

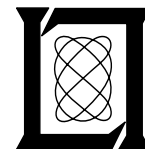
**Project Report
ATC-348**

Encounter Models for Unconventional Aircraft Version 1.0

**M.W. Edwards
M.J. Kochenderfer
J.K. Kuchar
L.P. Espindle**

10 April 2009

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS



Prepared for the Department of Homeland Security under Air Force Contract FA8721-05-C-0002.

Approved for public release; distribution is unlimited.

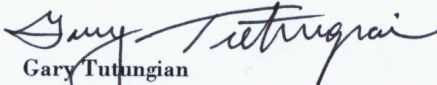
This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Department of Homeland Security under Air Force Contract FA8721-05-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER


Gary Tutungian
Administrative Contracting Officer
Acquisition Enterprise Division

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document
when it is no longer needed.

**Massachusetts Institute of Technology
Lincoln Laboratory**

**Encounter Models for Unconventional Aircraft
Version 1.0**

*M.W. Edwards
M.J. Kochenderfer
J.K. Kuchar
L.P. Espindle
Group 42*

Project Report ATC-348

10 April 2009

Approved for public release; distribution is unlimited

Lexington

Massachusetts

This page intentionally left blank.

ABSTRACT

Airspace encounter models, covering close encounter situations that may occur after standard separation assurance has been lost, are a critical component in the safety assessment of aviation procedures and collision avoidance systems. Of particular relevance to Unmanned Aircraft Systems (UAS) is the potential for encountering general aviation aircraft that are flying under Visual Flight Rules (VFR) and are not in contact with air traffic control. In response to the need to develop a model of these types of encounters, Lincoln Laboratory undertook an extensive data collection and modeling effort involving more than 96,000 unconventional aircraft tracks. The outcome of this effort was nine individual models encompassing ultralights, gliders, balloons, and airships. The models use Bayesian networks to represent relationships between dynamic variables and to construct random trajectories that are statistically similar to those observed in the data. The intruder trajectories can be used in fast-time Monte Carlo simulations to estimate collision risk.

The model described in this report is one of three developed by Lincoln Laboratory. A correlated encounter model has been developed to represent situations in which it is likely that there would be air traffic control intervention prior to a close encounter. The correlated model applies to encounters involving aircraft receiving Air Traffic Control (ATC) services and with transponders. An encounter with an intruder that does not have a transponder is uncorrelated in the sense that it is unlikely that there would be prior intervention by air traffic control. The uncorrelated model described in this report is based on global databases of pilot-submitted track data. This work is a follow-on to an uncorrelated conventional model developed from recorded radar tracks from aircraft using a 1200 transponder code. A byproduct of this encounter modeling effort was the extraction of feature distributions for unconventional aircraft. This provides an extensive collection of unconventional aircraft behavior in the airspace.

This page intentionally left blank.

ACKNOWLEDGMENTS

This work is sponsored by the Air Force under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

This report is the result of research and development sponsored by the United States Department of Homeland Security Science and Technology Directorate, the United States Air Force 303rd Aeronautical Systems Wing (303 AESW/XRX), and the Department of Defense Unmanned Aircraft System Airspace Integration Joint Integrated Product Team.

The authors greatly appreciate the support and assistance provided by John Appleby, Kevin “Spanky” Kirsch, David Thomasson, Joseph Kielman, and Merv Leavitt from the Department of Homeland Security, Maj. Luke Cropsey, Lt. Col. Kent Tiffany, Tony Salmonson, and Richard Graeff from the Air Force, and Laurence Newcome from MTSI.

The authors appreciate the support from Hill Air Force Base, including Jeff Richardson, Steven Schimmelpfennig, Richard Whitlock, Lt. Han Saydam, Lt. Tanuxay Keooudom, James Evans, and TSgt. Christopher Cosper.

Finally, the authors would also like to thank Lincoln Laboratory staff member Daniel Griffith for contributions to the airspace encounter model effort.

This page intentionally left blank.

TABLE OF CONTENTS

	Page
Abstract	iii
Acknowledgments	v
List of Illustrations	ix
List of Tables	xi
1. INTRODUCTION	1
1.1 Encounter Types	2
1.2 Unconventional Aircraft Types	3
1.3 Data Sources	7
1.4 Process Overview	8
2. MODELS	11
2.1 Model Variables	11
2.2 Initial Distribution	12
2.3 Transition Distribution	13
3. ESTIMATION	15
3.1 Track Quality Assurance	15
3.2 Track Smoothing	17
3.3 Interpolation	18
3.4 Ground Track Removal	18
3.5 Feature Extraction	19
3.6 Statistics Extraction	22
4. SUMMARY	25
A. SAFETY EVALUATION	27
A.1 Estimating NMAC Probability	27

A.2	Estimating Encounter Rate	29
A.3	Estimating Aircraft Density	30
B.	MODEL PARAMETERS	33
C.	UNCONVENTIONAL MODEL FEATURES	37
D.	TRACK LOCATION SUMMARY	43
E.	BAYESIAN NETWORKS	47
E.1	Definition	47
E.2	Sampling	47
E.3	Parameter Learning	47
E.4	Structure Learning	49
F.	VEHICLE MODELS AND NETWORK CANDIDATES	51
F.1	Network Candidates	51
F.2	Sampling Boundaries	68
	References	69

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Model hierarchy.	6
2	Modeling and simulation process overview.	9
3	Bayesian networks for the initial and transition distributions.	14
4	Estimation process flow.	16
5	Data quality assurance example.	17
6	Interpolation example.	18
7	Ground track removal example.	19
8	Unpowered heavier-than-air feature distributions.	21
9	Example of feature extraction.	22
10	Model convergence test and quantity of track data for each model.	24
C-1	Unpowered heavier-than-air altitude feature distribution.	38
C-2	Unpowered heavier-than-air detailed feature distributions.	39
C-3	Balloon feature distributions.	40
C-4	Skydiver feature distributions.	41
C-5	Airship feature distributions.	42
D-1	Western Hemisphere track location summary.	43
D-2	Eastern Hemisphere track location summary.	44
D-3	United States track location summary.	45

This page intentionally left blank.

LIST OF TABLES

Table No.		Page
1	Encounter model types	3
2	Cut points used for paraglider feature quantization	20
3	Paraglider sufficient statistics for airspeed given altitude layer	23
C-1	Bounds used for feature normalization	37
F-1	Summary of model graphical networks	51

This page intentionally left blank.

1. INTRODUCTION

One of the main challenges to integrating unmanned aircraft into the National Airspace System (NAS) is the development of systems that are able to sense and avoid local air traffic. If designed properly, these collision avoidance systems could provide an additional layer of protection that maintains or even enhances the current exceptional level of aviation safety. However, due to their safety-critical nature, rigorous assessment is required before sufficient confidence can exist to certify collision avoidance systems for operational use. Evaluations typically include flight tests, operational impact studies, and simulation of millions of traffic encounters to explore the robustness of the collision avoidance system. Key to these simulations are so-called encounter models that describe the statistical properties of the encounters in a way that represents what actually occurs in the airspace.

Each encounter generated from an encounter model specifies the initial positions and orientations of two aircraft and the nominal dynamic maneuvers (not affected by a collision avoidance system) that may occur leading up to the closest point of approach. Identical encounter situations are simulated with and without a collision avoidance system to determine the relative benefit of the system. Knowledge of the rates at which encounter situations occur in the NAS can be used to estimate the rate of near mid-air collisions per flight hour, thereby quantifying risk.

In these models, encounter situations are abstracted in the sense that there is no consideration of an explicit location or local airspace structure (e.g., airways, metering fixes, approach paths). Rather, the encounters represent what may be statistically expected to occur over the lifetime of a given system. If desired, a particular altitude layer can be specified and aircraft behavior will be representative of that particular region. Additionally, the flight path of one aircraft can be constrained to focus, for instance, on a particular departure profile or flight condition. It should be noted that the models cover approximately one minute before closest point of approach and so are not appropriate for large-scale air traffic impact studies (e.g., to examine sector loading or conflict rates); the focus here is on events in which loss of separation has already occurred between two aircraft and collision avoidance becomes paramount.

One system that has been rigorously tested using encounter models is the Traffic alert and Collision Avoidance System (TCAS). As part of the TCAS certification process in the 1980s and 1990s, several organizations tested the system across millions of simulated close encounters and evaluated the risk of a near mid-air collision (NMAC, defined as separation less than 500 ft horizontally and 100 ft vertically) [1–4]. This analysis ultimately led to the certification and U.S. mandate for TCAS equipage on larger transport aircraft. More recently, Eurocontrol and the International Civil Aviation Organization (ICAO) performed similar simulation studies for European and worldwide TCAS mandates [5, 6].

The design of a collision avoidance system requires a careful balance of preventing collisions and ensuring a low rate of unnecessary maneuvers. This balance is strongly affected by the types of encounter situations to which the system is exposed. It is therefore important that simulated encounters are representative of those that occur in the airspace. Hence, tremendous effort has been made by various institutions since the early 1980s to develop encounter models based on

radar data [1, 3, 7–10]. The primary contribution of this report is the application of encounter modeling to unconventional aircraft using a Bayesian statistical framework and which uses recent radar and global positioning system (GPS) receiver data. The advantage of applying a Bayesian framework is that it optimally leverages available data to produce a model that is representative of actual operations.

1.1 ENCOUNTER TYPES

The encounters covered by this model are those involving aircraft in the final stages before a collision. It is assumed that prior safety layers—e.g., airspace structure, Air Traffic Control (ATC) advisories or vectors—have failed to maintain standard separation distances between aircraft. The model is therefore useful in describing the types of situations that need to be addressed by a collision avoidance system, but will not address other separation aspects such as ATC communications or coordination.

Because they are by far the most likely to occur, only pairwise (two-aircraft) encounters are explicitly discussed in this model. If required, the probability of multiple aircraft encounters can be determined by analyzing the local traffic density. Random multi-aircraft encounters can be constructed, if desired, by combining several trajectories generated from the model presented here.

There are two fundamental types of encounters. In the first, both aircraft involved are cooperative (i.e., have a transponder) and at least one is in contact with ATC. It is then likely that at least one aircraft will receive some notification about the traffic conflict and begin to take action before a collision avoidance system gets involved. We term this type of encounter “correlated” because the trajectories of each aircraft may involve maneuvers that are correlated to some degree due to this prior intervention. The second type of encounter we term “uncorrelated” and involves at least one noncooperative aircraft (i.e., not using a transponder) or two aircraft flying under Visual Flight Rules (VFR) without flight following (i.e., using a transponder Mode A code of 1200). In these encounters, it is unlikely that air traffic control would become involved prior to the close encounter; rather the two aircraft must rely solely on visual acquisition (or some other collision avoidance system) at close range to remain separated. Such encounters tend to be uncorrelated since there is no coordinated intervention prior to the close encounter: the assumption is that the two aircraft blunder into close proximity. A complete evaluation of unmanned systems will require analysis using both correlated and uncorrelated models.

Aircraft in uncorrelated encounters can be further separated into conventional and unconventional. We broadly define conventional aircraft as powered propulsion vehicles having a fixed wing and seating more than one person while unconventional aircraft are generally smaller, usually seat one person but no more than two passengers, and are generally unpowered. Furthermore, unconventional aircraft rarely carry a transponder. This report focuses on uncorrelated unconventional encounter modeling. A general aviation study completed by the Federal Aviation Administration (FAA) found that more than 71%, 85%, 98% of light-sporting aircraft, gliders, and lighter-than-air vehicles, respectively, do not carry transponder equipment [11]. The ultralight category was not explicitly included in this study, but we estimate that transponder equipment would be no more than that for gliders. The specific aircraft types modeled are defined in Section 1.2.

TABLE 1
Encounter model types

C = correlated, U = uncorrelated conventional, X = uncorrelated unconventional			
		Aircraft of Interest	
		Discrete	VFR
Discrete		C	C
VFR		C	U
Noncooperative conventional (fixed-wing powered aircraft)		U	U
Noncooperative unconventional (balloon, glider)		X	X

Analysis presented in the preceding uncorrelated conventional document demonstrated that 1200-code tracks are a proper surrogate for certain classes of noncooperative traffic in the National Airspace System, but there are other categories of noncooperative targets for which they are not suitable [12]. For example, most balloons, ultralights, and gliders do not fly in a similar manner to transponder-equipped aircraft squawking 1200.

The challenge in developing models for unconventional aircraft such as balloons, ultralights, and gliders is the lack of high-quality radar data. Instead of radar data, this model uses GPS trajectories to build a model. It is not possible to extract the airspace density for unconventional aircraft because trajectories are voluntarily submitted by the individual pilot. Because knowledge of the relative airspace density, or rate at which unconventional aircraft occur in the airspace, is unknown, and the aircraft flight characteristics are strongly dependent on the aircraft type, separate models for each aircraft type must be created.

Table 1 shows which encounter model to use depending on the types of aircraft involved in the encounter. There are three types of encounter models shown in this Table: correlated (C), uncorrelated conventional (U), and unconventional (X). In terms of aircraft types, Discrete indicates a cooperative aircraft using a non-1200 Mode A transponder code, and VFR denotes a cooperative aircraft using the 1200 Mode A transponder code. Only the uncorrelated unconventional model (bold X) is discussed in this report; the other models are described in other reports [12, 13].

