend to decay; however, if the modes are unstable or neutrally stable (as in the case of integral compensation), uncontrollable drift can occur, leading to divergent control outputs, nuisance trips, and possible isolation of otherwise operable channels.

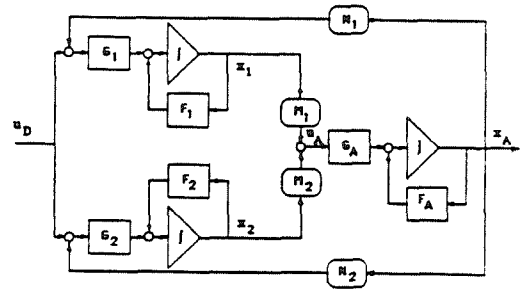


Figure 4. Model of a dual-redundant controller.

[1] In the present context, "sensor" implies the entire suite of sensors needed for control, and "computer" and "actuator" are defined similarly.

If there are sufficient cues to warn a human operator of system failure and plausible failure effects are slow enough to allow manual intervention, many of the benefits of parallel redundancy can be obtained by operating with a single control string, keeping an idle backup control string at the ready. The backup system can be similar or dissimilar to the primary system; however, if it is less capable, ability to perform the task will be degraded.

Parallel redundancy can protect against control-system component failures, but it does not address failures of plant components. Analytical redundancy provides a capability to improve tolerance to failures of both types. It does this with fewer additional components, flexible cross-strapping, and increased computation; as a consequence, there is greater reliance on the control computer, producing even greater need for computer reliability.

Analytical Redundancy

The principal functions of analytical redundancy are failure detection (through built-in-test alarms or off-nominal operation), failure identification (recognition of which components are failed), and control-system reconfiguration (adaptation to sensed or estimated failures). Detection and identification may be combined in built-in test functions. Although in-line monitors provide direct and rapid response to specific failures, it is impossible to provide full coverage of all failures by specialized instrumentation (which itself is subject to failure). A practical failure detection, identification and reconfiguration (FDIR) solution can be found in the control computer's ability to compare expected response to actual response, inferring component failures from the differences and changing either the structure or the parameters of the control system as a consequence.

Failure detection is exemplified by the generalized likelihood ratio test (Fig. 5) [23], which uses a Kalman-filter-like recursive equation to sense discrepancies in system response. The test compares the probability of the estimator's actual measurement residual [z - h(•)] with its expected value, detecting a jump that can be related to failure. It is very sensitive to off-nominal performance and is easy to implement; however, the test does not produce a tight indication of the failed element, and modeling errors can hamper detection [24].

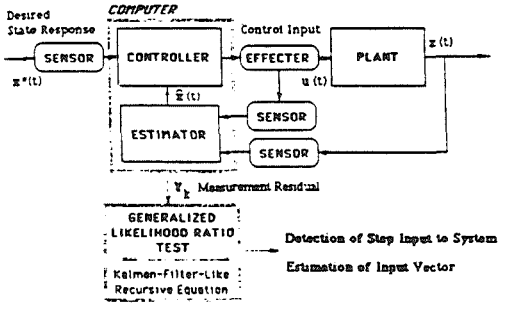


Figure 5. Failure detection: generalized likelihood ratio test.

Failure identification may require a more specific test, such as multiple-model hypothesis testing (Fig. 6) [25,26]. Each failure hypothesis (including that of no failure) is modeled in a Kalman filter, and the most likely hypothesis (based on probability estimates [4]) indicates the failure state. This is a computationally intensive technique, as not only the failed device must be hypothesized but the type, magnitude, and (if taken to the extreme) even the time of the failure must be modeled as well.

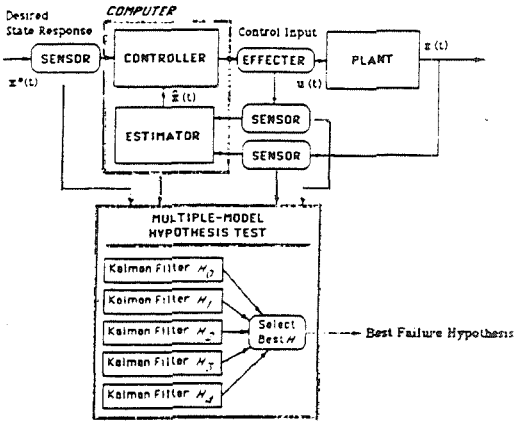


Figure 6. Failure identification: multiple-model hypothesis test.

Consider a modified form of the generic control structure:

$$\Delta u_k = - S_c C_k[\xi_k] + b_c \qquad (18$$

$$\xi_{k+1} = \Psi_k \xi_k + \Theta_k \Delta u_k + K_k[S_s z_k - h(\hat{x}_k, u_k) + b_s] \qquad (19$$

S_s and S_c are scale-factor matrices on the measurements and control, and b_s and b_c are bias vectors. Within this framework, we can identify the elements of the control system that need to be modified following various failures, as in Table 1. If the plant is altered, it may be necessary to change the internal model (Ψ, Θ), as well as the estimation and control gains (K, C), and so on for the remaining failure types. Precise failure identification is an important antecedent of control reconfiguration. Both "hard" (fast) and "soft" (slow) failures must be expected, and logic must accommodate command inputs (set-point transients), disturbances, and measurement noise [27].

Table 1
Failure Types and Related Control-Law Parameters

Failure	Parameter
Plant Alteration	Ψ, Θ, K, C
Actuator Failure	u, Θ, C
Sensor Failure	z, h, K
Bias Shift	b_s or b_c
Scale Factor Shift	S_s or S_c
Saturation Limit Change	K or C
Drift	b_s or b_c
Open Circuit	u, Θ, C, and/or z, h, K
Hardover/Stuck	Open Circuit, plus b_s and/or b_c
Noise	K

Reconfiguration attempts to retain nominal stability and performance characteristics. At a minimum, this requires that on-design controllability and observability (e.g., [4]) be preserved. There is a tradeoff between speed of reconfiguration, computer storage requirements, and flexibility of reaction. Controller structures and parameters for all conceivable failed states can be generated off-line and stored for eventual use; however, this ap-

proach could require an enormous memory. Conversely, on-line design requires minimal storage and (in principle) can adjust to unanticipated failures, but design algorithms must be executed and their results accepted soon enough to provide sufficient failure tolerance. With failed sensors, reconstruction of missing measurements may increase state-estimate errors; with failed effecters, the remaining actuators may have to operate with larger displacements and rates [28]. If the plant is open-loop-unstable, higher control activity combined with existing control-saturation limits may reduce the state space within which closed-loop stability can be assured [29,30].

Artificial Intelligence

Control theory and artificial intelligence both strive to harness mathematics and logic for practical problem solving, but control theory finds its origins in dynamics and electronics, while artificial intelligence springs from biology, psychology, and computer science. Failure-tolerant control systems can benefit from blending these perspectives. Two approaches have been followed in the field of artificial intelligence. *Artificial neural networks* are motivated by input-output and learning properties of living neural networks, although in application the network becomes an abstraction that may bear little resemblance to its biological namesake. *Expert systems* mimic the intelligent functions of an expert or group of experts. Initially, artificial neural networks appeared impractical because computers of the day were too slow and massive, and methods for training neural networks (e.g., *perceptrons* and *adalines*) were thought to be unworkable [31,32]. In the intervening years, the expert system approach proved to be quite achievable; hence, it received major emphasis in both theoretical development and applications. New insights about learning and improved electronics have restored interest in neural networks.

Expert Systems

Expert systems are computer programs that use heuristic relationships and facts as human experts do. The tasks and requirements of such systems (Table 2 [33]) are important for reconfigurable control systems, but there is a need to go beyond the usual limitations of static expert systems. Interpretation, diagnosis, monitoring, prediction, planning, and design must be cyclical, dynamic processes that can reconfigure the control system in "real time" (i.e., with negligible delay).

Table 2
Functions of an Expert System

Task	Requirements
Interpretation	Correct, consistent, complete analysis of data
Diagnosis	Fault finding
Monitoring	Recognition of alarm conditions
Prediction	Reasoning about time, forecasting the future
Planning	Defining actions to achieve goals
Design	Creating objects that satisfy requirements

The expert system offers a useful formalism for failure-tolerant control because it can consider diverse data sources and subproblem abstractions. The expert system can combine qualitative and quantitative reasoning, heuristics and statistics [34]. Failure indicators may be continuous variables generated by measurements or estimators, or they may be discrete variables from in-line monitors or discrete-event models. Indicators are the outputs of *productions*, routines with unique input-output characteristics that produce goal conditions from initial conditions. Hence, the expert system can be implemented as a *production system* or a *rule-based system* consisting of a *data base*, a *rule base*, and a rule interpreter (or *inference engine*) [35]. A production system generates actions predicated on the data base, which contains measurements as well as stored data or operator inputs.

A rule-based failure-tolerant control system contains FDIR logic in expert-system format (Fig. 7). The expert system is an adjunct to the nominal control structure, which remains the most efficient means of effecting precise control. From the control perspective, the expert system performs its decision-making tasks in a concentric *outer loop*; from the expert-system perspective, control activity is a *side effect* that supports decision making.

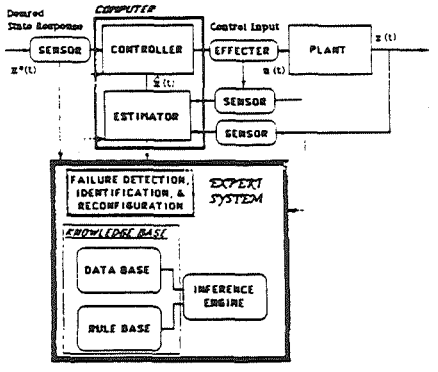


Figure 7. Expert-system approach to analytical redundancy.

An expert system performs *deduction* using *knowledge* and *beliefs* expressed as parameters and rules (Fig. 8). *Parameters* have values that either are external to the expert system or are set by rules. An "IF-THEN" *rule* evaluates a *premise* by testing values of one or more parameters related by logical "ANDs" or "ORs," as appropriate, and it specifies an *action* that set values of one or more parameters. The rule base contains all the rules of the expert system, and the inference engine performs its function by searching the rule base. Given a set of premises (evidence of the current state), the logical outcome of these premises is found by a data-driven search (*forward chaining*) through the rules. Given a desired or unknown parameter value, the premises needed to support the fixed or free value are identified by a goal-directed search (*backward chaining*) through the rules. Querying (or firing) a rule when searching in either direction may invoke procedures that produce parameter values as side effects.

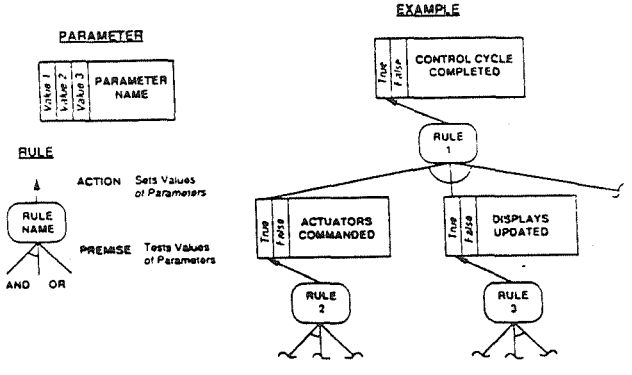


Figure 8. Graphical representation of expert system knowledge.

Both search directions are used in a rule-based control system [36]. Backward chaining drives the entire process by demanding that a parameter such as *CONTROL CYCLE COMPLETED* have a value of *true*. The inference engine works back through the rules to identify other parameters that allow this and, where necessary, triggers side effects like estimation and control to set these parameters to the needed values. Backward chaining also is invoked to learn the value of *ABNORMAL BEHAVIOR DETECTED*, be it *true* or *false*. Conversely, forward chaining indicates what actions can be taken as a consequence of the current state. If *SENSOR MEASUREMENTS REASONABLE* is *true*, and *ALARM DETECTED* is *false*, then failure identification and reconfiguration side effects can be skipped on the current cycle.

Rules and parameters can be represented as *objects* or *frames* that have identities and attributes. For example, a rule can be expressed as the ordered list (*NAME, STATUS, PREMISE, ACTION, ACTION PARAMETERS, PREMISE PARAMETERS, TRANSLATION*), while a parameter may take the form (*NAME, USING RULES, UPDATING RULES, ALLOWABLE VALUES, TRANSLATION*). Most of these attributes are self-explanatory. *STATUS* indicates the state of the rule, such as "not been tested," "being tested," "tested, and premise is *true*," "tested, and premise is *false*," or "tested, and premise is *unknown*." *ALLOWABLE VALUES* provides a mechanism for detecting false logic. *TRANSLATION* provides a natural-language explanation for display to the operator. Specific rules and parameters are represented by lists in which names and attributes are replaced by their values. The attribute lists contain not only values and logic but additional information for the inference engine. This information can be used to compile *parameter-rule-association lists* that speed execution [37].

Frames provide useful parameter structures for related productions, such as analyzing the origin of one or more failures in a complex, connected system [38]. The *dependency graph* of Fig. 9 showing relationships between actuators and their power supplies can be represented by the random-order list ((*OBJECT Name*) (*ATTRIBUTE*$_1$ *Value*$_1$) (*ATTRIBUTE*$_2$ *Value*$_2$) (...)), a more flexible form than the previous structure. In this application, the (*ATTRIBUTE Value*) lists are (*A-KIND-OF Device*), (*ANTERIOR <-OR> Device<s>*), (*POSTERIOR<-OR> Device<s>*), (*CRITICALITY Number*), and (*UNITS Number*). Frames possess an inheritance property; thus the object ((*OBJECT Pivoting Actuator*) (*A-KIND-OF Actuator*) (*ANTERIOR Hydro-Reservoir*) (*POSTERIOR-OR* (*Swashplate Pitching-Link*))) lays claim to the properties of ((*OBJECT Actuator*) (*A-KIND-OF Hydraulic Device*) (*UNITS* (*I* 2))). A two-step process estimates the failure state. In local failure analysis, forward chaining assesses the impact of known malfunctioning units, and backward chaining finds possible causes of the anomalies. In global failure analysis, local failure models are combined, an inclusion property prunes redundant models, and a heuristic evaluation based on criticality, reliability, extensiveness, implications, level of backtracking, and severity produces a list of most likely failure models.