1.2 UNCONVENTIONAL AIRCRAFT TYPES

Unconventional aircraft are broadly separated into heavier-than-air and lighter-than-air categories. Heavier-than-air aircraft stay aloft by a rigid or flexible wing creating lift while lighter-than-air aircraft remain buoyant due to an enclosed pocket of gas or air which is less dense than the surrounding air. These categories are further divided into individual aircraft types for modeling purposes. Brief descriptions of the aircraft which were modeled are presented below and Figure 1 illustrates this hierarchy. Powered light-sporting aircraft are implicitly included within the uncorrelated conventional model. The aircraft modeled may not be physically constrained to the ultralight

or light-sporting categories as defined by the FAA, but they are categorized as such based on typical flight envelope. For example, paragliders are not observed to travel faster than 60 kt, which is also the regulation for maximum airspeed for an ultralight. In total, nine aircraft models were created and are denoted in bold italics below.

- **Heavier-than-Air**

- *Unpowered Ultralight*: Climbing is achieved in localized sources of air updrafts such as thermals, travel no faster than 60 kt, and seat one to two people attached to the exterior of the lifting body. These aircraft are usually foot launched and a significant portion of flight is devoted to climbing.
 - ***Paraglider***: The pilot is attached to a flexible lifting wing by a harness. The wing shape and relative position of the pilot is similar to the parachute stage of a skydiving flight.
 - ***Flexible Wing Hang Glider***: The lifting wing resembles a kite and the pilot is usually placed in a head forward position.
 - ***Rigid Wing Hang Glider***: Similar to the flexible wing hang glider, but the wing is formed of a rigid plastic or composite.¹
- *Powered Ultralight*: The primary method for propulsion and climbing is an open or ducted fan behind the pilot. This additional component is the deviation from the unpowered ultralight.
 - ***Paramotor***: These aircraft are similar to paragliders with the addition of a powered fan for propulsion. Although powered types of hang gliders and other ultralights exist, no data were obtained. Paramotors are sometimes referred to as powered paragliders.
- *Unpowered Light-Sporting*: Similar to unpowered ultralights, climbing is usually achieved in thermals, but have considerably faster cruise speeds, up to 120 kt. They similarly seat one to two people, but generally have heavier airframes and enclosed cockpits.
 - ***Glider***: Also commonly referred to as sailplanes, they often have large wing spans and wing aspect ratios with an enclosed cockpit. Some gliders carry intermittent powered propulsion for a short period of climbing flight, but they are usually towed to an altitude of 5000–10,000 ft above ground level (AGL).

- **Lighter-than-Air**

- *Balloon*: Balloons have an envelope filled with hot air or gas placed above a smaller gondola, capsule or basket. They have no means of directly controlling horizontal motion. Therefore, we assume no horizontal component of true airspeed.

¹Rigid wing and flexible wing hang gliders are differentiated as separate vehicles by the Fédération Aéronautique Internationale (FAI). The FAI designations for flexible and rigid hang gliders are FAI1 and FAI5, respectively and are used interchangeably.

- ***Weather Balloon:*** The gas envelope of weather balloons are typically 10–30 ft in diameter, contain either helium or hydrogen gas and are used primarily for atmospheric research. They are launched twice daily at hundreds of sites in the NAS and can reach altitudes of more than 100,000 ft. These balloons do not carry people, but may provide a significant hazard to the safety of other aircraft if not avoided. Weather balloons, and the availability of their tracks, are the most prominent of the high-altitude balloon class and are thus considered a surrogate for all uncontrolled high-altitude balloons.
- ***Hot Air Balloon:*** Hot air balloons typically have an envelope filled with warm air. The larger hot air balloons can be more than 100 ft in diameter and can carry upwards of a dozen people. Balloons do exist with gas as the primary buoyancy source, but no data were found, thus hot air balloons are assumed a surrogate.
- ***Airships:*** They are differentiated from balloons by being able to maneuver in the horizontal direction. Buoyancy is provided by a large envelope, up to 800 ft in length, filled with helium or hydrogen. Large control surfaces and ducted fans provide control and propulsion, respectively. Airships are usually seen loitering above larger cities.

- **Other**

- ***Skydiver:*** The only aircraft type modeled without an airframe, they are lifted to an altitude reaching 15,000 ft or higher by a powered aircraft, fall for up to 30 seconds at or near terminal velocity and then release a parachute to slow their fall to the ground. Although the parachute itself is similar to the paraglider wing, the skydiver does not try to climb and the parachute is simply used to slow the decent rate. No distinction is made between nominal skydivers and skydivers with wing suits—lifting structures attached to the free-falling skydiver.

There are several aircraft types and airspace objects that could be considered uncorrelated unconventional aircraft. These other unconventional objects are not included in the model for one of two reasons. First, objects that remain at an altitude very low to the ground, less than approximately 200 ft. Second, objects that are too small (less than five pounds) or rare do not pose a significant hazard to larger manned or unmanned vehicles. Objects which are very low to the ground include kites, moored balloons, manned jet packs and parasailors—paraglider type aircraft attached to a water vehicle via a cable. The small and/or rare objects include birds, small recreational unmanned rockets and falling debris.

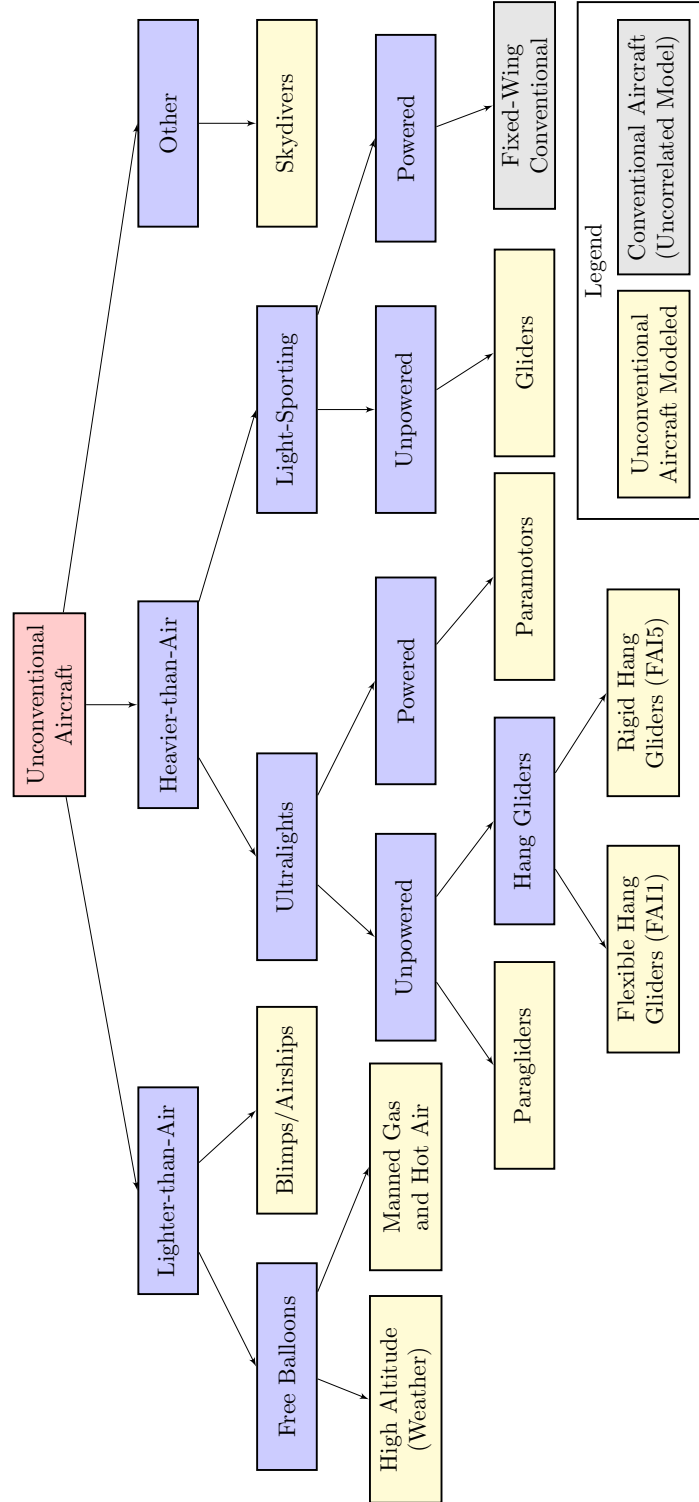


Figure 1. Model hierarchy with aircraft modeled by the uncorrelated unconventional model. The uncorrelated conventional model encompasses light-sporting fixed-wing conventional aircraft.

1.3 DATA SOURCES

1.3.1 Global Position System Receiver Data

With the growing popularity of low-cost GPS receivers and servers, many pilots have started uploading their aircraft tracks online. These websites, many of them attracting pilots internationally, were originally created for recreational use or for competition with other pilots. For example, the Online Contest (www.onlinecontest.org) is a database of archived GPS tracks sponsored by the FAI for competition among unpowered aircraft [14].² Certified GPS receivers and a track file format standard are required to ensure quality and prevent tampering. The update rates for these raw GPS tracks ranges from 1 s to 30 s and the precision of the GPS position reports is assumed to be within a relative precision bound of 30 ft [15].

There are two caveats when using the data. First, the tracks may not represent the total population of data for these vehicles as they are often submitted for competition. We assume that these data are representative of the aircraft platforms modeled as the data correspond very closely with several expected performance parameters such as cruise airspeed and minimum sink speed. Second, the tracks are submitted voluntarily, so although we can extract the position for each of the tracks, it is not possible to extract the airspace density for these vehicles. We assume that the performance and characteristics of each aircraft are independent of the location of the track and therefore, several international tracks are used together with those in the NAS. Appendix D includes a summary of the location of the tracks both internationally and within the continental United States although it should not be used for collision avoidance system analysis.

1.3.2 Radar Data

The radar data used to help build the airship model comes from the 84th Radar Evaluation Squadron (RADES) at Hill AFB, Utah. RADES receives radar data from FAA and Department of Defense sites throughout the United States. They maintain continuous real-time feeds from a network of sensors, including long-range ARSR-4 radars around the perimeter of the United States and short-range ASR-8, ASR-9, and ASR-11 radars in the interior. Radar ranges vary from 60 to 250 NM. Individual radar tracks were manually extracted based on prior knowledge (i.e., airships over populated areas and sporting events). In this report, radar data was only used to create the airship model as these aircraft often use transponders, especially when located above populated areas.

1.3.3 Weather Balloon Data

Regularly published weather balloon data allow for a separate, higher quality data source than the voluntarily contributed tracks used to create the other models. Although tracks can be

²Several other online sources were used to compile the models. For paraglider, hang glider and paramotor tracks, the Paragliding Forum was the source (www.paraglidingforum.com). Glider tracks were provided by the Soaring Server (soaringweb.org). Several hot air balloon tracks were used from the general GPS database at Every Trail (everytrail.com). Skydiver tracks were provided by Paralog, a skydiver performance competition website (www.paralog.net).

found for other high atmosphere gas balloons, weather balloon data is readily available. The data used to build this model consists of 6-second averaged measurements which are recorded daily by the National Weather Service upper-air observing system. There are upwards of 100 operational observing sites, each of which launch approximately two balloons per day. These 100 sites are spaced by no more than 200 miles [16]. The whole quantity of data, which consists of decades of tracks, is not necessary to create a robust model as only the vertical profile is required. We selected 12 sites spaced equally by about 400 miles for an eight week span. This resulted in approximately 650 tracks, each lasting about one hour. The vertical position is measured by pressure measurements at 250 Hz or by slower rate GPS measurements. These are then published at averaged 6-second intervals.

1.4 PROCESS OVERVIEW

Figure 2 diagrams the steps involved in processing position data to build the encounter model and generate encounters for simulation. The high-level approach is to model nominal aircraft trajectories using Markov models represented by Bayesian networks (Section 2). The first step in constructing a Markov model involves extracting features, such as turn rate and vertical rate, from the data. From these features, sufficient statistics are collected to describe the distribution over maneuvers and other properties of trajectories. Bayesian model selection methods are used to search for the best network structure that represents the observed data (Section 3). The best network structure is then selected and the associated sufficient statistics are obtained to generate new trajectories that are representative of those observed. Finally, the trajectories are used in a dynamic simulation to evaluate encounters between aircraft with or without a collision avoidance system. Appendix A discusses using the encounter model for large-scale safety analysis and collision risk estimation.

This work largely follows the uncorrelated conventional method and reference to the prior uncorrelated model documentation will be made where appropriate for brevity [12]. Specifically, discussions of trajectory sampling and generation, as well as encounter simulation are not included within this document. The unconventional model is composed of nine individual aircraft models. As it would be cumbersome to show examples and excerpts from each model, we selected representative examples. The model structures and cutpoints for each model are illustrated in Appendix F.

A digital representation of the sufficient statistics and model structures described in this report are available from Lincoln Laboratory. Additionally, a Matlab software package is available to generate random trajectories using the data tables.

Throughout this report we use the standard units used in aviation. In particular, altitudes are in feet, positions are in nautical mile coordinates, vertical rates are in feet per minute, turn rates are in degrees per second, airspeeds are in knots (true airspeed), and accelerations are in knots per second. Time is reported in seconds.