Expert systems process lists, so it is not surprising that LISP (LISt Processing) is the computer language of choice for preliminary development. However, LISP is not a fast, efficient language and is ill-suited to real-time applications. Moreover, a rule-based control system uses numerical algorithms that are most effectively coded in languages like Pascal, C, or FORTRAN. Consequently, *knowledge-base translation* from LISP to a procedural language is a useful (if not necessary) adjunct of rule-based control system design. This not only speeds program execution, it integrates control and decision-making processes, revealing new possibilities for incorporating diagnostic procedures in failure detection and identification [39].

Rule-based control systems must make decisions under uncertainty, and they can do so either by invoking *certainty-equivalent logic*, which is analogous to a well-known concept of stochastic optimal control, or by uncertainty management in the decison-making process. In the LQG regulator, uncertainties due

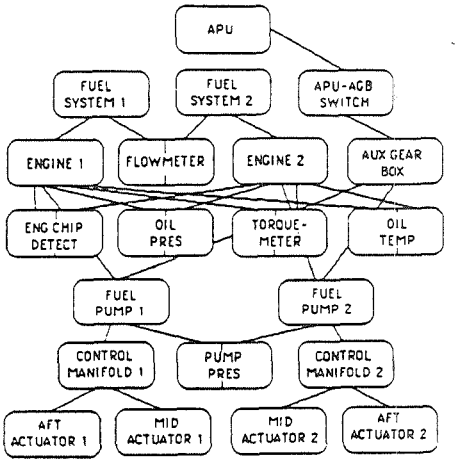


Figure 9. Dependency graph of a hydraulic control system.

to disturbances and measurement error are processed in the estimator, and the feedback control law operates on the state estimate as if it were the actual state [4]. The optimal control gains for the stochastic and deterministic cases are identical. Because the rule-based control system described above makes its best estimates of the failure state in the control logic, the expert system controlling FDIR can treat these results deterministically, realizing little or no improvement from further uncertainty processing. If inner-loop estimation is decidedly sub-optimal, uncertainty management can help, using probability theory, Dempster-Shafer theory, possibility theory, certainty factors, or the theory of endorsements [40]. Bayesian belief networks [41], which propagate event probabilities up and down a causal tree, have particular appeal for failure-tolerant control and are being applied in a related program to assist aircraft crews in avoiding hazards [42].

Teaching the expert system the rules and parameters that generalize the decision-making process from specific knowledge (the process of *induction*) is another concern. Here, we have followed two approaches at Princeton. The first is called *rule recruitment* [43], and it involves the manipulation of "dormant rules" (or *rule templates*). Each template possesses a fixed premise-action structure and refers to parameters through "pointers." Rules are constructed and incorporated in the rule base by defining links and modifying parameter-rule-association lists. Learning is based on repeated simulations of the controlled system with alternative failure scenarios. Learned parameter values then can be defined as "fuzzy functions" [44] contained in rule premises. The second approach [45] has two parts: analysis of variance identifies the factors that make statistically significant contributions to the decision metric, and the "ID3" algorithm [46] extracts rules from the training set by inductive inference. The rules take the form of *decision trees* that predict the performance of alternative strategies.

Expert systems are incorporated in the FDIR process to accommodate declarative functions, leaving reflexive functions to the estimation and control laws [43]. *Declarative action* requires a deep understanding of cause and possible effect. *Reflexive action* is automatic, quickly relating stimulus to response. Both are needed in intelligent failure-tolerant control.

Artificial Neural Networks

Artificial neural networks consist of nodes that simulate the neurons and weighting factors that simulate the synapses of a living nervous system. They are good candidates for performing a variety of reflexive functions in failure-tolerant control systems because they are potentially very fast (in parallel hardware implementation), they are intrinsically nonlinear, they can address problems of high dimension, and they can learn from experience. From the biological analogy, the neurons are modeled as switching functions that take just two discrete values; however, "switching" is softened to "saturation" in common usage, not only to facilitate learning of the synaptic weights but to admit the modeling of continuous functions.

The neural networks receiving most current attention are memoryless expressions that approximate functions of the form

$$y = f(x) \qquad (20$$

where x and y are input and output vectors and $f(\cdot)$ is the (possibly unknown) relationship between them. Neural networks can be considered *generalized spline functions* that identify efficient input-output mappings from observations [47,48]. Rather than approximating eq. 20 by a series, an N-layer neural network (Fig. 10) represents the function by recursive operations,

$$x^{(k)} = s^{(k)}[W^{(k-1)}x^{(k-1)}] \triangleq s^{(k)}[\eta^{(k)}] \text{ , } k = 1 \text{ to } N \text{ .} \qquad (21$$

where $y = x^{(N)}$ and $x = x^{(0)}$. $W^{(k-1)}$ is a matrix of weighting factors determined by the learning process, and $s^{(k)}[\cdot]$ is an activation-function vector whose elements are scalar, nonlinear functions $\sigma_i(\eta_i)$ appearing at each network node:

$$s^{(k)}[\eta^{(k)}] = [\sigma_1(\eta_1^{(k)}) \ldots \sigma_n(\eta_n^{(k)})]^T \qquad (22$$

One of the inputs to each layer may be a unity threshold element that biases the activation-function output.

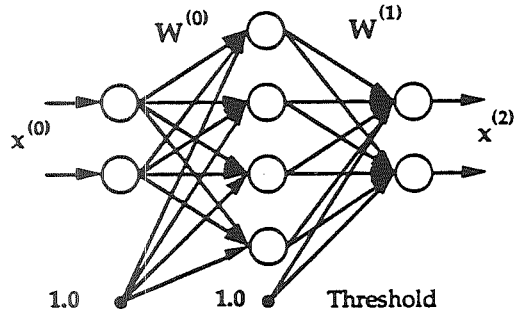


Figure 10. Backpropagation Feed-Forward Neural Network.

The *sigmoid* is commonly used as the artificial neuron. It is a saturating function defined variously as $\sigma(\eta) = 1/(1 + e^{-\eta})$ for output in (0,1) or $\sigma(\eta) = (1 - e^{-2\eta})/(1 + e^{-2\eta}) = \tanh \eta$ for output in (-1,1). Recent results indicate that any continuous mapping can be approximated arbitrarily closely with sigmoidal networks containing a single hidden layer (N = 2) [49,50]. It appears that certain symmetric functions, such as the *radial basis function* ($\sigma(\eta) = e^{-\eta^2}$) or the derivative of the sigmoid have even better convergence properties. Backpropagation learning algorithms for the elements of $W^{(k)}$ typically involve a gradient search (e.g., [51]), although learning speed and accuracy are improved using the extended Kalman filter [52]. The Cerebellar Model Articulation Controller (CMAC) is an alternative neural network formulation with somewhat different properties but similar promise for application in control systems [53].

Equation 20 can represent many functions of importance in dynamics and control. For example, defining x as [$\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{w}(t)$, $\mathbf{p}$], eq. 1 takes that form; together with the implied integration, neural networks can model plant dynamics. A discrete-time model of truck dynamics is demonstrated in [54], and a means of using neural networks in system identification is described in [55]. With x = [$\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{w}(t)$, $\mathbf{n}(t)$, $\mathbf{p}$], the measurement vector (eq. 2) also could be represented. There is little advantage to expressing a linear control law like eq. 7 by a neural network; however, if the control gain matrix C is scheduled by operating point or time, that relationship could be modeled by a neural network. If a nonlinear control function such as $\mathbf{u} = \mathbf{u}(\mathbf{x}, \mathbf{x}_{desired}, t)$ is generated by optimization, nonlinear inversion, or model matching, it can be represented by a neural network [e.g., 54,56,57]. Consequently, neural networks can be incorporated in most of the control and FDIR techniques mentioned above.

Neural networks can be applied to failure detection and identification by mapping data patterns (or *feature vectors*) associated with failures onto detector/identification vectors (e.g., [58,59]). The network is trained to detect failure with the scalar output "1" corresponding to all failure patterns and "0" corresponding to no failure. During operation, a failure is indicated when the output exceeds some threshold near "1." To identify specific failures, the output is a vector, with a training value of "1" in the i^{th} element corresponding to the i^{th} failure mode. For M failure modes, either M neural networks with scalar outputs are employed or a single neural network with M-vector output is used; there are evident tradeoffs related to efficiency, correlation, and so on. The data patterns associated with each failure may require *feature extraction*, pre-processing that transforms the input time series into a feature vector. In [59], this was done by computing 24 Fourier coefficients of the input signal in a moving temporal window. When assessing FDI logic, feature extraction must be considered part of the neural-network computation.

Of course, not all of the suggested neural nets can learn on-line, as a training set must contain desired outputs as well as available inputs. In the cited examples, [54] and [56,58,59] use *off-line learning*, while [55] and [57] allow *on-line learning*. Reference 60 trains a neural network using an expert system that previously learned the desired control strategy. Once an initially trained system is on-line, the "off-line" training process could be executed in parallel with the on-line operation, allowing updates to be made. If the control process that generates on-line training data performs satisfactory control, the need for the neural network must be questioned. The goal should be to provide satisfactory failure tolerance with minimum hardware and software.

Neural networks intended to detect failures would learn little from monitoring normally operating plants. In any case, the neural-network learning rate is slow, probably too slow to expect neural networks of appreciable dimension to adapt to system failures in real time. Hence, the immediate application of neural networks in failure-tolerant control systems is to approximating nonlinear functions used by the FDIR methods introduced earlier. On-line learning can fine-tune this logic over a period of time.

Conclusions

Intelligent failure-tolerant control can improve the operating characteristics of systems. These improvements depend upon a good knowledge of the plant, reliable control elements, and sufficient observability and controllability following failures. Inherent robustness, the ability to accommodate failures without adaptation, is a highly desirable attribute, but it may not be sufficient to contain all system failures. Because split-second decision and reconfiguration may be required, a high degree of pre-training should be assumed; even intelligent systems cannot learn about new failure modes and respond to them properly at the same time (except by chance). Failure-tolerant systems must be able to distinguish between failures, disturbances, and modeling errors, responding to each in the proper way. Probability theory provides an underlying theme that unifies failure-tolerant design, from the probability of instability of robust systems, through the probability of failure of redundant systems, to the probability of correct FDIR response in analytical redundancy. Artificial intelligence is a useful adjunct to parallel and analytical redundancy, as expert systems and artificial neural networks offer new alternatives for both declarative and reflexive response to system failures.

Acknowledgment

This paper expands on a talk given at the General Motors Research Laboratories, Warren, MI, October, 1988. The present work has been supported by the Federal Aviation Administration and the National Aeronautics and Space Administration under Grant No. NGL 31-001-252 and by the Army Research Office under Contract No. DAAL03-89-K-0092.

References

1. Turing, A., "Computing Machinery and Intelligence," *Mind*, Vol. 59, No. 536, Oct 1950, pp. 433-460.

2. Searle, J.R., "Is the Brain's Mind a Computer Program?," *Scientific American*, Vol. 262, No. 1, Jan 1990, pp. 26-31,

3. Harris, W.H., and Levey, J.S., *New Columbia Desk Encyclopedia*, Columbia University Press, New York, 1975.

4. Stengel, R.F., *STOCHASTIC OPTIMAL CONTROL: Theory and Application*, J.Wiley & Sons, New York, 1986.

5. Osder, S., "The DC-9-80 Digital Flight Guidance System's Monitoring Techniques," AIAA Paper No. 79-1704, *Proceedings of the 1979 AIAA Guidance and Control Conference*, Boulder, CO, Aug 1979, pp. 64-79.

6. Soroka, E., and Shaked, U., "On the Robustness of LQ Regulators," *IEEE Transactions on Automatic Control*, Vol. AC-29, No.7, July 1984, pp. 664-665.

7. Wong, P.K., and Athans, M., "Closed-Loop Structural Stability for Linear-Quadratic Optimal Systems," *IEEE Transactions on Automatic Control*, Vol. AC-22, No.1, Feb 1977, pp. 94-99.

8. Kalman, R.E., "When is a Linear Control System Optimal?" *Transactions of the ASME, Journal of Basic Engineering*, Vol. 86, Mar 1964, pp. 51-60.

9. Anderson, B.D.O., and Moore, J.B., *Linear Optimal Control*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

10. Lehtomaki, N.A., Sandell, N.R., and Athans, M., "Robustness Results in Linear-Quadratic-Gaussian Based Multivariable Control Designs," *IEEE Transactions on Automatic Control*, Vol. AC-26, No.1, Feb 1981, pp. 75-93.

11. Doyle, J.C., "Guaranteed Margins for LQG Regulators," *IEEE Transactions on Automatic Control*, Vol. AC-23, No.4, Aug 1978, pp. 756-757.

12. Doyle, J.C., and Stein, G., "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," *IEEE Transactions on Automatic Control*, Vol. AC-26, No.1, Feb 1981, pp. 4-16.

13. Safonov, M.G., Laub, A.J., and Hartmann, G.L., "Feedback Properties of Multivariable Systems: The Role and Use of the Return Difference Matrix," *IEEE Transactions on Automatic Control*, Vol. AC-26, No.1, Feb 1981, pp. 47-65.

14. Dorato, P., ed., *Robust Control*, IEEE Press, New York, 1987.

15. Doyle, J. C., "Analysis of Feedback Systems with Structured Uncertainties," *IEE Proceedings*, Vol. 129, Part D, No. 6, pp. 242-250, Nov 1982.

16. Stengel, R.F., "Some Effects of Parameter Variations on the Lateral-Directional Stability of Aircraft," *Journal of Guidance and Control*, Vol. 3, No. 2, Apr 1980, pp. 124-131.

17. Stengel, R.F., and Ryan, L., "Stochastic Robustness of Linear-Time-Invariant Control Systems," to appear in *IEEE Transactions on Automatic Control*.

18. Ray, L.R., and Stengel, R.F., "Stochastic Stability and Performance Robustness of Linear Multivariable Systems," *Proceedings of the 1990 American Control Conference*, San Diego, May 1990, pp. 462-467.

19. Miller, I., and Freund, J.E., *Probability and Statistics for Engineers*, Prentice-Hall, Englewood Cliffs, 1977.

20. Westermeier, T.F., "Redundancy Management of Digital Fly-By-Wire Systems," *Proceedings of the 1977 Joint Automatic Control Conference*, San Francisco, June 1977, pp. 272-277.

21. Bouricius, W.G., *et al*, "Reliability Modeling for Failure-Tolerant Computers," *IEEE Transactions on Computers*, Vol. C-20, No. 11, Nov 1971, pp. 1306-1311.

22. Stengel, R.F., "Some Effects of Bias Errors in Redundant Flight Control Systems," *Journal of Aircraft*, Vol. 10, No. 3, Mar 1973, pp. 150-156.

23. Willsky, A., and Jones, H.L., "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-21, No. 1, Feb 1976, pp. 108-112.