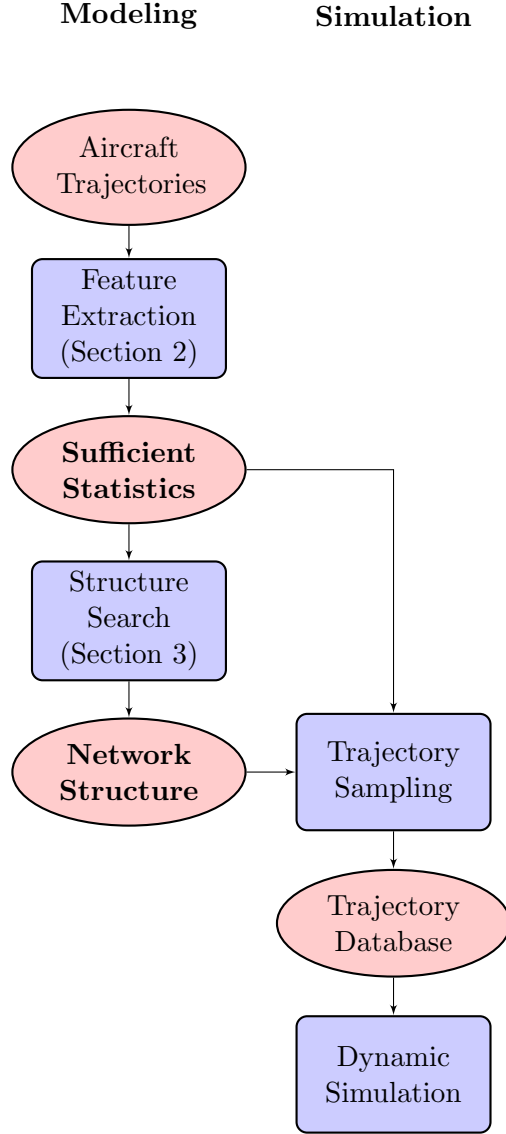


Figure 2. Modeling and simulation process overview. The sufficient statistics and network structure (in bold) are the main elements provided as part of this work.

This page intentionally left blank.

2. MODELS

We model nominal flight, i.e., flight without avoidance maneuvering, using a Markov process represented by a dynamic Bayesian network. A Markov process is a stochastic process where the probability distribution over future states is conditionally independent of past states given the present state. In other words, one only needs to know the present state to predict the next state.

The states in the model specify how the position, altitude, and airspeed of each aircraft involved in an encounter change over time. In particular, each state specifies a vertical rate \dot{h} , turn rate $\dot{\psi}$, and airspeed acceleration \dot{v} . Given an initial airspeed v , position (x, y, h) , heading ψ , vertical rate \dot{h} , altitude layer L , and airspace class A , one can determine from the model how the aircraft trajectories evolve over the course of the encounter.

One way to represent a Markov model is with an exhaustive state-transition matrix that specifies the probability of transitioning between all pairs of states. Unfortunately, the number of independent parameters (that is, independent probabilities) required to define the matrix grows super-exponentially with the number of variables defining the model. The more independent parameters there are in the model, the more data one needs to properly estimate their values. However, by using dynamic Bayesian networks, we can leverage conditional independence between some variables to greatly reduce the number of parameters. We can learn the structure of the dynamic Bayesian network by maximizing the posterior probability of the network structure given the data.

Appendix E provides the necessary background on Bayesian networks. The remainder of this section defines an encounter in further detail, introduces the variables used to describe an encounter, and presents the modeling structure.

2.1 MODEL VARIABLES

There are five variables in the uncorrelated unconventional encounter model:

- **Altitude layer L :** For 7 of the 9 aircraft models, airspace is divided into four altitude layers in a process similar to prior encounter models developed by Eurocontrol and to the uncorrelated conventional model. The first layer spans from 500 to 1200 ft AGL to capture aircraft in the traffic pattern or performing low-level maneuvers. The second layer spans a transition zone from 1200 to 3000 ft AGL, the cruise altitude where the hemispheric rule begins. The third layer spans from 3000 ft AGL to 5000 ft AGL covering a mix of low-altitude enroute and maneuvering aircraft. The fourth layer includes airspace above 5000 ft AGL and would cover most enroute VFR traffic. The models encompassing skydivers and weather balloons were shifted in order to accommodate the larger observed altitude range. Appendix F presents the boundaries for each model.
- **Airspeed v :** We model true airspeed and allow it to vary during flight.
- **Acceleration \dot{v} :** We allow airspeed acceleration to vary every second.

- **Turn rate $\dot{\psi}$:** Turn rate is permitted to change every second in our model.³
- **Vertical rate \dot{h} :** The vertical rate is permitted to change at every second.

An airspace class variable is not included as in the uncorrelated conventional model because international tracks were used. Because many of the variables are closely related due to physical constraints and flight characteristics (e.g., turn rate and vertical rate) is it important to properly represent correlations in the model. Independently sampling from distributions for turn rate and vertical rate, for example, would miss these important relationships. The remainder of this section explains how to model joint probability distributions over these variables to ensure proper consideration of correlations.

To generate an encounter, we first randomly sample from the joint distribution over the encounter variables to define the encounter geometry and initial conditions. Second, we use a Markov model to determine how the dynamic variables, i.e., turn rate, vertical rate, and airspeed acceleration, evolve during the course of the encounter. There are two corresponding separate probability distributions in the model: an initial distribution to set up an encounter situation, and a transition distribution to describe how the dynamic variables specifying the trajectories of the aircraft evolve over time. We allowed for a different graphical structure for each aircraft type as suggested by the data.

2.2 INITIAL DISTRIBUTION

The aircraft encounter model represents the distribution over the initial values of \dot{h} , $\dot{\psi}$, \dot{v} , v , and L as well as the time-varying history of \dot{h} , $\dot{\psi}$, and \dot{v} during the course of an encounter. Probabilistic relationships between these variables are represented using a Bayesian network. An example of such a relationship is the one between turn rate and vertical rate. Without properly capturing this dependency and other important dependencies, unrealistic situations may be generated, e.g., involving aircraft with simultaneously high climb rates and high turn rates. Initial position and altitude is determined through a separate process.

A Bayesian network (e.g., Figure 3a) includes a series of nodes represented by rectangles and directed arcs represented by arrows. Each node corresponds to a particular variable that may take on one of several discrete values with associated probabilities. Certain variables, such as altitude layer, are naturally quantized into a few discrete values. Continuous variables, such as vertical rate or turn rate, are described by a series of discrete bins from which a continuous value is later selected. Within each node, then, is a description of the possible values a variable can take and the probability that each value will occur. The directed arcs describe how the probabilities of one variable depend on other variables. Arrows leading into a particular node denote which parent nodes must first be evaluated in order to select the value of the given node. For example, referring to Figure 3a, the probabilities for node v depend on the value of node L ; the probabilities for node \dot{v}

³Approximately two-thirds of the skydiver tracks were obtained with vertical profiles and processed airspeed, so turn rate could not be estimated for these tracks. Therefore, the skydiver turn rate table is based on fewer samples than the other tables.

depend on the values of nodes v and L . In the latter case, for instance, there would be a conditional probability table describing the probability of each possible value of \dot{v} jointly conditioned on the values of v and L : $P(\dot{v} \mid L, v)$.

Because there are many possible ways variables can be connected in the model, it is necessary to use a quantifiable scoring process to evaluate the quality of each candidate network. For this model, we used a Bayesian scoring process (Appendix E) to evaluate different Bayesian network structures and choose a structure that optimally represents the trajectories we observed. Figure 3a shows the optimal structure for the glider and paraglider initial distributions. Appendix F shows other candidate network structures and their scores.

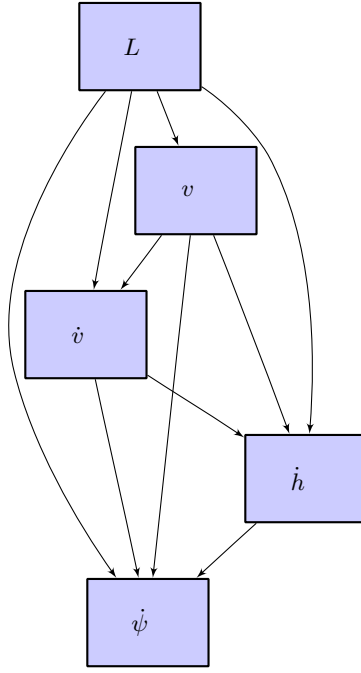
Given a structure, sufficient statistics extracted from data, and a Bayesian prior, we then sample from the Bayesian network to produce initial altitude layers, vertical rates, turn rates, airspeeds, and accelerations that are representative of those found in the data. The nodes and directed arcs used in the structural diagram show the order in which this sampling occurs. For example, based on the structure in Figure 3a, to determine the initial state of the aircraft we first randomly determine an altitude layer L . Once the altitude layer has been determined, we then randomly select airspeed v , and so on. Alternately, if we were interested in a specific case, we could assign outright an altitude layer for a particular study and then randomly select values for the remaining variables.

2.3 TRANSITION DISTRIBUTION

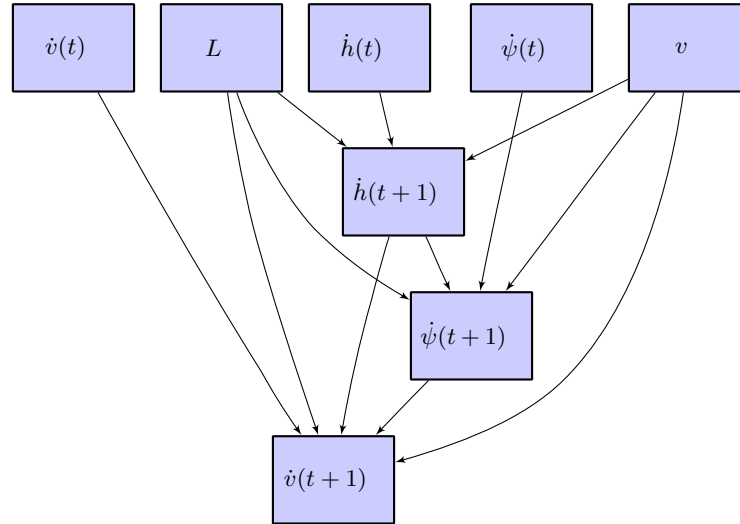
We use a separate Bayesian network to model how the variables \dot{h} , $\dot{\psi}$, and \dot{v} evolve over time. In this network, we have a first layer that represents the state of the system at the present time step and a second layer that represents the state of the system at the next time step. There are dependencies between layers and within the second layer. Such a two-layer temporal Bayesian network is known as a dynamic Bayesian network [17, 18]. Parameter and structure learning in dynamic Bayesian networks is similar to regular Bayesian networks (Appendix E).

Figure 3b shows the structure used for the transition distribution. Again, we chose the highest-scoring network structure among our candidate network structures (see Appendix F). Given a structure, sufficient statistics extracted from data, and a prior, we can then sample from the Bayesian network to determine the next vertical rate, turn rate, and airspeed acceleration command.

In general, time steps in dynamic Bayesian networks may be of any duration, but for our modeling effort we chose steps of 1 s. Shorter time steps allow for more frequent variations in airspeed, vertical rate, and turn rate, but they require more computation per unit of simulation time. Time steps of 1 s balance maneuver complexity with computation and are also appropriate given typical timescales of dynamic maneuvers. A complete trajectory is constructed by updating the aircraft state in 1 s intervals. Within each interval, the three derivative variables \dot{h} , $\dot{\psi}$, and \dot{v} are treated as target values and held constant. A dynamic model is used to compute and update the aircraft state at each time step based on these piecewise-constant target values. The dynamic model is independent of this encounter model and not discussed in this report.



(a) Initial distribution.



(b) Transition distribution

Figure 3. Bayesian networks representing the variable dependency structure for the paraglider and glider initial and transition distributions. Networks for each model and all network candidates are presented in Appendix F.

3. ESTIMATION

At a high level, the modeling process involves processing a large volume of track data and extracting features of aircraft trajectories. Features include static variables that specify an encounter (such as altitude layer or initial airspeed) and multiple, dynamic variables that describe aircraft motion (such as turn rate, vertical rate, and airspeed acceleration every second). To aid in data processing, each feature was quantized into several bins and counts were taken of the frequency with which each bin was occupied by the data. Based on these counts (called sufficient statistics in the Bayesian network literature), probability tables were then constructed so that each feature can be randomly generated such that the overall trajectories are representative of those observed in the data.

Accordingly, the inputs to this process are raw reports (latitude, longitude, altitude and time) and the outputs are probability tables that specify the likelihood that a given feature will take on a value within a bin corresponding to each table cell. This section describes the processing required to transform GPS tracks into sufficient statistics that may be used to model uncorrelated encounters. Although this section will focus on GPS track processing, it is similar to the radar processing method which is used to create the uncorrelated conventional model. The airship model was formed from radar tracks, and thus followed the method for radar processing. Figure 4 outlines the multiple-stage feature extraction process; the yellow stages indicate the additional processes required for GPS track processing while the blue stages show the original steps required for radar track processing used in the uncorrelated conventional model. Gray stages are processes that were unneeded for this model and were removed.

3.1 TRACK QUALITY ASSURANCE

As the tracks for building this model originated at several different locations and with varying levels of quality, a process for assuring the quality of tracks was defined. A small portion of the tracks had large time gaps between measurements. We also found that several tracks were created using GPS receivers which varied update rate as a function of speed. For example, if the GPS receiver was not moving, such as on the ground before takeoff, the update rate would be very sparse (such as 30–60 s between measurements), while the update rate could be as high as one measurement per second during high dynamic periods. After reviewing several tracks, we determined that small time gaps and large displacements per time period would not necessarily result in a low quality track. We defined several thresholds for determining the quality of a track.

- **Measurement Dropouts:** Large gaps in time were removed because track features could not be determined, except for the assumed average, during a gap. Tracks with more than two instances of gaps exceeding 60 s were thrown out. Tracks with one or two instances were split into multiple tracks. Similarly, tracks with fewer than two instances of negative time differences between sequential measurements were split into multiple tracks for further processing. Negative time differences were usually found near the beginning of a track which is caused by a GPS clock update following the first GPS fix.

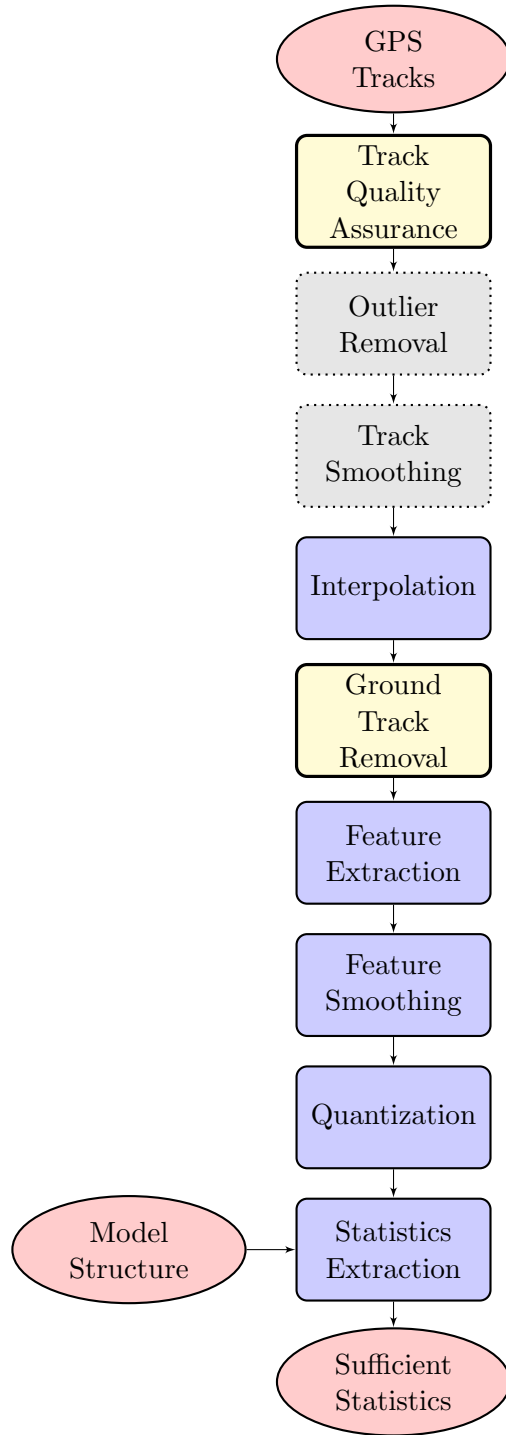


Figure 4. Estimation process flow. Gray nodes are processes that were removed and yellow nodes are processes that were added to the conventional processing method.

- **Measurement Interval:** Tracks with measurement update rates greater than 12s were removed to ensure at least one measurement during high rate turns and loops. We found that many of the unpowered vehicles have a maximum turn rate of 20 degrees per second. With fewer than two measurements during these turns, there is not a reliable method for recreating the necessary features in a turn, specifically turn rate and airspeed. Also, 12s is the scan rate for the en route radars used in the uncorrelated conventional model.
- **Outlier Removal:** A threshold was set on the maximum allowable positional displacement per unit time. If the raw true airspeed was greater than 120 kt (200 ft/s) for ultralights or 240 kt (400 ft/s) for skydivers and gliders, the track was split around these large displacements. Up to five of these large displacements were allowed per track. The upper limits were chosen as it is very unlikely for the aircraft modeled to travel faster than these limits.

Figure 5 illustrates an example of this process. In this instance, there are two instances where there is significant time difference between measurements. The filter splits the track into three separate tracks around these deviations for processing and statistical extraction.

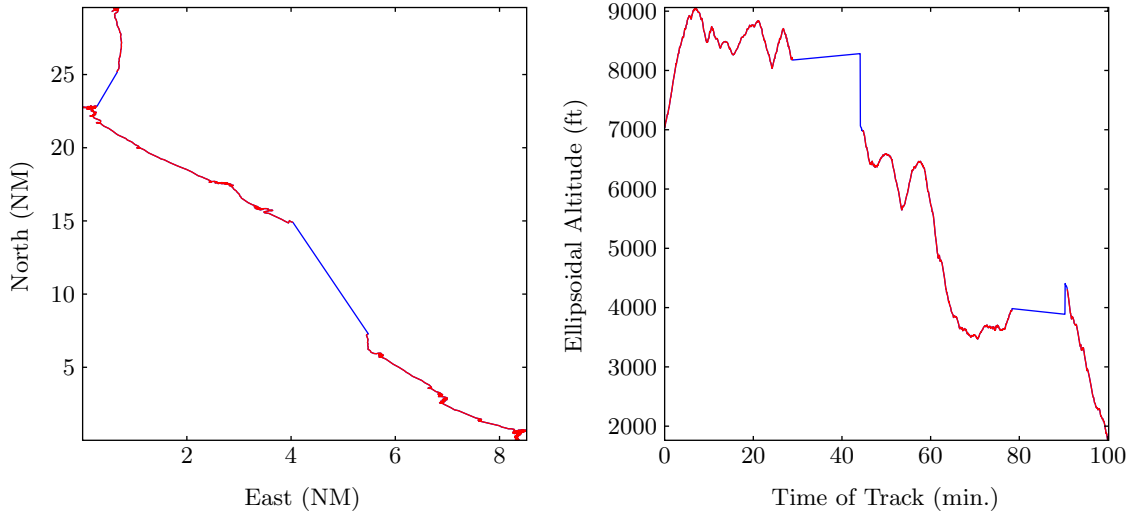


Figure 5. Data quality assurance example for a flexible hang glider (FAI1). Blue lines show the track before preprocessing. Red lines show the multiple tracks after the data quality assurance process.

3.2 TRACK SMOOTHING

The next step after outlier removal in the uncorrelated conventional model was track smoothing. The track smoothing method reduced the noise in the radar position measurements. However, we did not employ this method of smoothing to create this model for two reasons. First, high turn rates would be smoothed such that the aircraft would appear to be moving slower (by as much as 15 kt). This did not affect the turn rate as the same turn was completed in the same time period. Second, GPS data are usually more precise than radar and quantized barometric pressure data.

3.3 INTERPOLATION

The time interval between measurements in the data is typically much longer than the 1 s time step of the dynamic Bayesian network. The GPS tracks ranged in period from 1 to 30 seconds with the majority in the range 3 to 12 seconds. Hence, we interpolate to estimate the parameters in our dynamic Bayesian network. We chose a piecewise-cubic Hermite interpolation scheme that preserves monotonicity and shape [19]. An example of this interpolation method is illustrated in Figure 6.

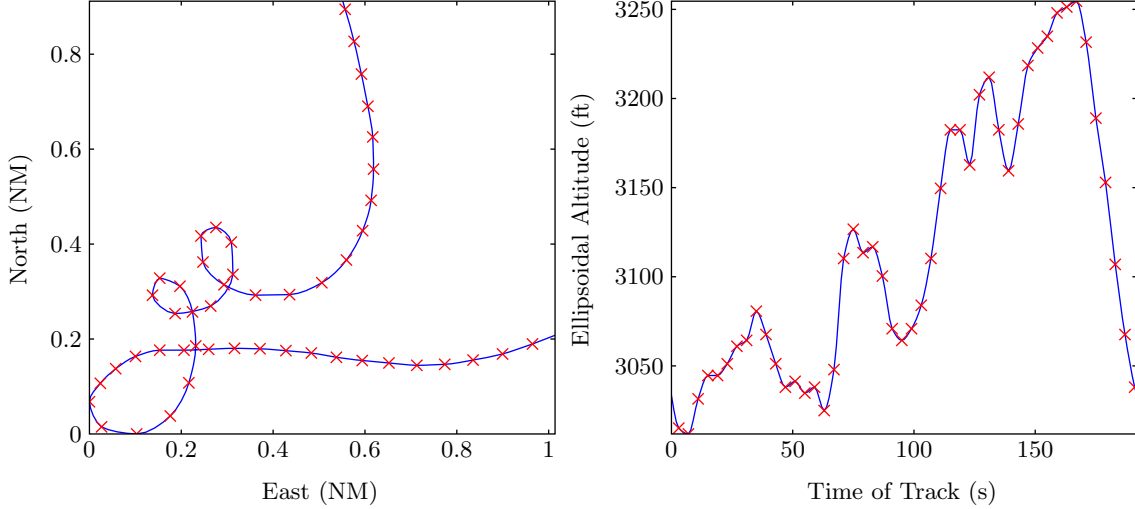


Figure 6. Interpolation example for a glider. Red crosses indicate original data points while the blue curve shows the interpolation.

3.4 GROUND TRACK REMOVAL

The uncorrelated conventional model processing excluded measurements below 500 ft AGL to avoid including portions of a track while the vehicle was on the ground. If these measurements are included in the statistical model, they cause the feature distributions to be unnecessarily skewed toward zero airspeed, acceleration, turn rate, altitude layer and vertical rate. We did not implement a minimum altitude threshold as a large portion of the unconventional tracks were observed to be under 500 ft (see Figure C-1). As many of the aircraft types did not otherwise have sufficient data to provide a statistically significant model, we kept the tracks under 500 ft provided they were not on the ground. However, we observed that many unconventional aircraft pilots would keep recording their position after landing and also start recording before takeoff, so these track portions needed to be removed.

We detected ground tracks by implementing a minimum threshold on airspeed at the first and last 150 s of a track. If the airspeed was below 4 kt for 30 or more seconds, we estimated that the aircraft was on the ground. We removed the portion of the track from the track beginning or

end to the point where the airspeed exceeded 4 kt. As the last step prior to feature extraction, we removed tracks with durations less than 100 s. An example of this process is shown in Figure 7.

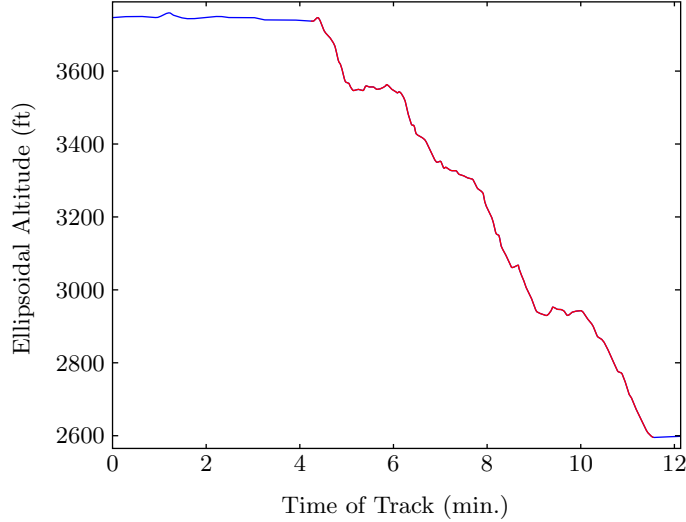


Figure 7. Ground track removal process for a paraglider. Blue lines show the track before preprocessing and the portion of the track that was removed. Red lines show the track after the ground track removal process.

3.5 FEATURE EXTRACTION

Feature extraction involves converting an interpolated track into sequences of quantized bins representing altitude layers, airspeeds, vertical rates, turn rates, and accelerations.

- **Altitude layer:** Altitude above ground level determines the altitude layer in the model. We estimate altitude by subtracting an estimate of ground elevation from altitude. We substituted pressure altitude for GPS altitude if GPS altitude was corrupt or did not exist. Our estimates of ground elevation come from Digital Terrain Elevation Data (DTED) provided by the National Geospatial-Intelligence Agency (NGA). We use DTED Level 0, which has post spacing of 30 arcseconds (approximately 900 meters).

- **Airspeed:** The true airspeed at time t is estimated by

$$v(t) = \sqrt{(x(t+1) - x(t))^2 + (y(t+1) - y(t))^2 + (h(t+1) - h(t))^2}.$$

- **Vertical rate:** The vertical rate is estimated from the interpolated altitudes reports. The vertical rate at time t is given by $\dot{h}(t) = h(t+1) - h(t)$.
- **Turn rate:** We first compute the heading along the interpolated track. The heading at time t is given by $\psi(t)$ and corresponds to the direction from $(x(t), y(t))$ to $(x(t+1), y(t+1))$. To compute the turn rate at time t , we find the acute change in heading between $\psi(t)$ and

$\psi(t+1)$. Turns to the right have positive turn rates, and turns to the left have negative turn rates.

- **Acceleration:** To find the acceleration at a particular point, we average the change in airspeed per unit time looking forward one time step and looking back one time step.

We then smooth the extracted features and use the following general formula to transform a raw trajectory $(t_1, \mathbf{x}_1), \dots, (t_n, \mathbf{x}_n)$ to a smoothed trajectory $\mathbf{y}_1, \dots, \mathbf{y}_n$,

$$\mathbf{y}_i = \frac{\sum_j w(t_i, t_j) \mathbf{x}_j}{\sum_j w(t_i, t_j)}, \quad (1)$$

where $w(t_i, t_j)$ is a weighting function that monotonically decreases as the difference between t_i and t_j increases. For the weighting function, we use the following definition based on a Gaussian kernel with standard deviation σ ,

$$w(t_i, t_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t_i - t_j)^2}{2\sigma^2}\right). \quad (2)$$

For vertical rate, turn rate, airspeed, and acceleration, we set σ to 5s, 5s, 5s and 10s respectively. We chose these numbers large enough so that noise is removed from the measurements but low enough so that the underlying properties of the maneuvers are not lost.