24. Horak, D.T., "Failure Detection in Dynamic Systems with Modeling Errors," *Journal of Guidance, Control, and Dynamics*, Vol. 11, No. 6, Nov-Dec 1988, pp. 508-516.

25. Willsky, A., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, Vol. 12, No. 6, Nov 1976, pp. 601-611.

26. Basseville, M., "Detecting Changes in Signals and Systems -- A Survey," *Automatica*, Vol. 24, No. 3, May 1988, pp.309-326.

27. Merrill, W.C., DeLaat, J.C., and Bruton, W.M., "Advanced Detection, Isolation, and Accommodation of Sensor Failures -- Real-Time Evaluation," *Journal of Guidance, Control, and Dynamics*, Vol. 11, No. 6, Nov-Dec 1988, pp. 517-526.

28. Huang, C.Y., and Stengel, R.F., "Restructurable Control Using Proportional-Integral Implicit Model Following," *Journal of Guidance, Control, and Dynamics*, Vol. 13, No. 2, Mar-Apr 1990, pp. 303-309.

29. Hanson, G., and Stengel, R.F., "Effects of Displacement and Rate Saturation on the Control of Statically Unstable Aircraft," *Journal of Guidance, Control, and Dynamics*, Vol. 7, No. 2, Mar-Apr 1984, pp. 197-205.

30. Shrivastava, P.C., and Stengel, R.F., "Stability Boundaries for Aircraft with Unstable Lateral-Directional Dynamics and Control Saturation," *Journal of Guidance, Control, and Dynamics*, Vol. 12, No. 1, Jan-Feb 1989, pp. 62-70.

31. Minsky, M., and Papert, S., *Perceptrons: An Introduction to Computational Geometry*, M.I.T. Press, Cambridge, 1968 (rev. 1988).

32. McCorduck, P., *Machines Who Think*, W.H. Freeman, San Francisco, 1979.

33. Stefik, M., *et al*, "The Organization of Expert Systems, A Tutorial," *Artificial Intelligence*, Vol. 18, No. 2, Mar 1982, pp. 135-174.

34. Handelman, D.A., and Stengel, R.F., "Combining Quantitative and Qualitative Reasoning in Aircraft Failure Diagnosis," *AIAA Guidance, Navigation, and Control Conference*, Snowmass, CO, AIAA Paper No. 85-1905CP, Aug 1985.

35. Charniak, E., and McDermott, D., *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MA, 1985.

36. Handelman, D.A., and Stengel, R.F., "Combining Expert System and Analytical Redundancy Concepts for Fault-Tolerant Flight Control," *Journal of Guidance, Control, and Dynamics*, Vol. 12, No. 1, Jan-Feb 1989, pp. 39-45.

37. Handelman, D.A., and Stengel, R.F., "An Architecture for Real-Time Rule-Based Control," *Proceedings of the 1987 American Control Conference*, Minneapolis, MN, June 1987, pp. 1636-1642.

38. Huang, C.Y., and Stengel, R.F., "Failure Model Determination in a Knowledge-Based Control System," *Proceedings of the 1987 American Control Conference*, Minneapolis, MN, June 1987, pp. 1643-1648.

39. Handelman, D.A., and Stengel, R.F., "Perspectives on the Use of Rule-Based Control," in *Artificial Intelligence in Real-Time Control*, M.G. Rodd and G.J. Suski, ed., Pergamon Press, New York, 1989, pp.27-32.

40. Ng, K.-C., and Abramson, B., "Uncertainty Management in Expert Systems," *IEEE Expert*, Vol. 5, No. 2, Apr 1990, pp. 29-47.

41. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Palo Alto, CA, 1988.

42. Stratton, D.A., and Stengel, R.F., "Probabilistic Reasoning for Intelligent Wind Shear Avoidance," AIAA Paper No. 90-3437, *1990 AIAA Guidance, Navigation & Control Conference*, Portland, OR, Aug 1990.

43. Handelman, D.A., and Stengel, R.F., "Rule-Based Mechanisms of Learning for Intelligent Adaptive Flight Control," *Proceedings of the 1988 American Control Conference*, Atlanta, June 1988, pp. 208-213.

44. Tong, R., "A Control Engineering Review of Fuzzy Systems," *Automatica*, Vol. 13, No. 6, Nov 1977, pp. 559-569.

45. Belkin, B., and Stengel, R.F., "Quantitative Knowledge Acquisition for Expert Systems," *Proceedings of the Space Operations, Applications, and Research Symposium*, Albuquerque, NM, June 1990.

46. Quinlan, J.R., "Discovering Rules by Induction from Large Collections of Samples," in *Expert Systems in the Micro Electronic Age*, D. Michie, ed., Edinburgh University Press, Edinburgh, 1979, pp. 169-201.

47. Poggio, T., and Girosi, F., "Regularization Algorithms for Learning That Are Equivalent to Multilayer Networks," *Science*, Vol. 247, No. 4945, Feb 23, 1990, pp. 978-982.

48. Linse, D.J., and Stengel, R.F., "Neural Networks for Function Approximation in Nonlinear Control, *Proceedings of the 1990 American Control Conference*, San Diego, May 1990, pp. 675-679.

49. Funahashi, K.-I., "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, Vol. 2, 1989, pp.183-192.

50. Cybenko, G., "Approximation by Superposition of a Sigmoidal Function," *Mathematics of Control, Signals, and Systems*, Vol. 2, No. 4, 1989, pp. 303-314.

51. Rumelhart, D., Hinton, G., and Williams, R., "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing: Explorations in the Microstructure of Cognitions, Vol. 1: Foundations*, D. Rumelhart and J. McClelland, ed., MIT Press, Cambridge, 1986.

52. Singhal, S., and Wu, L., "Training Feed-Forward Networks with the Extended Kalman Algorithm," *Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, May 1989, pp. 1187-1190.

53. Albus, J.S., "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Vol. 97, Sept 1975, pp. 220-227.

54. Nguyen, D.H., and Widrow, B., "Neural Networks for Self-Learning Control Systems," *IEEE Control Systems Magazine*, Vol. 10, No. 3, Apr 1990, pp. 18-23.

55. Stengel, R.F., and Linse, D.J., "System Identification for Nonlinear Control Using Neural Networks," *Proceedings of the 1990 Conference on Information Sciences and Systems*, Princeton University, Mar 1990.

56. Fadali, M., *et al*, "Minimum-Time Control of Robotic Manipulators Using a Back Propagation Neural Network," *Proceedings of the 1990 American Control Conference*, San Diego, May 1990, pp. 2997-3000.

57. Miller, W.T., "Sensor-Based Control of Robotic Manipulators Using a General Learning Algorithm," *Journal of Robotics and Automation*, Vol. RA-3, No. 2, Apr 1987, pp. 157-165.

58. Passino, K., Sartori, M., and Antsaklis, P., "Neural Computing for Numeric-to-Symbolic Conversion in Control Systems," *IEEE Control Systems Magazine*, Vol. 9, No. 3, Apr 1989, pp. 44-52.

59. Naidu, S., Zafiriou, E., and McAvoy, T., "Use of Neural Networks for Sensor Failure Detection in a Control System," *IEEE Control Systems Magazine*, Vol. 10, No. 3, Apr 1990, pp. 49-55.

60. Handelman, D.A., Lane, S.H., and Gelfand, J.J., "Integrating Neural Networks and Knowledge-Based Systems for Intelligent Robotic Control," *IEEE Control Systems Magazine*, Vol. 10, No. 3, Apr 1990, pp. 77-87.

SYSTEMATIC METHODS FOR KNOWLEDGE ACQUISITION AND EXPERT SYSTEM DEVELOPMENT

Brenda L. Belkin* and Robert F. Stengel**

Princeton University
Department of Mechanical & Aerospace Engineering
Princeton, NJ, 08544

ABSTRACT

Nine cooperating rule-based systems, collectively called AUTOCREW, were designed to automate functions and decisions associated with a combat aircraft's subsystems. The organization of tasks within each system is described; performance metrics were developed to evaluate the workload of each rule base, and to assess the cooperation between the rule-bases.

Each AUTOCREW subsystem is composed of several expert systems that perform specific tasks. AUTOCREW's NAVIGATOR was analyzed in detail to understand the difficulties involved in designing the system and to identify tools and methodologies that ease development. The NAVIGATOR determines optimal navigation strategies from a set of available sensors. A Navigation Sensor Management (NSM) expert system was systematically designed from Kalman Filter covariance data; four ground-based, a satellite-based, and two on-board INS-aiding sensors were modelled and simulated to aid an INS. The NSM Expert was developed using the Analysis of Variance (ANOVA) and the ID3 algorithm. Navigation strategy selection is based on an RSS position error decision metric, which is computed from the covariance data. Results show that the NSM Expert predicts position error correctly between 45% and 100% of the time for a specified navaid configuration and aircraft trajectory. The NSM Expert adapts to new situations, and provides reasonable estimates of hybrid performance. The systematic nature of the ANOVA/ID3 method makes it broadly applicable to expert system design when experimental or simulation data is available.

INTRODUCTION

The computational and symbolic processing requirements for pilot-aiding systems pose many problems, e.g., multi-task scheduling and intersystem cooperation [1,2]. Expert systems, which are computer programs usually developed in a symbolic processing language such as LISP or PROLOG, have emerged to solve difficult domain-specific problems [3]. The designer generally extracts heuristics and specific knowledge from domain experts. This information is used to formulate a knowledge base consisting of parameters and rules. An inference engine uses rules to conduct a search and to set parameters, thereby inferring knowledge about the problem domain.

Problem identification in the flight environment and expert system development has led to the design and implementation of expert systems in the areas of navigation, emergency procedures, and air traffic control [4-6]. The introduction of multiple system concepts such as global blackboard architectures for information exchange between knowledge bases has been successful in aerospace implementations [2,7]. However, methods in multiple knowledge-base development, systems integration, and ensemble prototyping need further research if complex systems are to be developed for flight domain operations or other problem areas.

A logical task classification scheme is a key factor in the successful development of multiple rule-based systems. In this paper a logical structure for providing assistance to the pilot of a military aircraft was patterned on the functions of a World War II bomber crew [8]. Based on this model, an ensemble of nine cooperating rule-based systems called AUTOCREW [9,10] was developed. Each system figuratively emulates a crew member's task responsibilities.

To illustrate some of the issues involved in the design of an AUTOCREW subsystem, the problem of navigation sensor selection was studied in detail. With a large number of available navaid sensors, choosing an optimal or near-optimal sensor set becomes a large combinatorial problem. Convergence towards an optimal sensor configuration requires an exhaustive computer search utilizing simulation results as the basis for selection. In contrast, a small number of available navaids reduces the decision space considerably. Hence, a dilemma occurs; increasing sensor capability (and thus reliability and performance) increases decision-making complexity.

A Navigation Sensor Management (NSM) Expert System that controls the selection of multi-sensor configurations was developed for AUTOCREW. Two on-board navigation aids (Doppler radar and Air data sensor) and five radio navigation systems (Global Positioning System, LORAN, TACAN, DME, and VOR) were modelled and covariance results obtained using a U-D implementation of the Kalman Filter equations [9]. Up to three ground stations and four satellites were simulated for typical high-performance, commercial jet transport, and general aviation aircraft trajectories. Mixed sensor simulations also were performed. These results provided the combinatorial, multi-factor source required for knowledge-base development.

The Analysis of Variance (ANOVA) statistical technique [13] was applied to the covariance results to identify the factors that cause variation in navigation performance. Once the important factors were identified, the relationships between them were determined. The ID3 algorithm [14], an inductive inference technique based on the probabilistic occurrence of events, was used to find these attribute relationships. Details of the development of the NSM Expert are described in Ref. 9, and the main results are shown in this paper.

AUTOCREW DESCRIPTION

The nine AUTOCREW components [9,10] are COPILOT (flight control, aircraft performance, terrain following), ENGINEER (system diagnosis, reconfiguration, emergency procedures), NAVIGATOR (navigation sensor management, dynamic route planning), COMMUNICATOR (radio/data operations), OBSERVER (lookout and alarm, identification friend or foe) ATTACKER (offensive weapon control, target acquisition/prioritization), DEFENDER (defensive weapon control), and SPOOFER (electronic measures/countermeasures). The ninth rule-based system, the EXECUTIVE, coordinates mission-specific tasks and has knowledge of the mission plan. In the AUTOCREW scheme, the human aircraft Pilot acts in the capacity of mission coordinator. The Pilot has full control of the aircraft and its on-board systems; AUTOCREW can provide automatic assistance when requested. In total, the AUTOCREW system performs 511 tasks, is comprised of 406 rules, and uses 89 shared parameters [9].

Knowledge-Base Development of AUTOCREW Components

Five main task groups are performed by each AUTOCREW component: 1) tasks executed during an attack on the aircraft, (2) tasks executed during aircraft subsystem emergency or potential threat

---

* Formerly, Graduate Student, Princeton University. Currently, Member of Technical Staff, AT&T Bell Laboratories, 480 Red Hill Road, Middletown, NJ. 07748

** Professor of Mechanical & Aerospace Engineering

Presented at the 29th IEEE Conference on Decision and Control, Honolulu, HI, December 5-7, 1990.