In order to be modeled by a discrete Bayesian network, it is necessary to quantize the features. We quantize continuous values by defining a sequence of cut points c_1, \dots, c_n . Values less than c_1 are in the first bin, values greater than c_n are in the $(n+1)$ th bin, and values in the half-open interval $[c_{i-1}, c_i)$ are in the i th bin. The cut points we used for quantization are listed in Table 2.

TABLE 2

Cut points used for paraglider feature quantization

	Cut Points
L	1200, 3000, 5000
v	5, 10, 15, 20, 25, 30, 35
\dot{v}	-0.2, -0.05, 0.05, 0.2
\dot{h}	-750, -450, -150, 150, 450, 750
$\dot{\psi}$	-12.5, -7.5, -2.5, 2.5, 7.5, 12.5

The cut points were chosen to capture the variation of the features as shown in the histograms in Figure 8. We observed large variations in features among aircraft types; feature distributions for airships, skydivers, and balloons are presented in Appendix C. After observing the flight characteristics of unconventional aircraft such as paragliders and gliders, we found that these unconventional

aircraft behave quite differently than conventional aircraft. In general, they tend to fly slower and maneuver more often than the conventional aircraft. As a result, a wider vertical rate distribution is observed even though conventional powered aircraft are capable of a greater range of vertical rates. Figure 9 shows the result of feature extraction for a paraglider track.

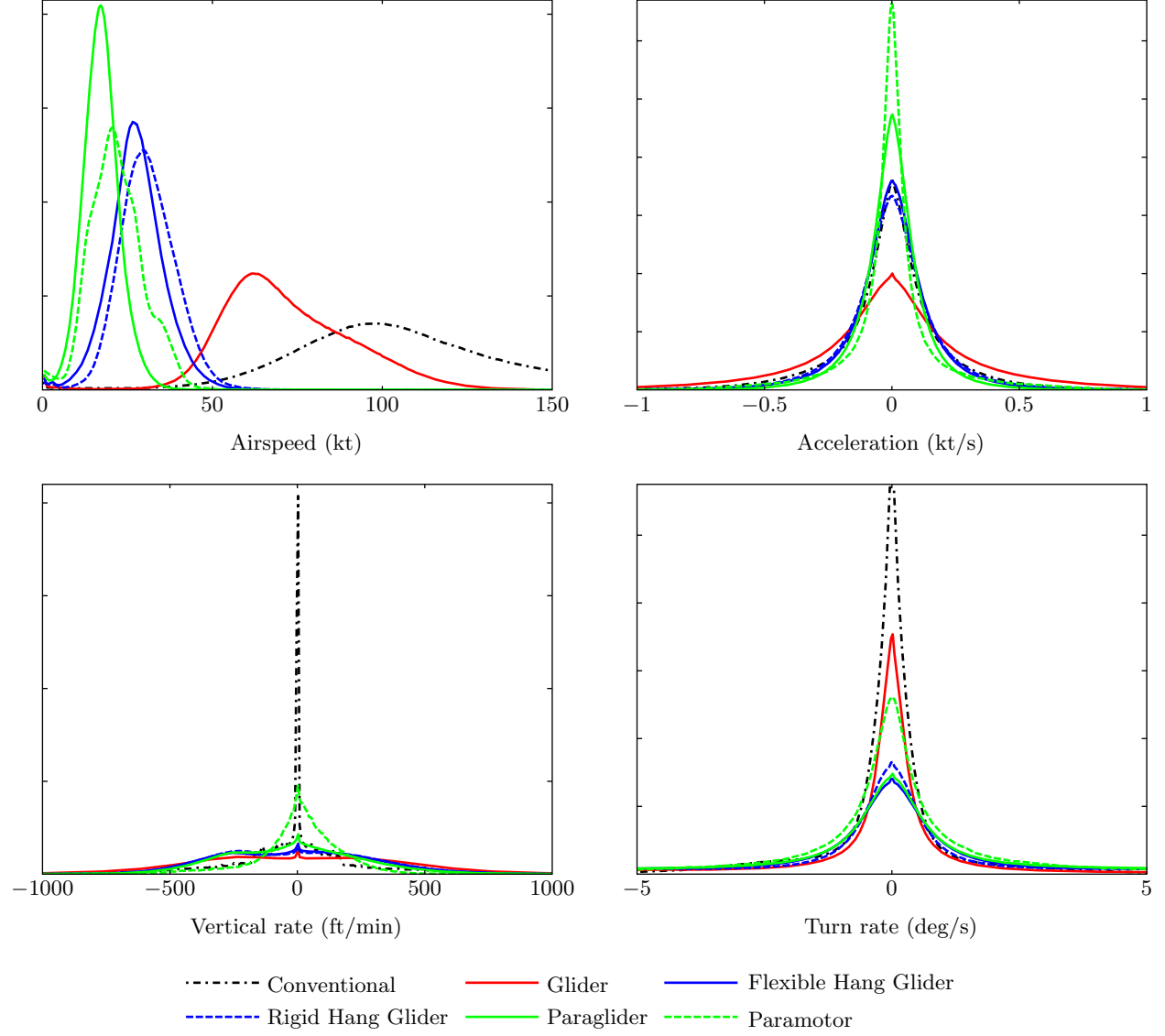


Figure 8. Unpowered heavier-than-air relative frequency feature distributions. The conventional aircraft category corresponds to the features of the uncorrelated conventional model which is shown for reference. A figure detailing the tails of these distributions along with altitude distributions is shown in Appendix C.

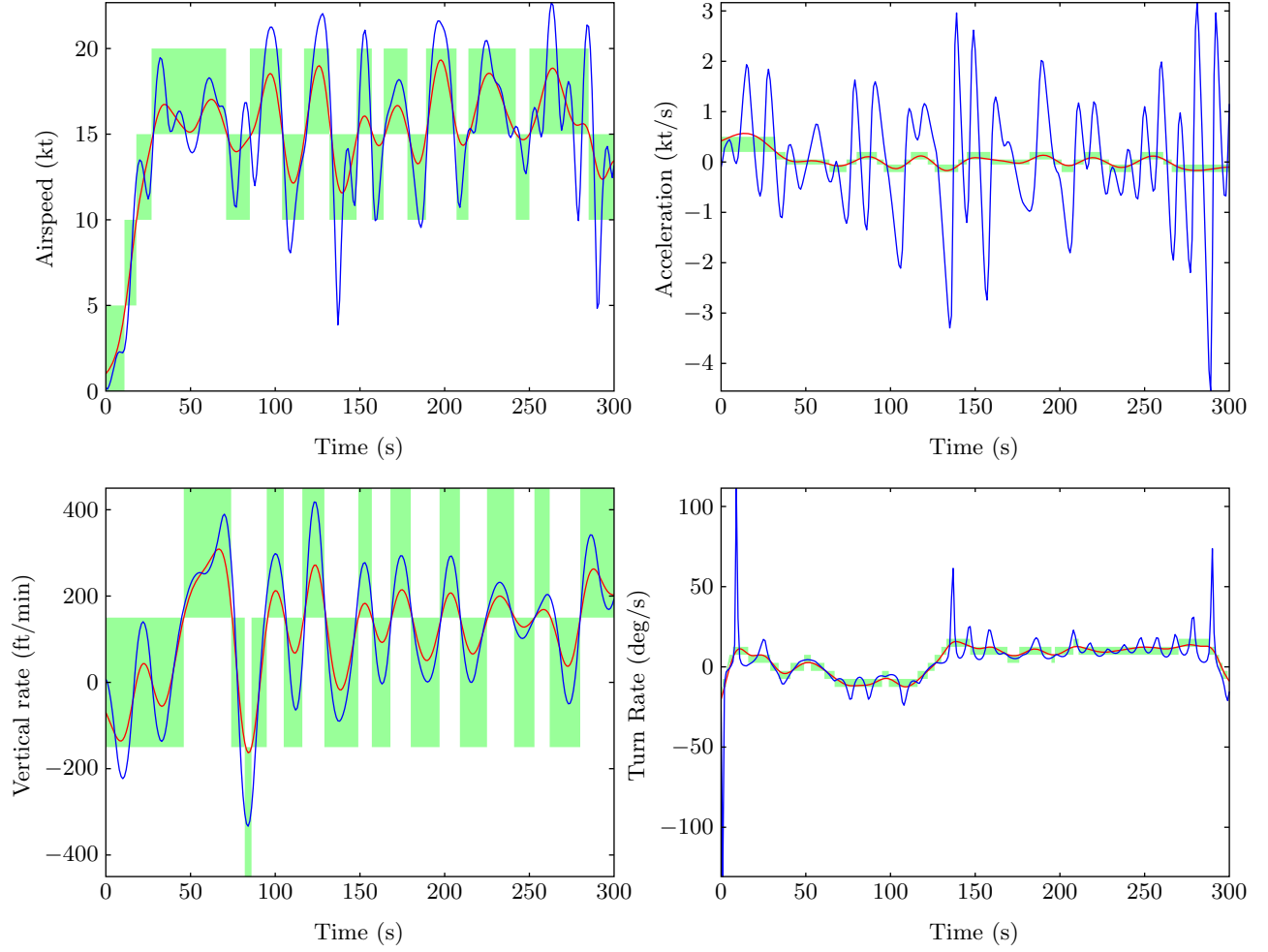


Figure 9. Example of feature extraction for a paraglider track. Blue lines show features before smoothing, and red lines show features after smoothing. The green blocks show the bins to which the features are assigned.

3.6 STATISTICS EXTRACTION

With structures for the initial and transition distributions and the quantized features from a set of tracks, we are able to collect the sufficient statistics to estimate the parameters for our model. For the two Bayesian networks, the sufficient statistics are simply the counts N_{ijk} of the various features (see Appendix E).⁴ Appendix B describes the sufficient statistics extracted from the data.

The counts N_{ijk} are then compiled into a separate table for each variable. For example, the table for airspeed v (which is itself dependent on altitude layer L) is shown in Table 3. Each cell entry in the table represents the counts N_{ijk} for that combination of parent bin L and bin of v .

⁴The counts are called *sufficient statistics* because they provide a summarization of the data that is sufficient to compute the posterior distribution from the prior. For an introduction to Bayesian statistics, see [20].

An electronic file available from Lincoln Laboratory contains all of the data required to construct these tables.

TABLE 3

Paraglider sufficient statistics for airspeed given altitude layer— $N(v \mid L)$

L (ft)	v (kt)							
	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	[35, 45]
[0, 1200)	3879844	14999543	47000749	51323443	16311008	2706447	440731	207174
[1200, 3000)	248110	3961586	23066750	44032743	25083482	6317280	1187811	244667
[3000, 5000)	41565	804999	7151851	18778961	16126654	6319400	1503668	298611
[5000, ∞)	10053	160906	1503769	4783136	5537742	3165563	1097369	296640

3.6.1 Model Convergence

To ensure that a sufficiently large data set was used, Figure 10 illustrates the convergence of the Bayesian network parameters as additional data are added. We test convergence by calculating the maximum percent deviation for each count in the statistical tables with the addition of data. Different amounts of data were collected for each aircraft type so we compared the last $1/32^{nd}$ of additional data—e.g., approximately 2679 hours for paragliders. Although not visible in Figure 10, paramotor, skydiver, hot air balloon and airship models contained 116, 198, 10 and 6 flight hours respectively. We set an upper bound on the convergence to 0.25%. This bound was set to approximately three times the convergence of the transition network for the uncorrelated conventional model. Any convergence above 0.25% was termed “unconverged”.

Only the airship model did not converge for the initial and transition network with this definition. Therefore, the airship model should be used with caution and we suggest that only the initial network be used—e.g., $dh(t + 1) = dh(t)$. This assumption is not unwarranted as the features for airships were observed to change more slowly than for the other vehicles. This convergence measure does not directly correlate to the amount of data collected. For example, the paramotor transition model converged to a greater extent than the flexible hang glider model. With fewer data, less connected graphical structures were favored by the scoring scheme. Thus, a model with fewer data and connections may converge quicker than a model with more data and a more complex structure.

To summarize, this section describes the process used to construct a model of uncorrelated aircraft trajectories based on GPS data. GPS data (latitude, longitude, altitude, and time) were processed through multiple filtering and tracking stages and then digested into a collection of conditional probability tables associated with Bayesian network structures. Each table cell represents the probability of a particular feature taking on a value within a certain quantized bin.

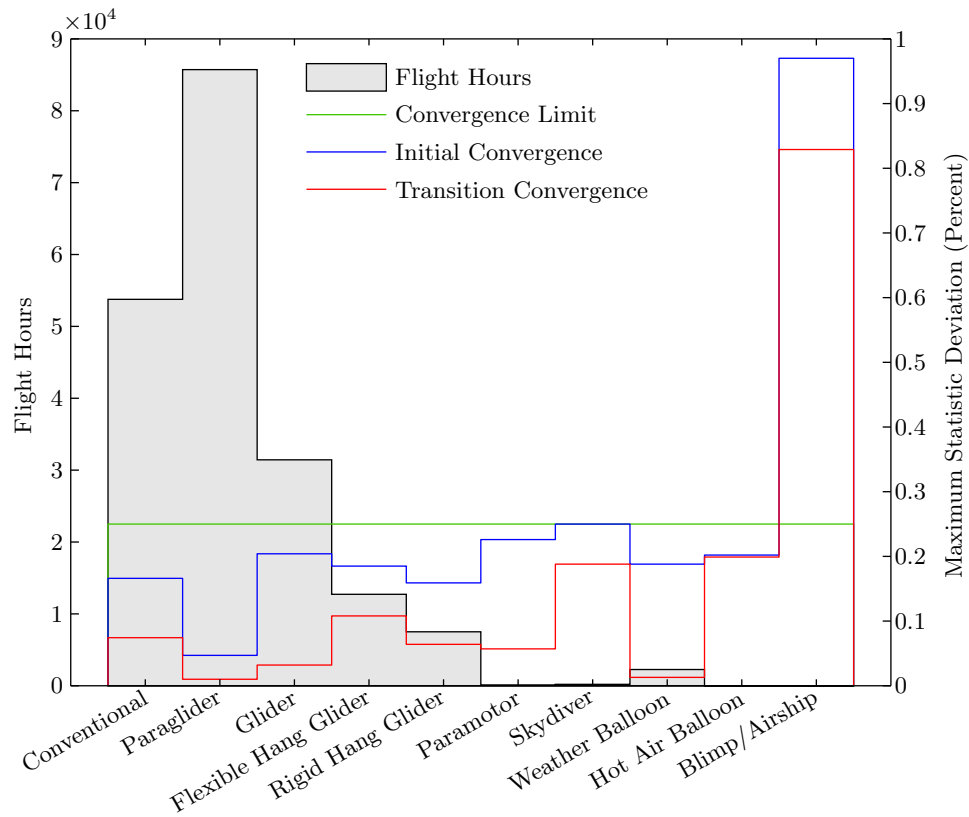


Figure 10. Model convergence test and quantity of track data for each model.

4. SUMMARY

This report presents a new approach to encounter modeling that is inclusive to unconventional aircraft. The approach involves modeling the dynamics of aircraft state based on a Markov model where the probability of the next state depends only upon the current state. One way to represent a Markov model is with an exhaustive state-transition matrix that specifies the probability of transitioning between all pairs of states. Unfortunately, the number of independent parameters required to define the matrix grows super-exponentially with the number of variables defining the model. The more independent parameters there are in the model, the more data needed to properly estimate their values. However, by using dynamic Bayesian networks, we leverage conditional independence between some variables to greatly reduce the number of parameters. We learn the structure of the dynamic Bayesian network by maximizing the posterior probability of the network structure given the data.

The work presented in this report focused on a model where the trajectories of the aircraft involved in the encounter are independent of each other prior to intervention by a collision avoidance system, human or automated. It extends the recent uncorrelated conventional model by including aircraft such as balloons, ultralights and gliders. This model assumes that aircraft blunder into close proximity without prior intervention. A separate, correlated encounter model has also been developed for aircraft under air traffic control, where intervention may impact the way in which aircraft encounter each other. Such a correlated encounter model is similar to the uncorrelated encounter model described here except the model defines the location of the aircraft at the closest point of approach.

This page intentionally left blank.

APPENDIX A SAFETY EVALUATION

This section explains how to estimate the near mid-air collision (NMAC) rate, denoted λ_{nmac} , based on a large number of simulations. We begin by observing that

$$\lambda_{\text{nmac}} = P(\text{nmac} \mid \text{enc})\lambda_{\text{enc}},$$

where $P(\text{nmac} \mid \text{enc})$ is the probability that an aircraft that enters the encounter cylinder⁵ penetrates the NMAC cylinder before exiting the encounter cylinder and λ_{enc} is the rate at which aircraft penetrate the encounter cylinder. The mean time between NMACs is simply $\lambda_{\text{nmac}}^{-1}$.

This section makes the following assumptions:

1. The density of air traffic outside the encounter cylinder is uniform in the local region being studied.
2. The trajectories of aircraft outside of the encounter cylinder are independent of the trajectories of aircraft within the encounter cylinder.

From these two assumptions, we explain how to compute $P(\text{nmac} \mid \text{enc})$ and λ_{enc} . Note that estimating the traffic density for unconventional aircraft requires a more focused assessment of a particular region and time period which is beyond the scope of this report. An estimate of the total NAS uncorrelated unconventional encounter rate is not currently possible as height finding capability does not exist for non-cooperative intruders in a large portion of the NAS. In locations where height data is available, the distribution over unconventional aircraft type should be defined when estimating NMAC rate; methods for determining this distribution are described in Section A.3. This step is not necessary when using the correlated or uncorrelated conventional models as they implicitly contain the relative aircraft category distribution. To estimate the density of unconventional aircraft in a certain region requires that all clutter, including ground vehicles and birds, be removed.

A.1 ESTIMATING NMAC PROBABILITY

By generating a large collection of encounters and determining which encounters lead to NMACs, we can estimate $P(\text{nmac} \mid \text{enc})$. Unfortunately, we cannot simply divide the number of sampled encounters that lead to NMACs by the total number of sampled encounters to estimate $P(\text{nmac} \mid \text{enc})$ due to the fact that our sampling scheme does not produce encounters from the same distribution that would occur in the airspace. In particular, we generate encounters with aircraft velocities distributed identically to the aircraft population at large, despite the fact that in reality the distribution of aircraft velocities given that an encounter is occurring favors high-speed aircraft. Although we sample from a distribution that is different from the true distribution when constructing

⁵The encounter cylinder is defined with respect to the position of one aircraft; the second aircraft is initialized on this cylinder to simulate encounters that are likely to result in an NMAC.

encounters, we can still use the samples to estimate $P(\text{nmac} \mid \text{enc})$ so long as we weight their results properly using an approach known as importance sampling [21]. We will begin by stating the weighting scheme and then prove that it is correct.

We refer to the sampled trajectories for AC1 and AC2 as \mathbf{z}_1 and \mathbf{z}_2 in an encounter with the requirement that both aircraft come from the same airspace class and altitude layer. The uncorrelated conventional documentation explained how to randomly select the position of AC2 relative to AC1, which we call \mathbf{x}_r , and the heading of AC2 relative to AC1, which we call ψ . Importance sampling allows us to make the following approximation based on N samples

$$P(\text{nmac} \mid \text{enc}) \approx \frac{1}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) \frac{V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})}{\bar{V}}.$$

The weight $V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})/\bar{V}$ corrects for the fact that our sampling distribution does not match the true distribution of encounter situations. The function $V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})$ is the average volume the encounter cylinder sweeps out per unit time when AC1 follows $\mathbf{z}_1^{(i)}$ and the airspace consists exclusively of aircraft following $\mathbf{z}_2^{(i)}$. In particular,

$$V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) = 4r_{\text{enc}}h_{\text{enc}} \underbrace{\frac{1}{\pi} \int_0^\pi \sqrt{(v_1^g + v_2^g \cos \psi)^2 + (v_2^g \sin \psi)^2} d\psi}_{\text{average relative horizontal speed}} + \pi r_{\text{enc}}^2 \underbrace{|\dot{h}_1 - \dot{h}_2|}_{\text{average relative vertical speed}},$$

where ψ is the relative heading of AC2, v_1^g and v_2^g are the initial ground speeds, and \dot{h}_1 and \dot{h}_2 are the initial vertical rates of $\mathbf{z}_1^{(i)}$ and $\mathbf{z}_2^{(i)}$. The constant \bar{V} is the average volume the encounter cylinder sweeps out per unit time

$$\bar{V} = \iint p(\mathbf{z}_1)p(\mathbf{z}_2 \mid \mathbf{z}_1)V(\mathbf{z}_1, \mathbf{z}_2) d\mathbf{z}_1 d\mathbf{z}_2.$$

Note that the distribution over \mathbf{z}_2 is conditional on \mathbf{z}_1 due to the constraint that AC1 and AC2 must belong to the same airspace class and altitude layer. The constant \bar{V} can be estimated using N samples:

$$\bar{V} \approx \frac{1}{N} \sum_i V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}). \quad (\text{A-1})$$

Now that we have defined our weighting scheme, we will now prove that it is correct. From the laws of probability,

$$\begin{aligned} P(\text{nmac} \mid \text{enc}) &= \iint P(\text{nmac} \mid \mathbf{z}_1, \mathbf{z}_2, \text{enc}) p(\mathbf{z}_1, \mathbf{z}_2 \mid \text{enc}) d\mathbf{z}_1 d\mathbf{z}_2 \\ &= \iiint P(\text{nmac} \mid \mathbf{z}_1, \mathbf{z}_2, \psi, \mathbf{x}_r, \text{enc}) p(\psi, \mathbf{x}_r \mid \mathbf{z}_1, \mathbf{z}_2, \text{enc}) p(\mathbf{z}_1, \mathbf{z}_2 \mid \text{enc}) d\mathbf{z}_1 d\mathbf{z}_2 d\psi d\mathbf{x}_r. \end{aligned}$$

We may approximate $P(\text{nmac} \mid \text{enc})$ using Monte Carlo sampling. Since it is difficult to sample from $p(\mathbf{z}_1, \mathbf{z}_2 \mid \text{enc})$ directly, we sample \mathbf{z}_1 and \mathbf{z}_2 from the distribution represented by our

Bayesian network subject to the constraint that both aircraft come from the same airspace class and altitude layer, and weight the samples appropriately:

$$P(\text{nmac} \mid \text{enc}) \approx \frac{1}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) \frac{p(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)} \mid \text{enc})}{p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})}.$$

We know that

$$\begin{aligned} p(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)} \mid \text{enc}) &= \frac{p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})}{\lambda_{\text{enc}}} \lambda_{\text{enc} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}} \\ &\propto p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)}) \lambda_{\text{enc} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}} \\ &\propto p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)}) V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}). \end{aligned}$$

We may normalize to obtain

$$\begin{aligned} p(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)} \mid \text{enc}) &= p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) / \iint p(\mathbf{z}_1)p(\mathbf{z}_2 \mid \mathbf{z}_1)V(\mathbf{z}_1, \mathbf{z}_2) d\mathbf{z}_1 d\mathbf{z}_2 \\ &= p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) / \bar{V}. \end{aligned}$$

We may substitute and simplify to obtain

$$\begin{aligned} P(\text{nmac} \mid \text{enc}) &\approx \frac{1}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) \frac{p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) / \bar{V}}{p(\mathbf{z}_1^{(i)})p(\mathbf{z}_2^{(i)} \mid \mathbf{z}_1^{(i)})} \\ &\approx \frac{1}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) \frac{V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})}{\bar{V}}, \end{aligned}$$

which corresponds to the weighting scheme we defined.

A.2 ESTIMATING ENCOUNTER RATE

Estimating the encounter rate requires knowing the density of traffic outside the encounter cylinder. This density, ρ , can be expressed in aircraft per NM³. The rate at which aircraft enter the encounter cylinder is the product of ρ and the average volume of new airspace the encounter cylinder sweeps through per unit time, \bar{V} , which was discussed in Section A.1:

$$\lambda_{\text{enc}} = \rho \bar{V}. \quad (\text{A-2})$$

The \bar{V} in Equation A-2 depends upon the size of the encounter cylinder and on the average velocities of aircraft involved in encounters. Finally, encounter rates for conventional traffic and for uncorrelated unconventional aircraft also need to be considered to obtain an overall collision risk estimate.

We may compute λ_{nmac} as follows:

$$\begin{aligned}
\lambda_{\text{nmac}} &= P(\text{nmac} \mid \text{enc})\lambda_{\text{enc}} \\
&= \rho \bar{V} P(\text{nmac} \mid \text{enc}) \\
&\approx \frac{\rho \bar{V}}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) \frac{V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})}{\bar{V}} \\
&= \frac{\rho}{N} \sum_i P(\text{nmac} \mid \mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \psi^{(i)}, \mathbf{x}_r^{(i)}, \text{enc}) V(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}).
\end{aligned}$$

The discussion so far in this appendix has focused on computing NMAC rate for a single model, but we have several different models for capturing the behavior for different categories of unconventional aircraft. The overall NMAC rate for unconventional aircraft is the sum of the NMAC rates for each category. In order to estimate the NMAC rates for each category, of course, one needs to know the density of each aircraft category in the airspace. If the overall unconventional aircraft density is known but the density of each individual aircraft category is unknown, an upper bound on the NMAC rate can still be estimated as follows:

$$\lambda_{\text{nmac}} \leq \rho \max_c \bar{V}_c P(\text{nmac} \mid \text{enc}, c), \quad (\text{A-3})$$

where \bar{V}_c is the \bar{V} associated with aircraft category c and $P(\text{nmac} \mid \text{enc}, c)$ is the probability that an encounter leads to an NMAC in the model associated with aircraft category c .

A.3 ESTIMATING AIRCRAFT DENSITY

We can use tracks from a primary surveillance radar to estimate the density of aircraft without transponders. Special care is required to remove clutter and bird tracks from the radar output, otherwise the density of noncooperative traffic may be overestimated. Unless one is satisfied with an upper bound on the NMAC rate as computed in Equation A-3, it is necessary to then associate the various primary-only aircraft tracks with the different categories of unconventional aircraft.

Before presenting the equations used for unconventional aircraft classification, we begin with an example. For simplicity, assume that we have prior knowledge that only two aircraft types exist in the region of our primary surveillance radar—gliders and paragliders. Using the paraglider and glider models developed earlier, we can assign a probability that the track corresponds to a paraglider or a glider. If we have one measurement point and the altitude at this point is 500 ft, we would quantize this to the bin encompassing 500 ft, $[0, 1200)$ and assign the relative probability from the sufficient statistics tables—0.44 for the paraglider and 0.07 for the glider. Normalizing and assuming no measurement error, the probability that the track is a paraglider is 0.87 while the probability that it represents a glider is 0.13. If we then know the airspeed of the track, 18 kt, then we can refine our estimate. Using the airspeed dependence on altitude layer in the model, we can estimate the probability of both a 500 ft altitude and 18 kt airspeed—0.166 for a paraglider and 0.001 for a glider. Given our model, we find that it is more likely that this measurement point represents a paraglider than a glider.