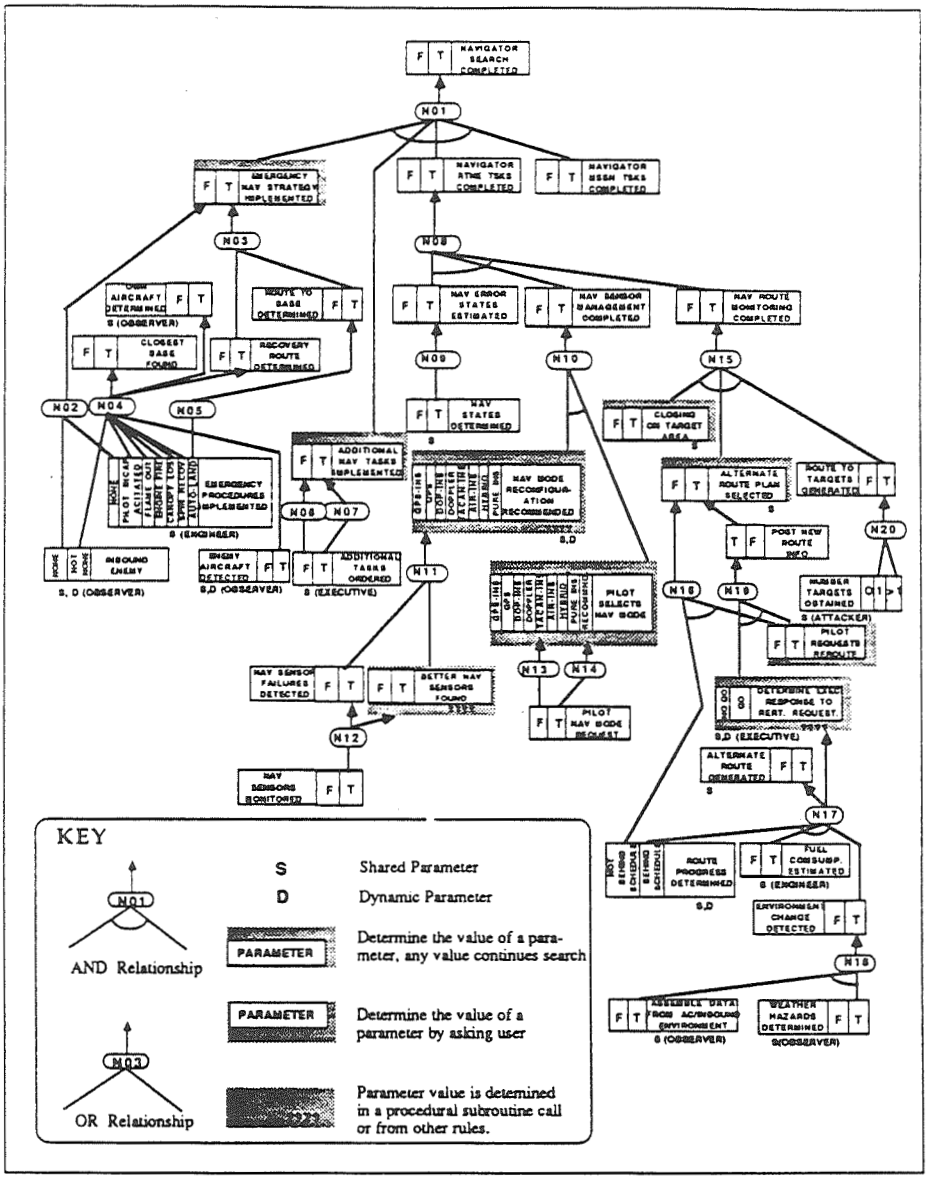


Figure 1. Knowledge Base of AUTOCREW Member NAVIGATOR

situations, (3) additional tasks ordered by the EXECUTIVE, (4) tasks executed on a routine basis, and (5) mission-specific tasks that (for the most part) are executed once and that require a high degree of cooperation among the various systems. Mission-specific tasks are divided into groups appropriate to each mission phase [10].

Figure 1 shows the graphical representation of AUTOCREW's NAVIGATOR knowledge base (see Refs. 9 and 10 for details of the AUTOCREW knowledge bases). The NAVIGATOR's tasks include navigation sensor management, navigation state and error estimation, and dynamic route planning. As seen in the figure, detection of an inbound enemy triggers the NAVIGATOR to perform defensive tasks. Rule N04 shows that the NAVIGATOR finds the nearest friendly airfield able to serve the aircraft from its stored airbase information.

The NAVIGATOR also uses the OBSERVER's information on friendly aircraft in the area and plans a recovery route. It is extremely important that the aircraft's location is known accurately in order to avoid such danger zones as Surface-to-Air Missile (SAM) belts. Tactical aircraft carry inertial navigation systems (INS) for high-frequency navigation information. A Kalman Filter can be used to integrate external navaids with an INS to estimate navigation errors. The *integrated* or *hybrid* INS gives accurate high-frequency navigation information with bounded errors in the state estimates (Rule N09). The error magnitudes depend on the accuracy of the navaid selected. The NAVIGATOR monitors ground-based and satellite radio navaids

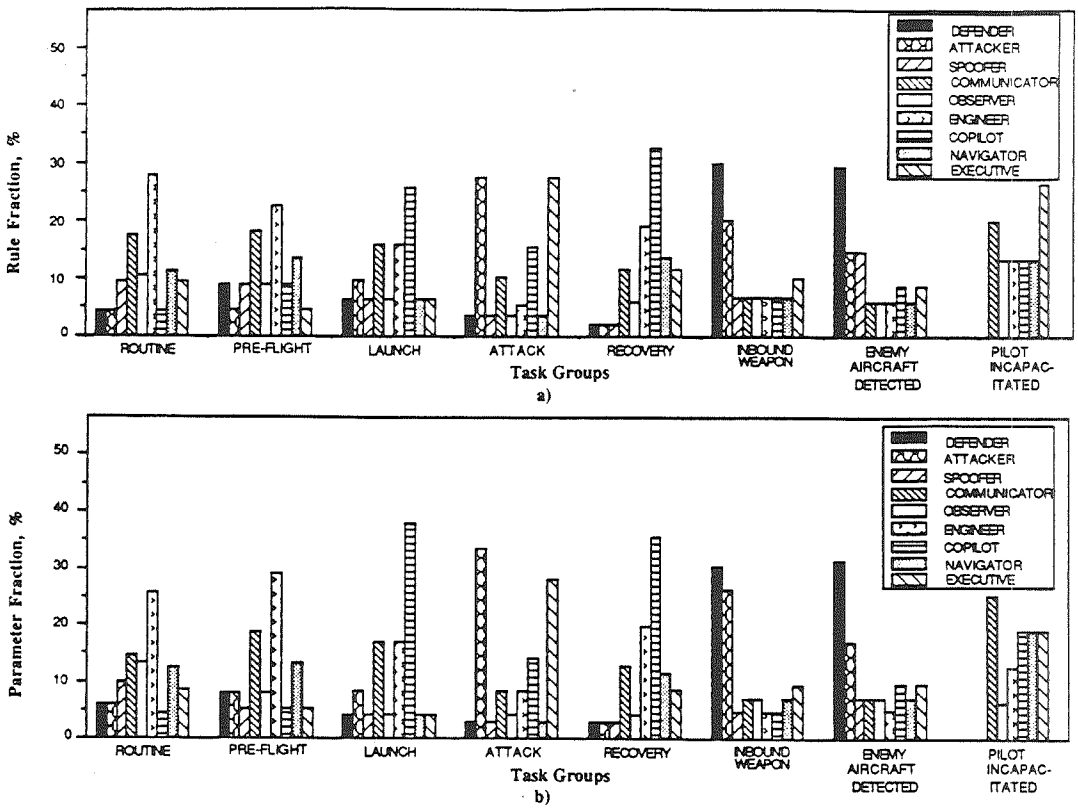


Figure 2. AUTOCREW Workload - Mission-Specific Tasks and Emergency Scenarios

in the area to determine if they are sufficiently operational for integration purposes (Rule 11). From this list of navaids, the NAVIGATOR recommends the best combination meeting the computational resource constraints. The Pilot can change the NAVIGATOR's selection if desired (Rule N14).

The NAVIGATOR also is responsible for routinely monitoring flight plan progress. The NAVIGATOR first determines if the aircraft is closing on the target area. This information initiates additional tasks in the remaining AUTOCREW knowledge bases. The NAVIGATOR keeps a record of aircraft positions, restricted zones, and areas of weather disturbances as reported by the OBSERVER (Rule N18). The NAVIGATOR generates an alternate flight path for the following conditions: flight path conflicts, flight plan behind schedule, and fuel considerations. The EXECUTIVE is consulted when the NAVIGATOR recommends an alternate route. The EXECUTIVE considers the effect of the route change on the overall mission plan and aircraft's safety, and it makes a GO/NO GO recommendation (Rule N17). The Pilot can accept or reject the EXECUTIVE's recommendation in Rule N19. If the new route is selected, the NAVIGATOR posts the route plan details on the display. As noted above, the COPILOT changes its steering commands in response to the route change. When the ATTACKER acquires enemy targets and the plan is to engage, the NAVIGATOR generates routes to each target, as seen in Rule N20.

AUTOCREW Simulation Results

Simulation and comparative workload results for two mission scenarios are given in this section. The scenarios are inbound SAM attack on the aircraft and Pilot incapacitation. Results are presented for five mission phases (routine, pre-flight, launch, attack, and recovery) and three emergency conditions (inbound weapon, enemy aircraft detected, and pilot incapacitated).

AUTOCREW workloads for three emergency conditions were obtained and compared with the routine workload, referred to as "cruise mode" (Table I). For the three scenarios the workload progressively increases as the situation gets worse. There is a dramatic increase in workload between cruise and attack modes with an inbound SAM, and there is an additional increase between the inbound SAM scenario (cruise), and inbound SAM with the Pilot incapacitated (cruise). The workload nearly doubles for an inbound SAM when the aircraft is already engaging another aircraft (attack mode).

Table I Scenario Workload-Increase Comparisons

Scenario	Workload Increase, %	
	Tasks	Decision-Making
Cruise Mode	-----	-----
Inbound SAM/Cruise Mode	26.3	28.1
Inbound SAM & Pilot Incapacitated/Cruise Mode	39.5	38.6
Inbound SAM/Attack Mode	77.2	75.2

A comparison of the workload distribution among the AUTOCREW members for each of the given mission phases is shown in Fig. 2. The Rule Fraction is the ratio of number of rules fired during the mission phase for the specified AUTOCREW member to the total number of mission phase rules in all AUTOCREW knowledge bases [9]. The Parameter Fraction is the ratio of number of tasks performed during the mission phase for the specified AUTOCREW component to the total number of mission phase tasks in all AUTOCREW knowledge bases. These metrics were used to characterize the knowledge base

workloads. The ENGINEER, COMMUNICATOR, OBSERVER, NAVIGATOR and EXECUTIVE have the largest amount of routine tasks to perform each search cycle. During the pre-flight phase, the ENGINEER, COMMUNICATOR, and NAVIGATOR are the busiest components, as each prepares the aircraft for takeoff. About 40% of all the launch phase tasks are performed by the COPILOT, as shown in Fig. 2, while the next largest task load is executed by the COMMUNICATOR and ENGINEER (each having 17% of the total launch workload). Most of the attack phase work is done by the ATTACKER (33%), whereas the recovery phase is dominated by the COPILOT's (35%) activities. Most of the workload during an inbound weapon attack is performed by the DEFENDER (about 30% in Fig. 2). The ATTACKER's workload also increases at this time, as it monitors the DEFENDER's firepower capability.

There is evidence of more SPOOFER decision-making when an enemy aircraft is detected than when the OBSERVER detects an inbound weapon. This is due to its Electronic Counter Measures (ECM) strategy consultation with the EXECUTIVE. The COPILOT's workload also increases during these two emergency conditions; in both scenarios, the COPILOT selects an appropriate evasive maneuver. These selections are based on the aircraft capability information provided by the ENGINEER. The EXECUTIVE's workload also increases at this time to assist the Pilot in making decisions. When the Pilot is incapacitated, the EXECUTIVE becomes the primary decision-maker, as shown in Fig. 2. The ENGINEER's major task during this scenario is to monitor the Pilot's condition, resulting in a 10% workload increase.

DEVELOPMENT OF A NAVIGATION SENSOR MANAGEMENT EXPERT SYSTEM

The remainder of this paper focuses on the issues involved in designing the navigation sensor selection expert in NAVIGATOR. This example demonstrates how simulation data and statistical analysis can be used to systematically design an expert system.

To determine the performance of the various hybrid navaid combinations, the following U-D Kalman filter simulations were performed:

1. Single-type hybrids: GPS, LORAN, TACAN, DME, VOR, Doppler Radar, or Air Data Sensor aiding an INS
2. Number of stations used in a single-type hybrid
3. Multi-type hybrids: Combinations of different navaid types aiding an INS
4. Aircraft trajectories simulated: High-performance, commercial, general aviation

The root sum of squares (RSS) of the north and east component errors was selected for the hybrid system performance metric. The primary function of this expert system is to select the external navaid sensors that provide the smallest possible RSS position error from a large set of available sensors. The Analysis of Variance (ANOVA) technique [13] is used to identify the factors that make statistically significant contributions to the decision metric. Then, the ID3 algorithm determines the relationships between these factors [13,14].

Identifying Important Factors Using ANOVA

The RSS position error histories from over 200 covariance simulations were obtained, and the data were used in an ANOVA four-factor navaid experiment. The goal of the experiment was to identify which of the factors (navaid type, number of ground stations, trajectory effects, performance history) and their interactions had statistically significant impacts on the RSS position error. The factor values used in the ANOVA experiment were:

Navaids={VOR, DME, LORAN, TACAN, GPS} (Model 1)
Number of Stations={One, Two, Three}
Trajectory Type={High-Performance, Commercial Transport, General Aviation}
Performance History = {Intervals: I, II, III, IV}

Since each trajectory consists of four, 15-minute legs, the Performance History (or "time interval") factor refers to the RSS error obtained within each 15-minute time frame. Four single-station, six double-station, and four triple-station hybrids were simulated using combinations of four different Stations (named Stations A-D) [9,12].

The ANOVA technique was applied to the simulation results as follows: First, the mean value of the RSS position error and the variance using all simulation data were computed. The ANOVA model decomposes the variance into a sum of variances, each associated with a potentially contributing factor and factor interactions. For example, if a two-factor ANOVA experiment using navaid type and number of ground stations was performed, the total variation of the RSS position error in the complete set of simulation results would be:

$$SS_{TOTAL} = SS_{NAV} + SS_{STAT} + SS_{NAV\text{-}STAT\ INT} + SS_{UNEXPLAINED} \quad (1)$$

Here, SS_{TOTAL} is the total variation in the RSS position error based on all simulations, SS_{NAV} is the variation in the RSS position error due to differences in navaid type, SS_{STAT} is the variation in the RSS position error due to the number of ground stations, and $SS_{NAV\text{-}STAT\ INT}$ is the variation in the RSS position error due to interaction effects. $SS_{UNEXPLAINED}$ is the variation in the RSS position error that cannot be attributed to navaid type, the number of stations, or interactions between these two factors.

Equation (1) was expanded for the four-factor navaid experiment defined by Model 1 [9,12]. Using the computed sums-of-squares, Scheffé multiple comparison tests were performed to identify the specific differences within the groups (e.g., RSS position error differences between different navaid types).

Extracting Rules Using Induction: The ID3 Algorithm

The ID3 Algorithm uses inductive inference to extract rules [14] from an example set. The problem space is described in terms of attributes, where each attribute is characterized by a set of values that define the possible "states." For example, in Model 1, the navaid type and number of ground stations were shown to be attributes affecting RSS position error. The ID3 algorithm defines the shape of the decision tree as it identifies the most important attribute at each decision node. The algorithm uses an Information-Theoretic Measure (ITM) that minimizes the number of tests (attribute nodes) necessary to define the decision tree. The ID3 algorithm uses the ITM in a splitting strategy [14] to decide which attribute provides the most information from the example set. (See Refs. 9, 12, 14 for details of the ID3 algorithm.)