More formally, we wish to find the unconventional aircraft category c that maximizes $P(c | D)$, the probability that the track data D was generated by aircraft category c . From Bayes' rule, we know $P(c | D) \propto P(c)P(D | c)$. The prior $P(c)$ is our degree of belief that a track is associated with aircraft category c before any observation of track characteristics. In our paraglider and glider example, we used a uniform prior, i.e., $P(c)$ is equal for all aircraft categories.

The likelihood $P(D | c)$ is equal to $P(D | c, G_c, D_c)$ where G_c is the graphical structures—i.e., the initial network and transition network—associated with aircraft category c and D_c is the data used to construct the model for aircraft category c . The following steps may be used to compute $P(D | c)$:

1. Convert the sufficient statistics N_{ijk} in the various tables associated with aircraft category c to conditional probabilities (using uniform Dirichlet prior $\alpha_{ijk} = 1$, see Appendix E):

$$P(x_i = k | \pi_j) = \frac{N_{ijk} + 1}{\sum_{k'=1}^{r_i} (N_{ijk'} + 1)},$$

where π_j is the j th instantiation of the parents of variable x_i in the network as described in Appendix E. This conversion to probabilities should be done for both the initial and transition networks associated with aircraft category c .

2. Compute the probability that the initial Bayesian network generated the first data point associated the track we are trying to classify. If the first data point is (x_1, \dots, x_n) then,

$$P(\mathbf{x}) = P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \pi_i),$$

as given in Equation E-1. The probability $P(x_i | \pi_i)$ is as computed in Step 1.

3. For each additional time step of our track in question, use the transition network to compute the probability of making the observed transition.
4. $P(D | c)$ is the product of the probabilities computed in Steps 2 and 3.

Because many of these probabilities are small, it may be easier in terms of numerical stability to compute the logarithms in Steps 2 and 3 and then sum the logarithms instead of multiplying the small probabilities together.

We classify our track as belonging to the category with the largest value for $P(c)P(D | c)$ or, equivalently, $\log P(c) + \log P(D | c)$. After classifying a collection of tracks captured by a primary surveillance radar, we may estimate the average density of the various aircraft categories. The density of aircraft category c is the number of flight hours associated with tracks classified as category c divided by the product of the surveillance time and volume. For example, if we observe 1 flight hour of aircraft category c in 10 hours of primary radar surveillance with a volume of 10 NM^3 , then the density is $1/(10 \times 10) = 0.01$ category c aircraft per NM^3 .

This page intentionally left blank.

APPENDIX B MODEL PARAMETERS

This appendix describes the sufficient statistics, N_{ijk} (see Appendix E), used to estimate the conditional probabilities associated with the initial and transition distributions. There are nine independent tables corresponding to each aircraft type modeled. Each file is named with the convention: `aircraft_v#` (e.g., `paraglider.v1.txt`).⁶ These text files, available electronically from Lincoln Laboratory, describe the following model parameters:

- **Variable labels:** A quoted, comma-delimited list specifies the variable labels, e.g., `\dot \psi i`, as would be used by L^AT_EX. There are different variable labels for the initial network and the transition network. The ordering of the variables in this list determines the ordering of the variables in the other tables. Note that the ordering of the variable labels does not necessarily correspond to the order in which they are sampled; a topological sort may be necessary before sampling.
- **Graphical structure:** A binary matrix is used to represent graphical structure. A 1 in the i th row and j th column means that there is a directed edge from the i th variable to the j th variable in the Bayesian network, the ordering of the variables are as defined in the variable labels section of the file. The text file specifies two graphical structures; one for the initial network and the other for the transition network. The element in the i th row and j th column is represented by $G(i, j)$.
- **Variable instantiations:** For each network, a list of integers specifies the number of instantiations that exist for each variable.
- **Sufficient statistics:** For each network, a list of integers specifies the sufficient statistics. We explain how to interpret the array of integers below.
- **Boundaries:** The boundaries of the variable bins are specified by a row of numbers. The variables A and L are not quantized because they are already discrete, and so boundaries do not exist. A `*` is used for these variables.
- **Resampling rates:** A list of numbers specify the resampling rates.

The list of numbers describing the sufficient statistics, N_{ijk} , requires explanation. The array is ordered first by increasing k , then increasing j , and then increasing i . Again, the variable ordering is as defined in the variable labels section of the file. One way to load the sufficient statistics into memory is to allocate an array of pointers to 2-dimensional matrices. There would be 5 matrices for the initial network and 8 matrices for the transition network; 2 and 3 for the balloon model respectively. The dimensions of each matrix is $r_i \times q_i$, or the number of instantiations of the variable

⁶For this version of the model, the following text files are included: `balloon.v1.txt` (hot air balloons), `blimp.v1.txt` (airships), `fai1.v1.txt` (flexible wing hand gliders), `fai5.v1.txt` (rigid wing hand gliders), `glider.v1.txt` (gliders), `paraglider.v1.txt` (paragliders), `paramotor.v1.txt` (paramotors), `skydiving.v1.txt` (skydivers), and `weatherballoon.v1.txt` (weather balloons).

by the number of instantiations of the parents of the variable (see Appendix E).⁷ The counts may be read directly into the matrices from the file, starting with the first column of the first variable to the last column of the last variable.

Instead of reading the sufficient statistics into an array of matrices stored in memory, one can reference the elements in the parameters file directly. For some specified variable i , parental instantiation j , and variable instantiation k , the value N_{ijk} is given by the following element in the list

$$k + r_i(j - 1) + \sum_{i'=1}^{i-1} q_{i'} r_{i'} , \quad (\text{B-1})$$

where q and r are as specified in Appendix E.⁸

It is important to clarify the ordering of the parental instantiations. If the variables X_1, \dots, X_n are instantiated to bins b_1, \dots, b_n , the parental instantiation of variable X_i is given by

$$j = 1 + \sum_{i'=1}^n G(i', i)(b_{i'} - 1) \prod_{i''=1}^{i'-1} r_{i''}^{G(i'', i)} . \quad (\text{B-2})$$

For example, suppose that a variable has three parents. The first parental instantiation will assign all parents to their first bin. The second parental instantiation will assign the first parent (as defined by the ordering in the variable labels portion of the file) to its second bin and the other two parents to their first bin. The sequence continues until all of the parents are instantiated to their last bins.

The following is a fragment of the parameter file `fai5_v1.txt`. The lines that describe the sufficient statistics are truncated due to length.

```
# labels_initial
"L", "v", "\dot v", "\dot h", "\dot \psi"
# G_initial
0 1 1 1 1
0 0 1 1 1
0 0 0 1 1
0 0 0 0 1
0 0 0 0 0
# r_initial
4 8 5 7 7
# N_initial
5022663 10489676 8567362 2944746 134183 58943 144237 447463 1902038 ...
# labels_transition
"L", "v", "\dot v(t)", "\dot h(t)", "\dot \psi(t)", "\dot v(t+1)",
```

⁷For the transition network, note that the matrices for the variables that are not associated with time $t + 1$ are empty.

⁸In the transition network, q_i is as defined in Appendix E for the nodes representing variables at time $t + 1$. For the other nodes, i.e., the static variables and the variables at time t , q_i is set to zero.

```

"\dot h(t+1)", "\dot \psi(t+1)"
# G_transition
0 0 0 0 0 1 1 1
0 0 0 0 0 1 0 1
0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 1
0 0 0 0 0 0 0 0
# r_transition
4 8 5 7 7 5 7 7
# N_transition
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 18 1 0 0 0 4 1 0 0 0 0 0 ...
# boundaries
# boundaries
*
0 5 10 15 20 28 36 44 56
-0.75 -0.5 -0.15 0.15 0.5 0.75
-1050 -750 -450 -150 150 450 750 1050
-17.5 -12.5 -7.5 -2.5 2.5 7.5 12.5 17.5
# resample_rates
0 0 0.0983538 0.0877061 0.16609

```

This page intentionally left blank.

APPENDIX C

UNCONVENTIONAL MODEL FEATURES

Perhaps one of the most interesting outcomes of this model is the determination of how several types of unconventional aircraft actually behave. More than 99.8% of all position reports, excluding skydivers and weather balloons, were observed below 10,000 ft AGL. We further observed that when compared to the uncorrelated conventional model, heavier-than-air unconventional aircraft tend to be slower and favor unsteady flight—the acceleration, vertical rate and turn rates deviate from zero to a greater extent than for the conventional aircraft.

The feature distributions are normalized such that the integrals over the entire defined bounds are equal and not necessarily the span shown. The bounds used for these distributions are shown in Table C-1. The first set in the table is the default and any values that follow are deviations from the default. Distributions were smoothed for models that contained little data—specifically, paramotors and hot air balloons.

TABLE C-1

Bounds used for feature normalization

L	[0, 10000]	Skydiver: [0, 15000], Weather Balloon: [0, 150000]
v	[0, 300]	Skydiver: [0, 200]
\dot{v}	[-2, 2]	Skydiver: [-3, 3]
\dot{h}	[-3000, 3000]	Skydiver: [-15000, 1000]
$\dot{\psi}$	[-16, 16]	

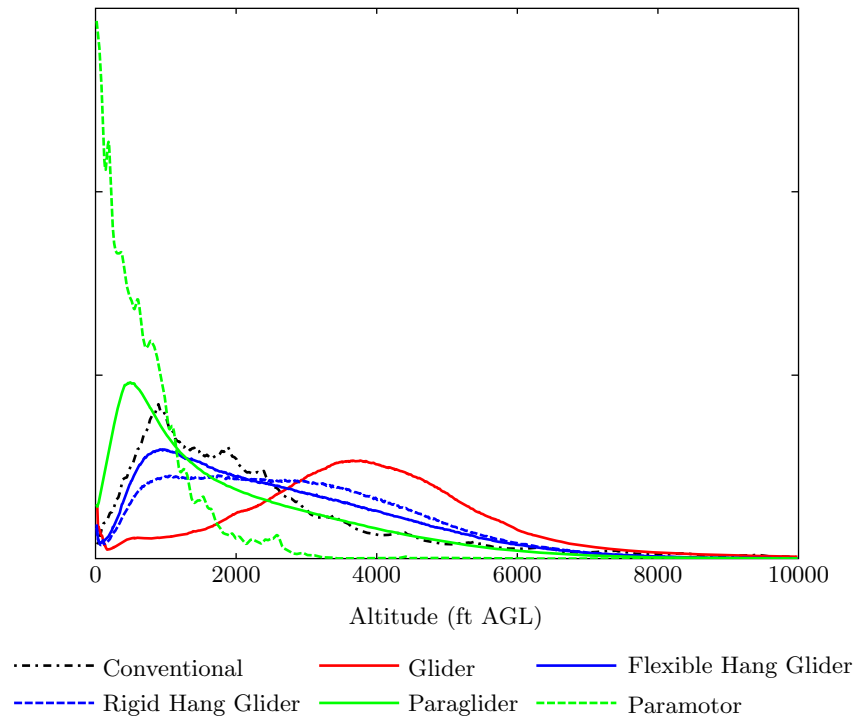


Figure C-1. Unpowered heavier-than-air altitude feature distribution.

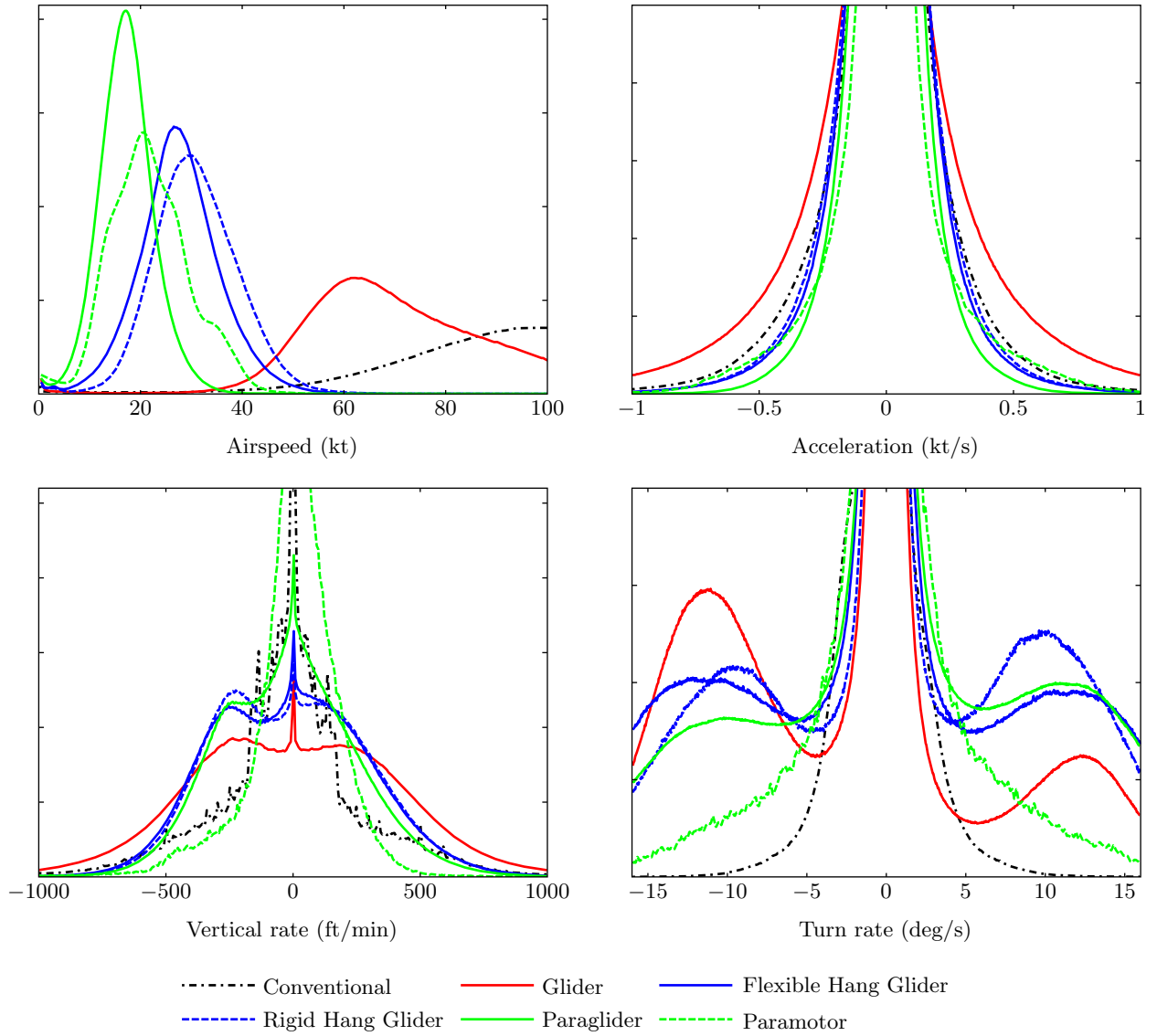


Figure C-2. Unpowered heavier-than-air detailed relative frequency feature distributions. Note that the vertical rate, turn rate and acceleration distributions are magnified such that the peaks extend beyond the figure bounds.