Developing the ID3 Attribute Framework Using ANOVA Results

The Model 1 covariance simulations were used to extract decision trees for the NSM Expert system. Eleven attributes were defined for the ID3 framework: Navaid type, trajectory leg, aircraft groundspeed, number of ground stations, minimum geodetic distance from station, maximum geodetic distance from station, (Max - Min) distance on trajectory leg, maximum line-of-site angle from station, minimum line-of-sight angle from the station, direction of flight (approaching or receding) relative to station, and RSS position error class on previous trajectory leg. The ID3 algorithm determined how these attributes were related to each other and to the final RSS position error.

The classification scheme chosen to represent the RSS position error endnode in the NSM decision trees is depicted in Table II. The velocity, distance, and line-of-sight angles were expressed in terms of ranges instead of individual values, so the expert system weights trends more heavily than specific examples. This makes the expert system more adaptable to new conditions, because matches between the actual and knowledge-base cases can be obtained more frequently.

Table II RSS Position Error Classification Scheme

[High] Error Range, N. Mi.	Code	Accuracy [Medium] Error Range, N. Mi.	Code	[Low] Error Range, N. Mi.	Code
0.0-0.02	c-1	0.10-0.20	c-6	1.0-1.5	c-15
0.02-0.04	c-2	0.20-0.30	c-7	1.5-2.0	c-16
0.04-0.06	c-3	0.30-0.40	c-8	2.0-2.5	c-17
0.06-0.08	c-4	0.40-0.50	c-9	2.5-3.0	c-18
0.08-0.10	c-5	0.50-0.60	c-10	3.0-3.5	c-19
		0.60-0.70	c-11	3.5-4.0	c-20
		0.70-0.80	c-12	4.0-4.5	c-21
		0.80-0.90	c-13	4.5-5.0	c-22
		0.90-1.00	c-14	> 5.00	c-23

The example set was developed using the attribute framework described above. The RSS position errors for each simulation were classified on each trajectory leg using the scheme in Table II. The ID3 example base was then created from each single-, double-, and triple-station simulation.

NSM Decision Trees

The NSM example set was divided into 17 smaller example sets. The GPS and on-board navaid examples were grouped into one expert, whereas the ground-based navaid examples were divided according to navaid type and time (15-minute intervals). The ID3 algorithm constructed decision trees for each of the 17 small expert systems to comprise the larger NSM Expert. The total number of examples used to develop the NSM Expert System was 932, based on 260 Kalman Filter covariance simulations. An additional 37 simulations were performed to obtain an RSS error-estimate decision tree for navaid-type combinations. The NSM expert system prompts the user for a set of flight conditions commensurate with the attribute/value lists used in the example set, and the resulting RSS classification code is returned to the user from the decision tree.

Figure 3 shows decision trees for single-, double-, and triple-TACAN station combinations on the second 15-minute trajectory leg. Here, the majority of the testing nodes are trajectory parameters (distance, LOS angle, direction of flight with respect to the stations). The top or root node is the aircraft's direction of flight. This is expected because the distance and LOS angle attributes are dependent on directional motion. Distance, LOS angle, and groundspeed are results of the aircraft's motion; hence, they represent more specific problem parameters, and it is expected that these parameters appear deeper in the decision tree. Figure 3 also shows that distance, ground velocity, LOS angle, and hybrid performance history are significant factors in RSS error prediction.

The decision tree in Fig. 4 shows the expected position error range when different navaid types are integrated into a hybrid system. The RSS position errors for these simulations were averaged over the entire flight time for the high-performance trajectory. The tree is organized in terms of the navigation method used: (1) Distance-Velocity (ρ-V), (2) Bearing-Velocity (θ-V), (3) Distance-Bearing (ρ–θ), (4) Distance-Distance (ρ–ρ), (5) Bearing-Bearing (θ–θ), and (6) Velocity-Velocity (V-V). These results show that LORAN is a better distance-measuring navaid than DME and that Doppler Radar provides better navigation accuracy than the Air Data Sensor when ρ-V navigation is used. The ρ–θ results show that it is possible to obtain good performance when LORAN and VOR are used. The LORAN/DME hybrid gives better results than two DME stations but worse performance than two LORAN stations. By far the worst results are obtained using two VOR stations.

Performance of NSM Expert on Test Trajectories

Two high-performance trajectories were used in the performance evaluation of the NSM Expert. Multi-station hybrids were simulated on each test trajectory for the DME, VOR, TACAN, and LORAN systems. In total, 60 covariance simulations were performed for the

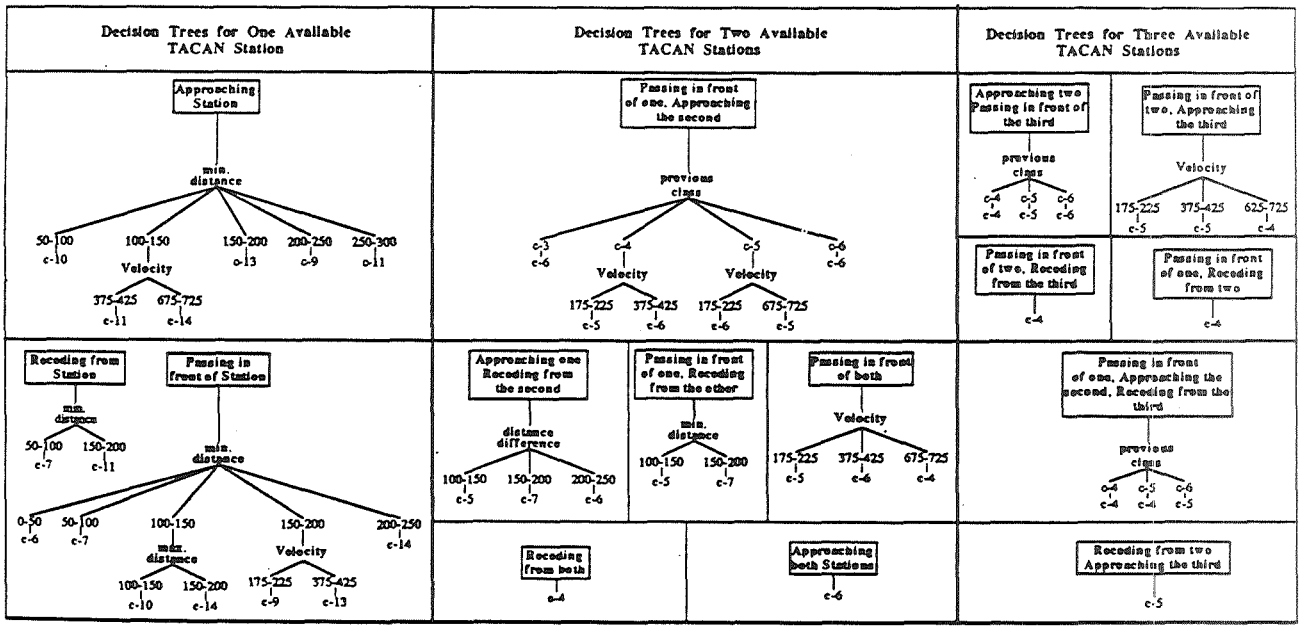


Figure 3. Decision Trees Predicting RSS Position Error Range for an INS Aided by TACAN During the Second 15 Minutes of Flight

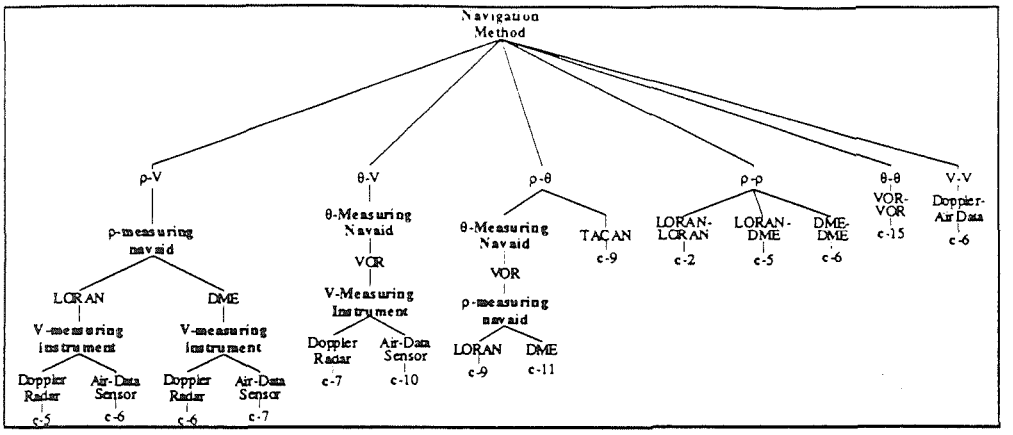


Figure 4. Decision Tree Predicting RSS Performance When Different Navaid Combinations are Used to Aid an INS

two test trajectories. The results for each simulation were classified on each trajectory leg according to the scheme in Table II. The total number of matches was counted on each trajectory leg for the seven navaid types studied. Figure 5 shows how well the NSM Expert predicts the RSS position error for each hybrid configuration. The *predictive performance metric* for each navaid is defined as the percentage of number of matches obtained from the total number of combinations tested for that navaid.

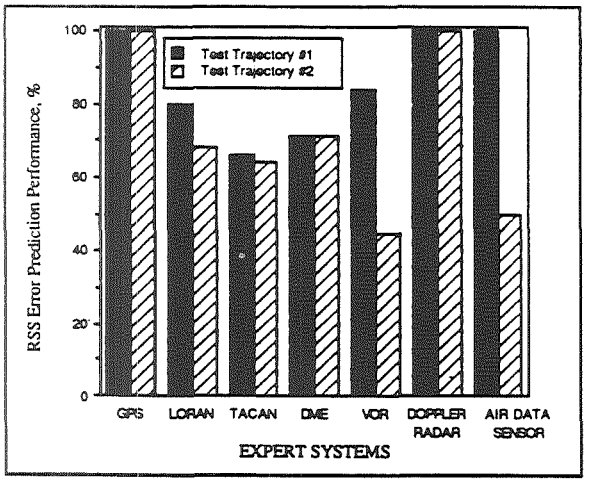


Figure 5. Performance of Navaid Experts on Test Trajectories

The NSM Expert performed very well on the two test trajectories. Figure 5 shows that the NSM Expert correctly predicts the RSS position error better than 70% of the time on test Trajectory #1. The system required only the trajectory information and its knowledge of hybrid system performance to make these predictions. However, its predictive capability on test Trajectory #2 is slightly worse for the LORAN hybrids (69%), considerably worse for the VOR (45%) and Air Data sensor hybrids (53%), and identical to the trajectory #1 results for the remaining configurations. Hence, the results from Trajectory #2 suggest that additional investigation into trajectory effects on VOR's and Air Data Sensor's performance may be necessary.

Selecting Navigation Strategies Using the NSM Expert System

The NSM-Recommended Navaid strategies were compared with covariance-determined strategies for various navigation sensor suite scenarios [9,12]. For example, the NSM Expert was asked to find the best two-measurement hybrid strategies for the following navaid sensors:

Sensor Scenario #1 = {2 GPS satellites, LORAN Station A, DME Station A, VOR Station C, Doppler Radar}

The results obtained by mixing combinations of the Scenario #1 sensors are shown in Fig. 6. The NSM Expert correctly identified the best, next-best, and third-best navigation strategies. Additional NSM Expert results are given elsewhere [9,12].

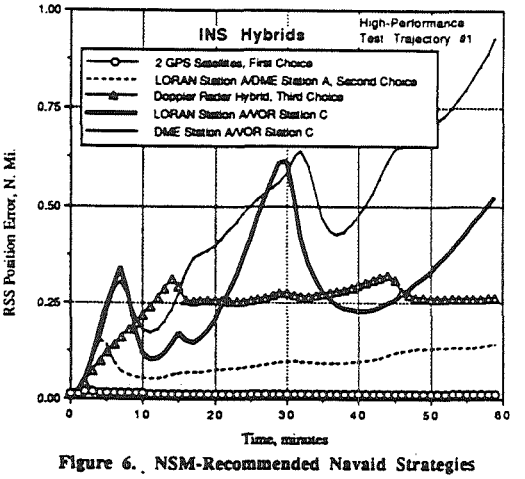


Figure 6. NSM-Recommended Navaid Strategies

In addition to recommending navigation strategies from a given set of sensors, several heuristic sensor selection schemes were determined from the ANOVA results, the Scheffé comparisons, and the ID3 decision trees.

For single-type hybrids whose ground stations are at or near the same location, the ranking in order of best to worst performance was determined from ANOVA and the Scheffé tests as follows:

Ranking heuristic for co-located navaids = {GPS, LORAN or TACAN, DME, VOR}

The ranking scheme above applies to single-station hybrids (or double-satellite GPS hybrids). The Scheffé comparisons show that a TACAN or LORAN selection depends on trajectory effects; this means that either a TACAN or LORAN hybrid could provide statistically lower RSS position errors depending on the aircraft flight path with respect to the stations being compared. The ranking heuristic is accurate with a 99% confidence level. If non-colocated stations are compared, trajectory effects must be considered for these single-station hybrids. The ANOVA and Scheffé comparison tests showed that double-station hybrids provide much smaller RSS position errors than single-station hybrids [9,11,12]. The performance difference between two and three stations is not as dramatic but is nonetheless statistically significant [12].

For multi-station hybrids of the same navaid type, the ANOVA and ID3 results provide the following ranking scheme from best to worst navigation performance as follows [9]:

Ranking heuristic for multi-station hybrids and on-board sensor hybrids = {4 GPS, 3 GPS, 3 LORAN, 2 LORAN, 3 TACAN, 2 TACAN, 3 DME, 2 DME, Doppler Radar, Air data sensor, 3 VOR, 2 VOR}

However, the decision trees highlight the importance of trajectory effects in RSS position error estimation.

CONCLUSIONS

This paper summarized two research efforts to develop expert systems using systematic methods: The design methodology for multiple cooperating rule-based systems, and the development of a navigation sensor management system. Nine modular rule-based systems, collectively called AUTOCREW, were designed to automate functions and decisions associated with a combat aircraft's subsystems. The nine AUTOCREW knowledge bases were designed individually and areas of cooperation between the knowledge bases were identified. Each knowledge-base was designed using a graphical symbology that clearly illustrated the relationships between functional and decision-making tasks. Performance metrics were developed to evaluate the workload of each knowledge base and to assess the cooperation between the rule-bases.

The NAVIGATOR's sensor management task was treated in detail. The performances of seven navigation systems aiding a medium-accuracy Inertial Navigation System (INS) were investigated using Kalman Filter covariance analyses. A Navigation Sensor Management Expert System was formulated from covariance simulation data using the Analysis-of-Variance (ANOVA) method and the ID3 algorithm. ANOVA results show that statistically different position accuracies are obtained when different navaids are used, the number of navaids aiding the INS is varied, the aircraft's trajectory is varied, and the performance history is varied. The ID3 algorithm determines the NSM Expert's classification "rules" in the form of decision trees. The performances of these decision trees were assessed on two arbitrary trajectories, and the results demonstrate that the NSM Expert adapts to new situations and provides reasonable estimates of the expected hybrid performance. The NSM Expert demonstrates that carefully-planned simulation experiments can be used to develop a fully-operational expert system with a designer-specified performance effectiveness.

ACKNOWLEDGMENTS

This project was sponsored by the U.S. Army Research Office under Contract No. DAAG29-84-K-0048, and supported by NASA and the FAA under Grant No. NGL31-001-252. This work was partially supported by the Natural Sciences and Research Council of Canada (NSERC) through the "1967" Scholarship. Lawrence Rosen's many helpful comments on this manuscript are also gratefully acknowledged.

REFERENCES

[1] D.A. Handelman and R.F. Stengel, "A Theory for Fault-Tolerant Flight Control Combining Expert System and Analytical Redundancy Concepts", in *Proceedings of the 1986 AIAA Guidance, Navigation, and Control Conference*, 1986, pp. 375-384.

[2] W.K. Ericksen, "The Blackboard Model: A Framework for Integrating Multiple Cooperating Expert Systems", in *Proceedings of the 5th Computers in Aerospace Conference*, 1985, pp. 33-40.

[3] F. Hayes-Roth, D. Waterman, and D. Lenat, *Building Expert Systems*, Addison-Wesley Publishing Company, Reading, MA., 1983.

[4] A.D. Pisarro and H.L. Jones, "An Expert System Approach to Adaptive Tactical Navigation", in *Proceedings of The First Conference on Artificial Intelligence Applications*, IEEE Computer Society, 1982, pp. 460-464.

[5] B.M. Anderson *et al*, "Intelligent Automation of Emergency Procedures in Advanced Fighter Aircraft", in *Proceedings of The First Conference on Artificial Intelligence Applications*, IEEE Computer Society, 1982, pp. 496-501.

[6] S.E. Cross, "Model-Based Reasoning in Expert Systems: An Application to Enroute Air Traffic Control", in *Proceedings of the AIAA/IEEE 6th Digital Avionics Systems Conference*, 1984, pp. 95-101.

[7] D.A. Handelman and R.F. Stengel, "An Architecture for Real-Time Rule-Based Control", in *Proceedings of the 1987 American Control Conference*, 1987, pp. 1636-1642.

[8] E. Jablonski, *Flying Fortress*, Doubleday, Garden City, New York, 1965, pp. 324-339.

[9] B. L. Belkin, *Cooperative Rule-Based Systems for Aircraft Navigation and Control, M.S.E. Thesis*, Report 1856T, Department of Mechanical and Aerospace Engineering, Princeton University, June 1989.

[10] B.L. Belkin and R.F. Stengel, "Cooperative Rule-Based Systems For Aircraft Control", in *Proceedings of the 26th IEEE Conference on Decision and Control*, Los Angeles CA, December 1987, pp. 1934-1940.

[11] B. L. Belkin and R. F. Stengel, "Quantitative Knowledge Acquisition for Expert Systems", Presented at the *Space Operations, Applications, and Research Symposium*, Albuquerque, NM, June 1990.

[12] B. L. Belkin and R. F. Stengel, "Knowledge Acquisition for Expert Systems Using Statistical Methods", Presented at the *51st AGARD Symposium of the Guidance and Control Panel on Knowledge Based System Applications for Guidance and Control*, Madrid, Spain, September 1990.

[13] G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experiments: An Introduction to Design, Data Analysis and Model Building*, John Wiley & Sons, New York, 1978.

[14] J. R. Quinlan, "Discovering Rules by Induction From Large Collections of Examples", in *Expert Systems in the Micro Electronic Age*, D. Michie, Editor, Edinburgh University Press, 1979, pp. 169-201.

[15] A. Gelb and A. Sutherland Jr., "Software Advances in Aided Inertial Navigation Systems", *Navigation, The Institute of Navigation*, Washington DC, Vol. 17, No.4, Winter 1970-71, pp. 358-369.

[16] R. F. Stengel, *Stochastic Optimal Control -- Theory and Application*, J. Wiley & Sons, New York, 1986.

Technical Notes and Correspondence

Stochastic Robustness of Linear Time-Invariant Control Systems

Robert F. Stengel and Laura R. Ray

Abstract—**A simple numerical procedure for estimating the *stochastic robustness* of a linear time-invariant system is described. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Confidence intervals for the scalar probability of instability address computational issues inherent in Monte Carlo simulation. Trivial extensions of the procedure admit consideration of alternate discriminants; thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions can also be estimated. Results are particularly amenable to graphical presentation.**

I. Introduction

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model.

Guaranteeing robustness has long been a design objective of control system analysis, although in most instances, insensitivity to parameter variations has been treated as a deterministic problem (see [1] for a comprehensive presentation of both classical and modern robust control). Stability (gain and phase) margins are useful concepts for designing robust single-input/single-output systems, addressing disturbance rejection and other performance goals, and they are amenable to the manual graphical procedures that preceded the widespread use of computers. With the help of computers, singular-value analysis has extended the frequency-domain approach to multiinput/multioutput systems (e.g., [2], [3]); however, guaranteed stability-bound estimates often are unduly conservative, and the relationship to parameter variations in the physical system is weak. Structured singular-value analysis [4] reduces this conservatism somewhat, and alternate treatments of structured parameter variations have been proposed (e.g., [5]–[7]), although these approaches remain deterministic. Reference [8] uses the term "stochastic robustness" to describe a stability bound based on Lyapunov methods and parameter perturbations that are modeled as stochastic sequences. This is a deterministic stability bound expressed in terms of the norm of a vector of noise variances. Elements of "stochastic stability" [9] have application to robustness but have yet to be presented in that context.

The notion of *probability of instability*, which is central to the analysis of stochastic robustness, was introduced in [10], with application to the robustness of the Space Shuttle's flight control system, and it is further described in [11]–[14]. This method determines the *stochastic robustness* of a linear time-invariant system by the probability distributions of closed-loop eigenvalues, given the statistics of the variable parameters in the plant's dynamic model. The probability that all of these eigenvalues lie in the open left-half s plane is the scalar measure of robustness.

With the advent of graphics workstations, the stochastic robustness of a system is easily computed by Monte Carlo simulation, and results can be displayed pictorially, providing insight into otherwise hidden robustness properties of the system. The method is computationally simple, requiring only matrix manipulation and eigenvalue computation, and it is inherently nonconservative, given a large enough sample space. Furthermore, the analysis of stochastic robustness is a logical adjunct to parameter-space control design methods [15]–[18]. Details of the approach and an example are given in the sequel.

Manuscript received October 14, 1988; revised September 30, 1989. Paper recommended by Associate Editor at Large, A. Benveniste. This work was supported by the FAA and the NASA Langley Research Center under Grant NGL 31-001-252.

The authors are with the Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544.

IEEE Log Number 9039313.

II. Probability of Instability

Consider a linear time-invariant (LTI) system subject to LTI control

$$\dot{x}(t) = F(p)x(t) + G(p)u(t) \tag{1}$$

$$y(t) = H(p)x(t) \tag{2}$$

$$u(t) = u_c(t) - CH(p)x(t) \tag{3}$$

$x(t)$, $u(t)$, $y(t)$, and p are state, control, output, and parameter vectors of dimension n, m, q, and r, respectively, accompanied by conformable dynamic, control, and output matrices F, G, and H, which may be arbitrary functions of p. $u_c(t)$ is a command input vector, and, for simplicity, the $(m \times n)$ control gain matrix C is assumed to be known without error. The n eigenvalues, $\lambda_i = \sigma_i + j\omega_i$, $i = 1$ to n, of the matrix $[F(p) - G(p)CH(p)]$ determine closed-loop stability and can be determined as the roots of the determinant equation

$$|sI_n - [F(p) - G(p)CH(p)]| = 0 \tag{4}$$

where s is a complex operator and I_n is the $(n \times n)$ identity matrix. While the explicit relationship between parameters and eigenvalues is complicated, estimating the probability of instability of the closed-loop system from repeated eigenvalue calculation is a straightforward task. Putting aside the mathematical intricacies, note that the probability of stability plus the probability of instability is one. Since stability requires all the roots to be in the open left-half s plane, while instability results from even a single right-half s plane root, we may write

$$\Pr(\text{instability}) \triangleq \bar{P} = 1 - \int_{-\infty}^{0} \text{pr}(\sigma)\, d\sigma \tag{5}$$

where σ is an n-vector of the real parts of the system's eigenvalues ($\lambda = \sigma + j\omega$), $\text{pr}(\sigma)$ is the joint probability density function of σ (unknown analytically), and the integral that defines the probability of stability is evaluated over the space of individual components of σ. Denoting the probability density function of p as $\text{pr}(p)$, (4) is evaluated J times with each element of p_j, $j = 1$ to J, specified by a random-number generator whose individual outputs are shaped by $\text{pr}(p)$. This *Monte Carlo evaluation* of the probability of stability becomes increasingly precise as J becomes large. Then

$$\int_{-\infty}^{0} \text{pr}(\sigma)\, d\sigma = \lim_{J \to \infty} \frac{N(\sigma_{\max} \le 0)}{J}. \tag{6}$$

$N[\cdot]$ is the number of cases for which all elements of σ are less than or equal to zero, that is, for which $\sigma_{max} \leq 0$, where σ_{max} is the maximum real eigenvalue component in σ. An important feature of this definition is that it does not depend on the eigenvalues and eigenvectors retaining fixed structures. As parameters change, complex roots may coalesce to become real roots (or the reverse), and modes may exchange relative frequencies. The only matter for concern is whether or not all real parts of the eigenvalues remain in the left-half s plane. For $J < \infty$, the probability of instability resulting from Monte Carlo evaluation is an estimate, denoted $\hat{\mathcal{P}}$.

Histograms and cumulative distributions for varying degrees of stability are readily given by the Monte Carlo estimate of $\int_{-\infty}^{\Sigma} \mathrm{pr}(\sigma)\, d\sigma$, where Σ represents a maximum real eigenvalue component, and $-\infty < \Sigma < \infty$. This histogram is a plot of $N[(\Sigma - \Delta) < \sigma_{max} \leq \Sigma]/J$ versus Σ; Δ is an increment in Σ, $N[\cdot]$ is the number of cases whose maximum real eigenvalue components lie in the increment, and J is the total number of evaluations. The histogram estimates the *stability probability density function*, $\mathrm{pr}(\Sigma)$, which is obtained in the limit for a continuous distribution of Σ as $\Delta \to 0$ and $J \to \infty$. The *cumulative probability distribution of stability*, $\Pr(\Sigma)$, is similarly estimated and presented as $N(\sigma_{max} \leq \Sigma)/J$ versus Σ, the exact distribution being achieved in the limit as $J \to \infty$. Consequently

$$\hat{\mathcal{P}} = 1 - \Pr(0). \tag{7}$$

There is, of course, no limitation on admissible specifications for the multivariate $\mathrm{pr}(p)$; it may be Gaussian or non-Gaussian, as appropriate. Rayleigh, correlated, and any other well-posed distributions are admissible, the principal challenge being to properly shape (and correlate) the outputs of the random-number generator. In practice, system parameter uncertainties are most likely to be bounded, as typical quality control procedures eliminate out-of-tolerance devices, and there are physical limitations on component size, weight, shape, etc. The uniform distribution is particularly interesting, as it readily models bounded uncertainty, and it is the default distribution of most algorithms for random-number generation. Given binary distributions for each parameter, in which the elements of p take maximum or minimum values with equal probability, the Monte Carlo evaluation reduces to 2^r deterministic evaluations, the result is exact, and the probability associated with each possible value of p is $1/2^r$. Similarly, the distribution for r parameters, each of which takes w values (i.e., for quantized uniform distributions), can be obtained from w^r evaluations; the probability of acquiring each value of p is $1/w^r$.

When has stochastic robustness been achieved? The answer is problem-dependent. $\mathrm{pr}(p)$ should be chosen to reflect physical limits of parameter uncertainty. In some applications involving bounded parameters, it will be possible to choose C such that $\hat{\mathcal{P}} = 0$, and that is a desirable goal; however, if admissible parameter variations are unbounded, if C is constrained, or if the rank of CH is less than n, the minimum $\hat{\mathcal{P}}$ may be greater than zero. C then must be chosen to satisfy performance goals and one of two robustness criteria: minimum $\hat{\mathcal{P}}$, or $\hat{\mathcal{P}}$ small enough to meet a reliability specification (e.g., one chance of instability in some larger number of realizations). One may object that the parameter distributions must be known or estimated for stochastic robustness analysis. However, if robustness estimates are strongly dependent on the statistics, then it is incumbent on the designer to know something about the statistics. Otherwise, the final control system may be either unduly conservative (at the expense of performance) or insufficiently robust in the face of real-world uncertainties.

II. Stochastic Root Locus

While it is not necessary to plot the eigenvalues of (4) to determine or portray stochastic robustness, stochastic root loci provide insight regarding the effects of parameter uncertainty on system stability. Consider, for example, a classical second-order system with characteristic equation

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0. \tag{8}$$

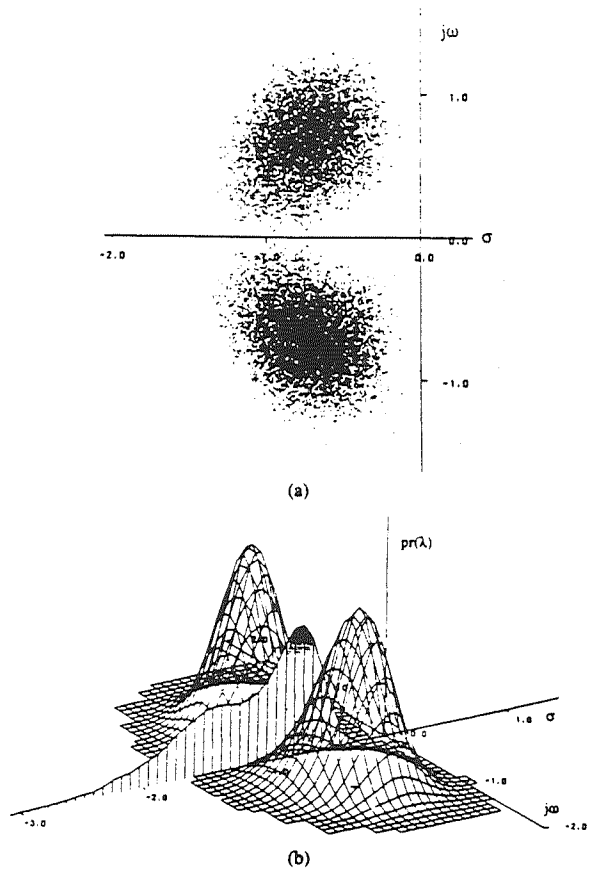


Fig. 1. Stochastic root loci of a second-order system with Gaussian damping ratio and natural frequency. $\zeta_0 = 0.707$, $\omega_{n0} = 1$; 50 000 Monte Carlo evaluations. The root density pr (λ) is given in units of roots/unit length along the real axis and roots/unit area in the complex plane. (a) Scatter plot. (b) Oblique three-dimensional representation.

Suppose that the damping ratio (ζ) and natural frequency (ω_n) are nominally 0.707 and 1, respectively, and that each may be a Gaussian-distributed random variable with standard deviation of 0.2. Root loci for individual parameter variations would follow classical configurations of root locus construction, with the heaviest density of roots in the vicinities of the nominal roots. If both ζ and ω_n are uncertain and uncorrelated (i.e., $p = [\zeta\ \omega_n]^T$), the possible root locations become "clouds" surrounding the nominal values [Fig. 1(a)]. Understanding of robustness issues can be gained by plotting the density of the roots in a third dimension above the root locus plot [Fig. 1(b)]. This can be done by simply dividing the s plane into subspaces (or "bins") and counting the number of roots in each bin as a sampled estimate of the *root density* ρ. The result is a multivariate histogram, with σ and ω serving as independent variables. Complex root bins are elemental areas, for which ρ_A is defined in units of roots/unit area. Real root bins are confined to the real axis; hence, ρ_L measures roots/unit length. The density of roots depicts the likelihood that eigenvalues vary from their nominal values, including branches on the real axis and in the right-half s plane for large enough variations of ζ and ω_n. Fig. 1 is based on 50 000 Monte Carlo evaluations, and numerical smoothing has been applied to account for sampling effects. An example of the histogram and cumulative distribution is given in Fig. 2. The probability-of-instability estimate ($\Sigma = 0$ on the cumulative distribution) is 8×10^{-5}.

When considering instability, distinction must be made between the number of *cases* with right-half plane roots and the number of *roots* in the right-half plane. For example, a third-order system can be unstable with 1, 2, or 3 roots in the right-half plane, yet N

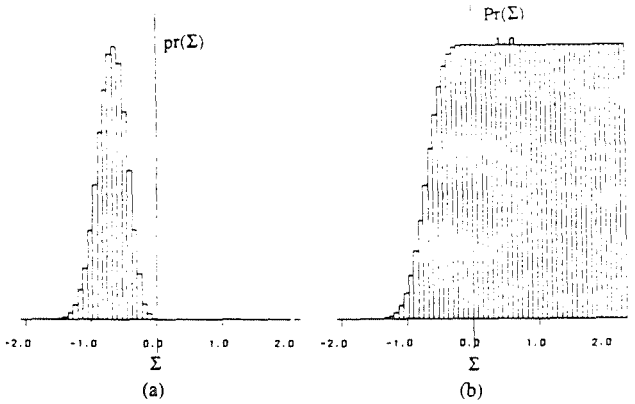


Fig. 2. (a) Histogram and (b) cumulative distribution for the second-order system with Gaussian damping ratio and natural frequency.

would be incremented by one in each case. A high-order system with real roots could be unstable with one or more roots in the same right-half plane bin. Again, N would be incremented by one, although the bin's ρ depends upon the number of roots it contains.

III. Confidence Intervals

Because it is an estimate, $\hat{\bar{P}}$ must have associated with it some notion of accuracy or relationship to the true underlying probability of instability $\bar{P}$. Confidence intervals relate $\hat{\bar{P}}$ to $\bar{P}$ by bounding the (unknown) true value with defined certainty. A confidence statement for $\bar{P}$ is

$$\Pr(L < \bar{P} < U) = 1 - \alpha \tag{9}$$

where (L, U) is the interval estimate (lower and upper bounds), $1 - \alpha$ is the confidence coefficient, and $\bar{P}$ lies within (L, U) with $100(1 - \alpha)\%$ confidence. Appropriate selection of the confidence coefficient and the number of evaluations determines the interval width.

The method used to compute confidence intervals depends on the underlying probability distribution of the variable in question. The probability of instability is a *binomial variable*, with the outcome of a trial taking one of two possible values (stable or unstable) for each Monte Carlo evaluation. $\bar{P}$ therefore has a binomial probability density and cumulative distribution given by

$$\text{pr}(x) = (J, x)\bar{P}^x(1 - \bar{P})^{J-x} \tag{10}$$

$$\Pr(X \le x) = \sum_{j \le x} (J, j)\bar{P}^j(1 - \bar{P})^{J-j} \tag{11}$$

where x is the number of occurrences of instability in J evaluations, and (J, x) is the binomial coefficient, $J!/x!(J - x)!$. The *binomial test* is applied to determine exact confidence intervals for binomial variables. The lower and upper confidence bounds are derived from the cumulative distribution and satisfy [19]

$$\Pr(X \le x - 1) = \sum_{j=0}^{x-1} (J, j)L^j(1 - L)^{J-j} = 1 - \frac{\alpha}{2} \tag{12}$$

$$\Pr(X \le x) = \sum_{j=0}^{x} (J, j)U^j(1 - U)^{J-j} = \frac{\alpha}{2} \tag{13}$$

where $1 - \alpha$ is the required confidence coefficient and X is the actual number of unstable cases after J evaluations ($X = J\hat{\bar{P}}$).

The validity of the Monte Carlo analysis depends on a number of simulation parameters: the number of eigenvalues computed, the number of varying parameters, and their probability distributions. However, by applying the binomial test, the derivation of explicit

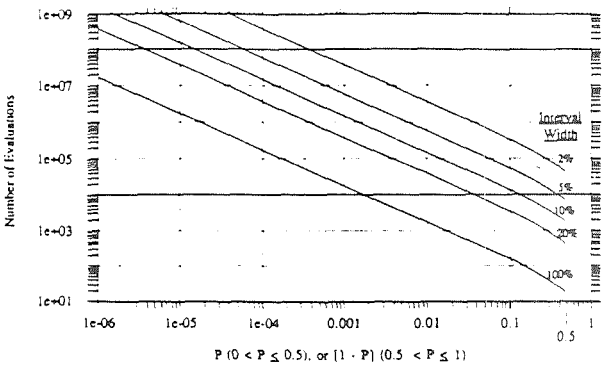


Fig. 3. Number of evaluations required for given confidence interval widths and confidence coefficient $1 - \alpha = 0.95$. Interval width is given as percent of $\bar{P}$. For $\bar{P} > 0.5$, the curves are symmetric, with the number of evaluations given by that for $1 - \bar{P}$.

relationships between simulation parameters and the required number of evaluations is avoided. Nevertheless, it offers a rigorous theory by which to calculate exact confidence intervals. The number of evaluations required for a specified confidence interval can be related to a single variable, $\bar{P}$, and this relationship is valid for any simulation parameters or application.

Fig. 3 presents the number of evaluations required for specified interval widths and a 95% confidence coefficient, given as a percentage of $\bar{P}$. An estimate $\hat{\bar{P}}$ based on a small number of evaluations can be used as the abscissa of Fig. 3 to forecast the total number of evaluations required for the desired interval width. A larger confidence coefficient shifts the curves of Fig. 3 to higher numbers of evaluations.

IV. Stochastic Robustness Example

An example of the application of stochastic robustness is based on the longitudinal dynamics and control of an open-loop-unstable aircraft. The Forward-Swept-Wing Demonstrator's aerodynamic center is forward of its center of gravity, resulting in static instability. Its stability matrix, control-effect matrix, and open-loop eigenvalues are

$$F = \begin{bmatrix} -0.02 & -0.3 & -0.4 & -32.2 \\ -0.0001 & -1.2 & 1 & 0 \\ 0 & 18. & -0.6 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{14}$$

$$G = \begin{bmatrix} -.04 & 35. \\ 0 & 0 \\ 0.2 & -0.2 \\ 0 & 0 \end{bmatrix} \tag{15}$$

$$\lambda_{1-4} = -0.1 \pm 0.057j, -5.15, 3.35. \tag{16}$$

The state components represent forward velocity, angle of attack, pitch rate, and pitch angle. The principal control surfaces are the canard control surface and the thrust setting. Possible uncertainties in aerodynamic and thrust effects as well as separate dynamic pressure (ρ and V) effects lead to a 12-element parameter vector (the remaining terms are kinematic, due to gravity, identically zero, or otherwise negligible)

$$p = [\rho\, V f_{11} f_{12} f_{13} f_{22} f_{32} f_{33} g_{11} g_{12} g_{31} g_{32}]. \tag{17}$$

Although velocity (V) and air-density (ρ) are essentially deterministic, including them as parameters gives the ability to look at flight condition perturbations around the nominal value and reduces correlation of the remaining parameters. ρ and V are modeled as uniform parameters, giving an indication of stochastic robustness over a

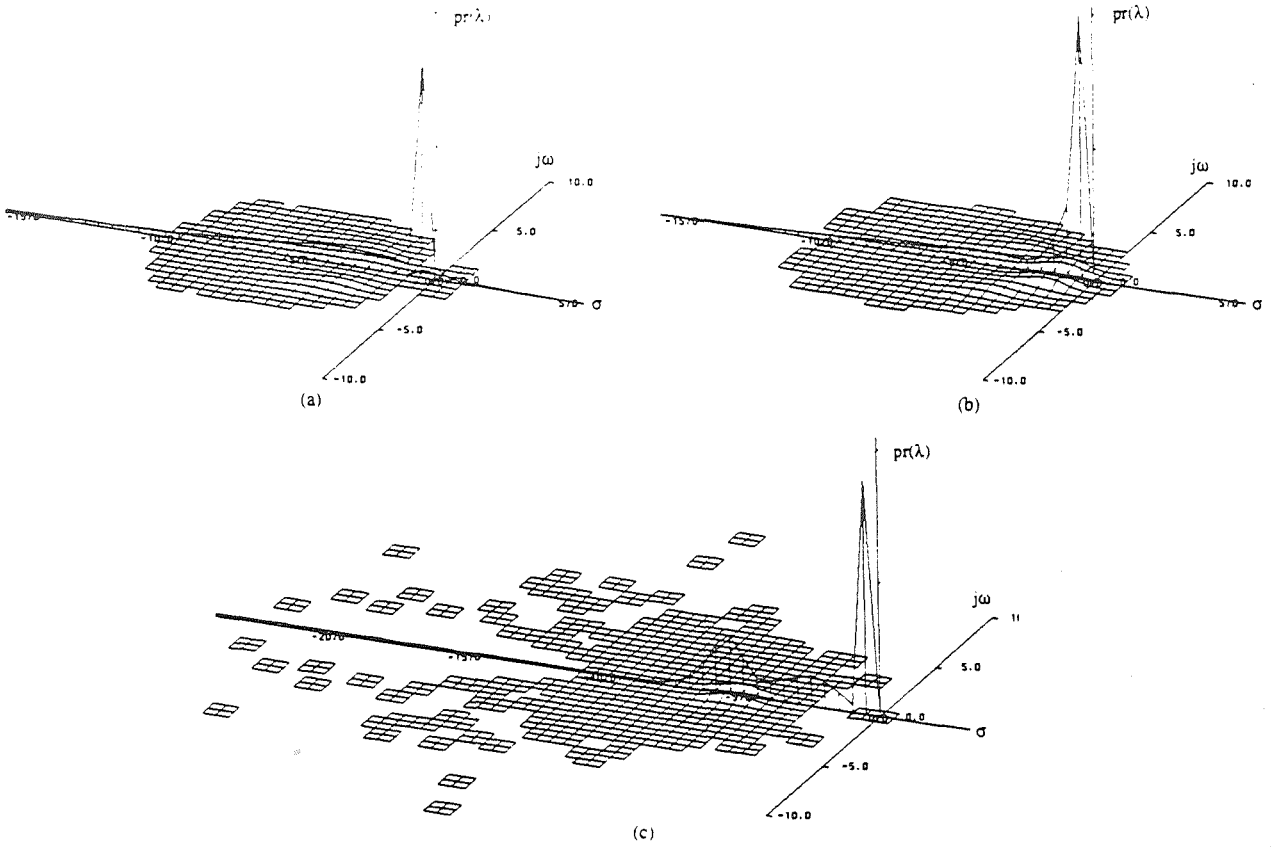


Fig. 4. Stochastic root loci for the Forward-Swept-Wing Demonstrator Aircraft with Gaussian parameters, Cases (a), (b), and (c). The root density pr (λ) is given in units of roots/unit length along the real axis and roots/unit area in the complex plane.

TABLE I
PARAMETERS FOR FORWARD-SWEPT-WING DEMONSTRATOR AIRCRAFT EXAMPLE

Case (a)	$C = \begin{bmatrix} 0.1714 & 130.26 & 33.165 & 0.364 \\ 0.984 & -11.387 & -2.968 & -1.133 \end{bmatrix}$	$Q = \text{diag}(1,1,1,0)$ $R = \text{diag}(1,1)$	$\lambda = -35.0, -5.14, -3.32, -0.0183$
Case (b)	$C = \begin{bmatrix} 0.0270 & 82.659 & 20.927 & -0.0638 \\ 0.0107 & -62.623 & -16.203 & -1.902 \end{bmatrix}$	$Q = \text{diag}(1,1,1,0)$ $R = 1000\,\text{diag}(1,1)$	$\lambda = -5.15, -3.36, -1.09, -0.0186$
Case (c)	$C = \begin{bmatrix} 0.1349 & 413.294 & 104.633 & -0.3191 \\ 0.0535 & -313.112 & -81.015 & -9.509 \end{bmatrix}$		$\lambda = -32.21, -5.15, -3.44, -0.01$

range of flight conditions. In terms of the elements p, F, and G are

$$F = \begin{bmatrix} \dfrac{-2gf_{11}}{V} & \dfrac{\rho V^2 f_{12}}{2} & \rho V_{13} & -g \\ \dfrac{-45}{V^2} & \dfrac{\rho V f_{22}}{2} & 1 & 0 \\ 0 & \dfrac{\rho V^2 f_{32}}{2} & \rho V f_{33} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (18)$$

$$G = \frac{\rho V^2}{2} \begin{bmatrix} g_{11} & g_{12} \\ 0 & 0 \\ g_{31} & g_{32} \\ 0 & 0 \end{bmatrix}. \quad (19)$$

Linear-quadratic controllers are designed according to three specifications: (a) $Q = \text{diag}(1,1,1,0)$, $R = \text{diag}(1,1)$; (b) $Q = \text{diag}(1,1,1,0)$, $R = \text{diag}(1000, 1000)$; and (c) the control gain matrix of Case (b) is multiplied by an arbitrary factor (5) to restore the closed-loop bandwidth to that of Case (a). The resulting control gain matrices and corresponding nominal closed-loop eigenvalues are given in Table I. These three cases have not been chosen to satisfy any particular performance criteria, but merely to demonstrate the impact of differing generalized design criteria on stochastic robustness. Furthermore, the designs are not meant to reflect acceptable control laws, as the high gains were purposely chosen to magnify robustness problems and to illustrate the application of stochastic robustness.

For illustration, ρ and V are $\pm 30\%$ uniform parameters, and the remaining elements of p are subject to independent 30% standard deviation Gaussian uncertainties. Fig. 4 shows the stochastic root loci for each case, based on 25 000 Monte Carlo evaluations. In each case, the eigenvalue near the origin is least affected by the parameter changes, and its peak dominates the distribution. In Cases (a) and (c), the eigenvalue farthest to the left (not shown in Fig. 4) has an enormous variance along the real axis. Interaction of roots

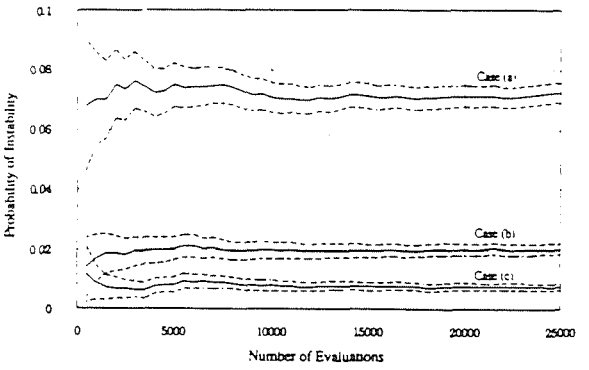


Fig. 5. 95% confidence intervals ($\alpha = 0.05$) based on the binomial test for the Forward-Swept-Wing Demonstrator Aircraft with Gaussian parameters, Cases (a), (b), and (c). Solid line represents $\hat{P}$ estimate, and dashed lines give confidence intervals.

around the origin causes instability. The corresponding probability-of-instability estimates and 95% confidence intervals are (a) 0.0724 (0.0692, 0.0756), (b) 0.0205 (0.0187, 0.0222), and (c) 0.0076 (0.0065, 0.0086). Robustness improves from Case (a) to (b) as control usage is restrained by high control weighting, and the ad hoc robustness recovery technique used in Case (c) gives additional improvement. Fig. 5 shows 95% confidence intervals and the $\hat{P}$ history with the number of evaluations. While confidence intervals for Cases (b) and (c) initially overlap, 25 000 evaluations are more than sufficient to rank the three cases in order of robustness.

For uniformly distributed parameters in $[0.7p, 1.3p]$, the extent of the parameter and eigenvalue distributions decreases substantially (Fig. 6). A comparison of Figs. 4 and 6 gives a better indication of the effects of Gaussian "tails" on the eigenvalue probability densities. $\hat{P}$ and confidence intervals for 25 000 evaluations are (a) 3.6E-4 (1.25E-4, 5.95E-4) and (b) and (c) zero (0, 1.48E-4). For 12-parameter binary distributions, 4096 evaluations produce exact results. $\hat{P}$ for each case is (a) 0.1191 and (b) and (c) zero.

Stochastic stability robustness is given as a function of the control design parameter ν, where the LQR control weighting matrix $R = \nu I_2$ (Fig. 7). Under the specified limits of parameter uncertainty, the distinction in stability and robustness versus design parameter statistics is apparent. Note that the qualitative result that $10 < \nu < 10^5$ provides the most robust designs is independent of the assumed probability distribution. In particular, the fact that the minimum-control-energy case ($\nu \to \infty$) represents the least robust design would not be obvious using standard robustness analysis techniques such as unstructured-singular-value analysis. While the exact relationship between parameter uncertainty and eigenvalue location is complicated, the increase in $\hat{P}$ beyond 10^5 is attributed to the shift in the closed-loop eigenvalues towards the imaginary axis. For example, at $\nu = 10^8$ the closed-loop eigenvalues are -5.15, -3.35, $-0.0104 \pm 0.057j$; the first two are almost identical to those of Case (b), while the complex pair is closer to the imaginary axis. The kind of results presented in Fig. 7 offer controller design insight and show nonobvious robustness characteristics.

V. Conclusion

Stochastic robustness offers a rigorous yet straightforward alternative to current metrics for control system robustness that is simple to compute and is unfettered by normally difficult problem statements, such as non-Gaussian statistics, products of parameter variations, and structured uncertainty. The approach answers the question, "How likely is the closed-loop system to fail, given limits of parameter uncertainty?" It makes good use of modern computational and graphic tools, and it is easily related to practical design considerations. The principal difficulty in applying this method to controlled systems is that it is computationally intensive; however, requirements are well within the capabilities of existing computers. The principal advantage of the approach is that it is easily implemented, and results have direct bearing on engineering objectives.

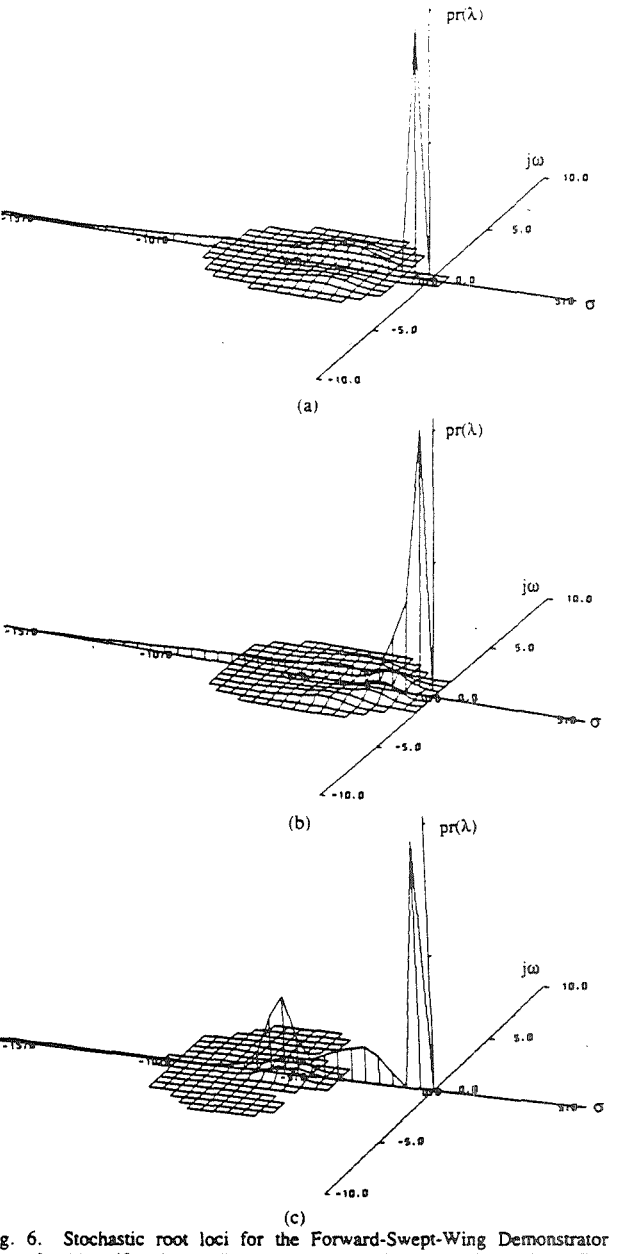


Fig. 6. Stochastic root loci for the Forward-Swept-Wing Demonstrator Aircraft with uniformly distributed parameters, Cases (a), (b), and (c). The root density pr (λ) is given in units of roots/unit length along the real axis and roots/unit area in the complex plane.

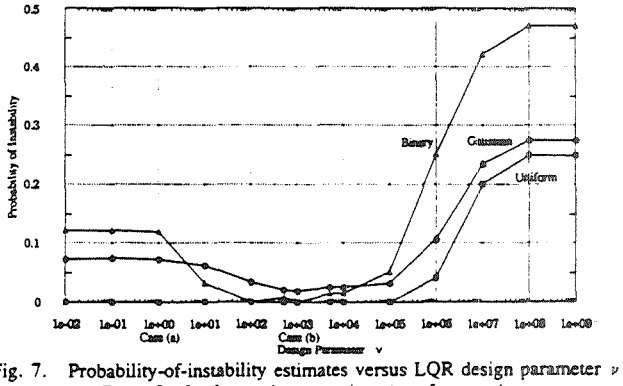


Fig. 7. Probability-of-instability estimates versus LQR design parameter ν ($R = \nu I_2$) for forward-swept wing aircraft example.

ACKNOWLEDGMENT

The authors are grateful to Prof. W. VanderVelde of M.I.T. for helpful discussions on this subject.

REFERENCES

[1] P. Dorato, Ed., *Robust Control*. New York: IEEE Press, 1987.
[2] N. R. Sandell, Jr., Ed., "Recent developments in the robustness theory of multivariable systems," Office of Naval Research, Rep. no. ONR-CR215-271-1F, Aug. 1979.
[3] N. A. Lehtomaki, N. R. Sandell, Jr., and M. Athans, "Robustness results in linear-quadratic-Gaussian-based multivariable control designs," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 1, pp. 75-93, Feb. 1981.
[4] J. C. Doyle, "Analysis of feedback systems with structured uncertainties," *IEE Proc.*, vol. 129, pt. D, no. 6, pp. 242-250, Nov. 1982.
[5] M. Tahk and J. L. Speyer, "Modeling of parameter variations and asymptotic LQG synthesis," *IEEE Trans. Automat. Contr.*, vol. AC-32, no. 9, pp. 793-801, Sept. 1987.
[6] R. K. Yedavalli and Z. Liang, "Reduced conservatism in stability robustness bounds by state transformation," *IEEE Trans. Automat. Contr.*, vol. AC-31, no. 9, pp. 863-866, Sept. 1986.
[7] I. Horowitz, "Quantitative feedback theory," *IEE Proc.*, vol. 129, pt. D, no. 6, pp. 215-226, Nov. 1982.
[8] E. Yaz, "Deterministic and stochastic robustness measures for discrete systems," *IEEE Trans. Automat. Contr.*, vol. 33, no. 10, pp. 952-955, Oct. 1988.
[9] H. J. Kushner, *Stochastic Stability and Control*. New York: Academic, 1967.
[10] R. F. Stengel, "Some effects of parameter variations on the lateral-directional stability of aircraft," *AIAA J. Guidance Contr.*, vol. 3, no. 2, pp. 124-131, Apr. 1980.
[11] ——, *Stochastic Optimal Control: Theory and Application*. New York: Wiley, 1986.
[12] R. F. Stengel and L. E. Ryan, "Stochastic robustness of linear control systems," in *Proc. 1989 Conf. Inform. Sci. Syst.*, Mar. 1989, pp. 556-561.
[13] ——, "Multivariate histograms for analysis of linear control system robustness," in *Proc. 1989 Amer. Contr. Conf.*, Pittsburgh, PA, June 1989, pp. 937-943.
[14] ——, "Application of stochastic robustness to aircraft control systems," in *Proc. 1989 AIAA Guidance, Navigation, Contr. Conf.*, Boston, MA, Aug. 1989, pp. 698-708.
[15] D. D. Siljak, *Nonlinear Systems, The Parameter Analysis and Design*. New York: Wiley, 1969.
[16] J. Ackermann, "Parameter space design of robust control systems," *IEEE Trans. Automat. Contr.*, vol. AC-25 no. 6, pp. 1058-1072, Dec. 1980.
[17] P. Putz and M. J. Wozny, "A new computer graphics approach to parameter space design of control systems," *IEEE Trans. Automat. Contr.*, vol. AC-32, no. 4, pp. 1058-1072, Apr. 1987.
[18] S. Boyd, V. Balakrishnan, C. Barratt, N. Khraishi, X. Li, D. Meyer, and S. Norman, "A new CAD method and associated architectures for linear controllers," *IEEE Trans. Automat. Contr.*, vol. 33, no. 3, pp. 268-283, Mar. 1988.
[19] W. J. Conover, *Practical Non-Parametric Statistics*. New York: Wiley, 1980.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1991	Conference Publication

4. TITLE AND SUBTITLE
Joint University Program for Air Transportation Research
1990-1991

5. FUNDING NUMBERS
WU 505-64-52-01

6. AUTHOR(S)
Frederick R. Morrell, Compiler

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23665-5225

8. PERFORMING ORGANIZATION REPORT NUMBER
L-17017

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001 and
Federal Aviation Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA CP-3131

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified–Unlimited

Subject Category 01

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes the research conducted during the academic year 1990-91 under the NASA/FAA sponsored Joint University Program for Air Transportation Research. The year end review was held at Ohio University, Athens, Ohio, June 20–21, 1991. The Joint University Program is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). Completed works, status reports, and annotated bibliographies are presented for research topics, which include navigation, guidance, and control theory and practice, aircraft performance, human factors, and expert systems concepts applied to airport operations. An overview of the year's activities for each university is also presented.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Avionics, Aircraft guidance, navigation, and control; Human factors; Neural networks; Expert systems	191
	16. PRICE CODE A09

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298(Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

NASA-Langley, 1991

National Aeronautics and
Space Administration
Code NTT

Washington, D.C.
20546-0001

Official Business
Penalty for Private Use, $300

SPECIAL FOURTH-CLASS RATE
POSTAGE & FEES PAID
NASA
Permit No. G-27

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return