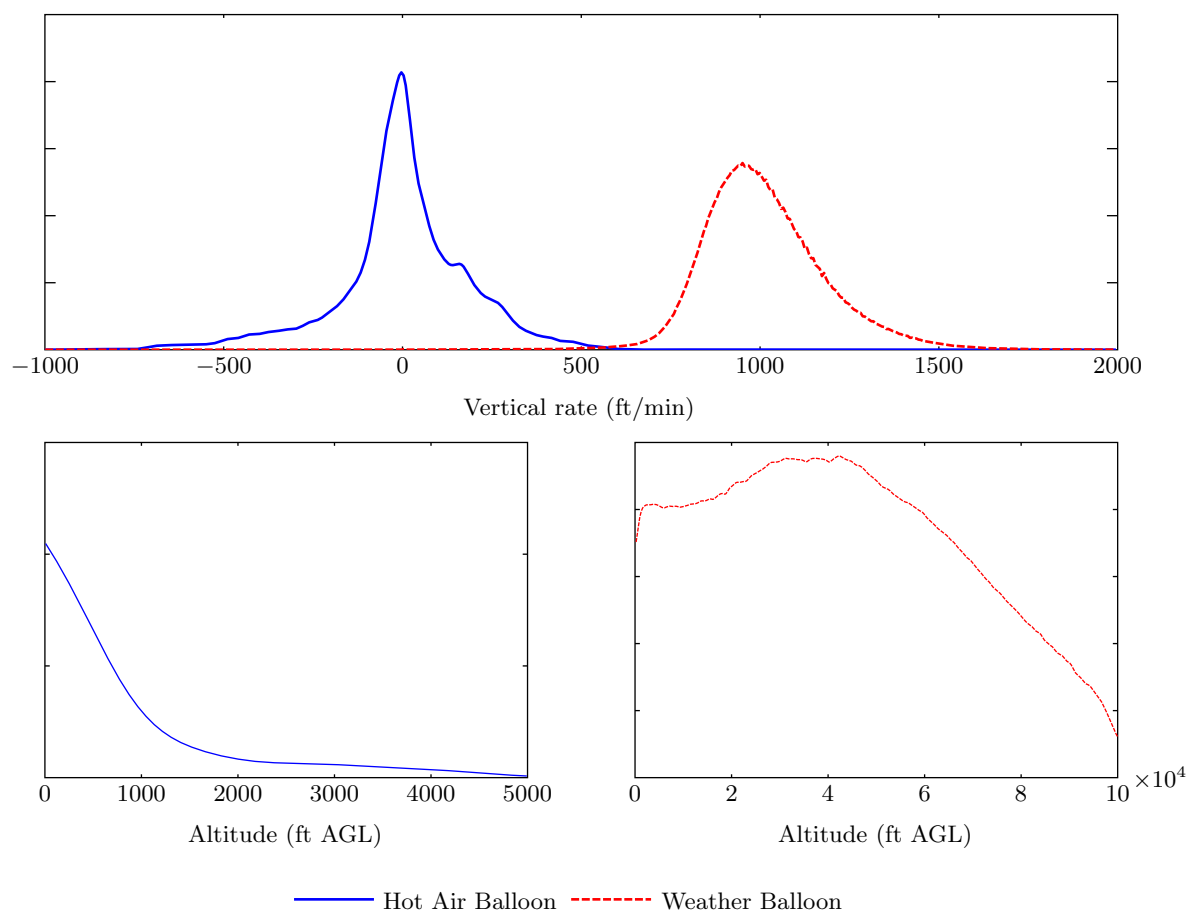


Figure C-3. Balloon feature distributions.

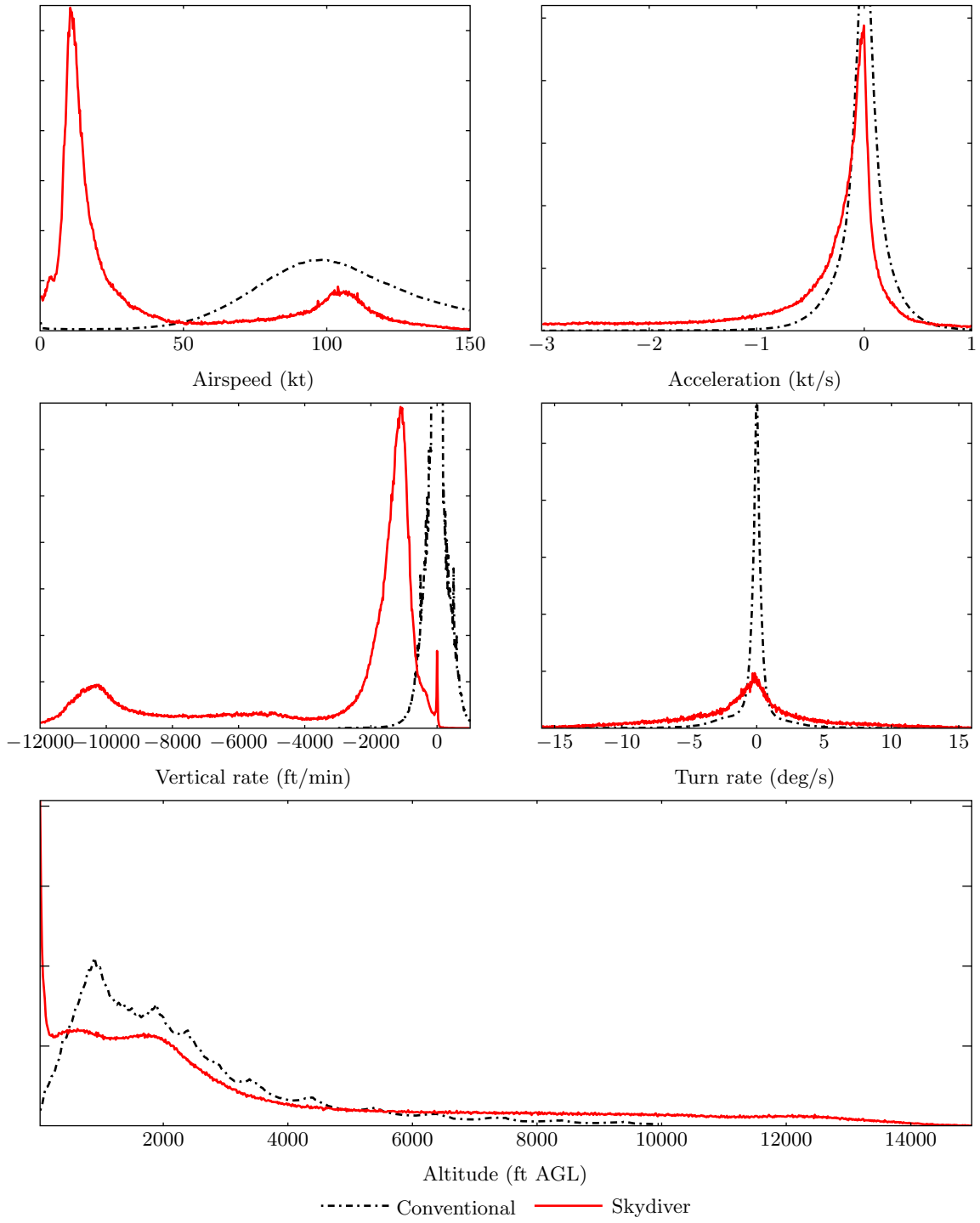


Figure C-4. Skydiver feature distributions.

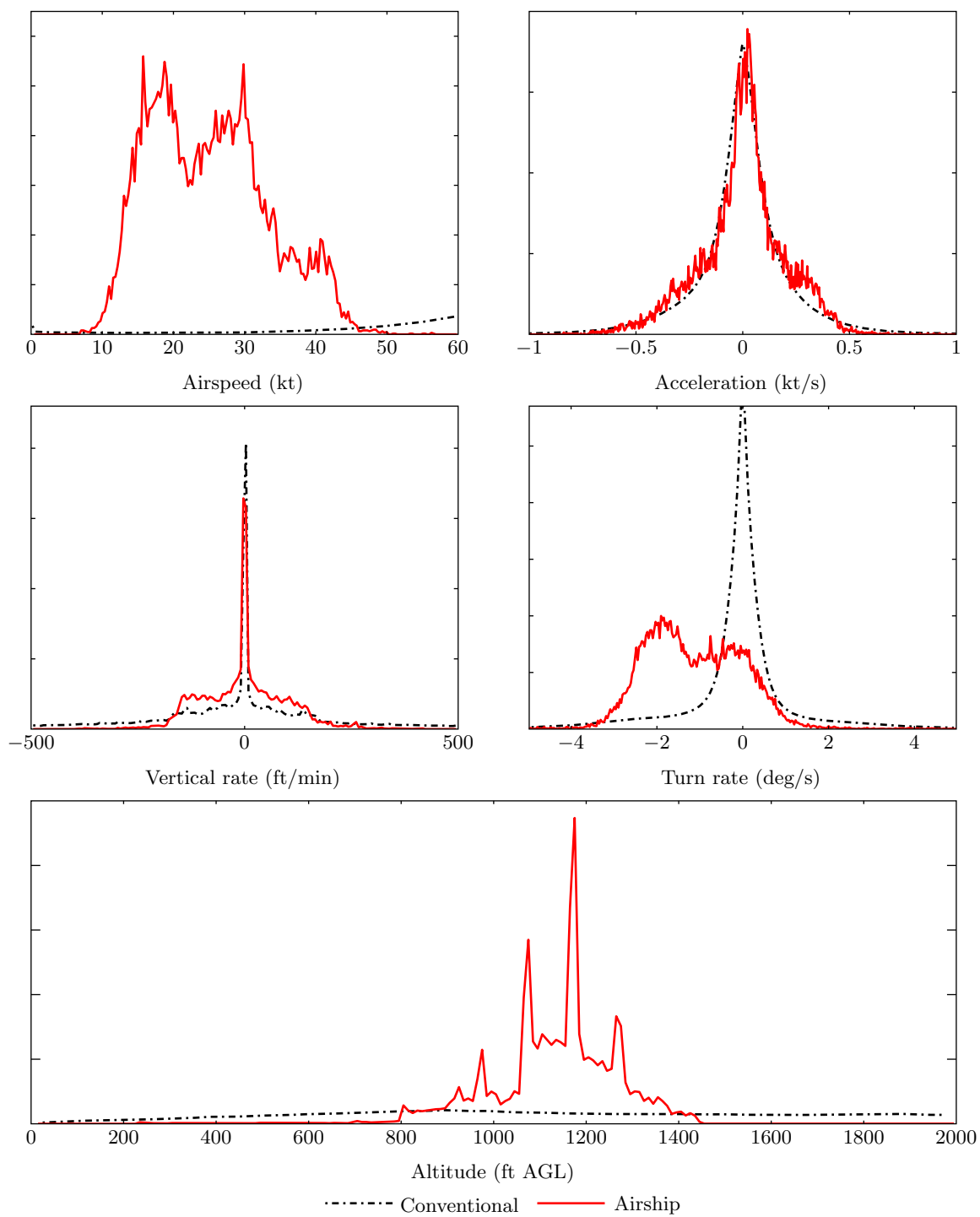


Figure C-5. Airship feature distributions.

APPENDIX D

TRACK LOCATION SUMMARY

Although density data cannot be extracted from the collected track data, a summary of track location is presented to show the diversity of tracks within the models. Both worldwide and United States location data are presented. The former is broken up into three bins per degree while the United States illustration is broken up into six bins per degree. We present a binary illustration—if an aircraft was located in a certain bin, the bin is highlighted. The majority of track coverage is observed in European airspace.

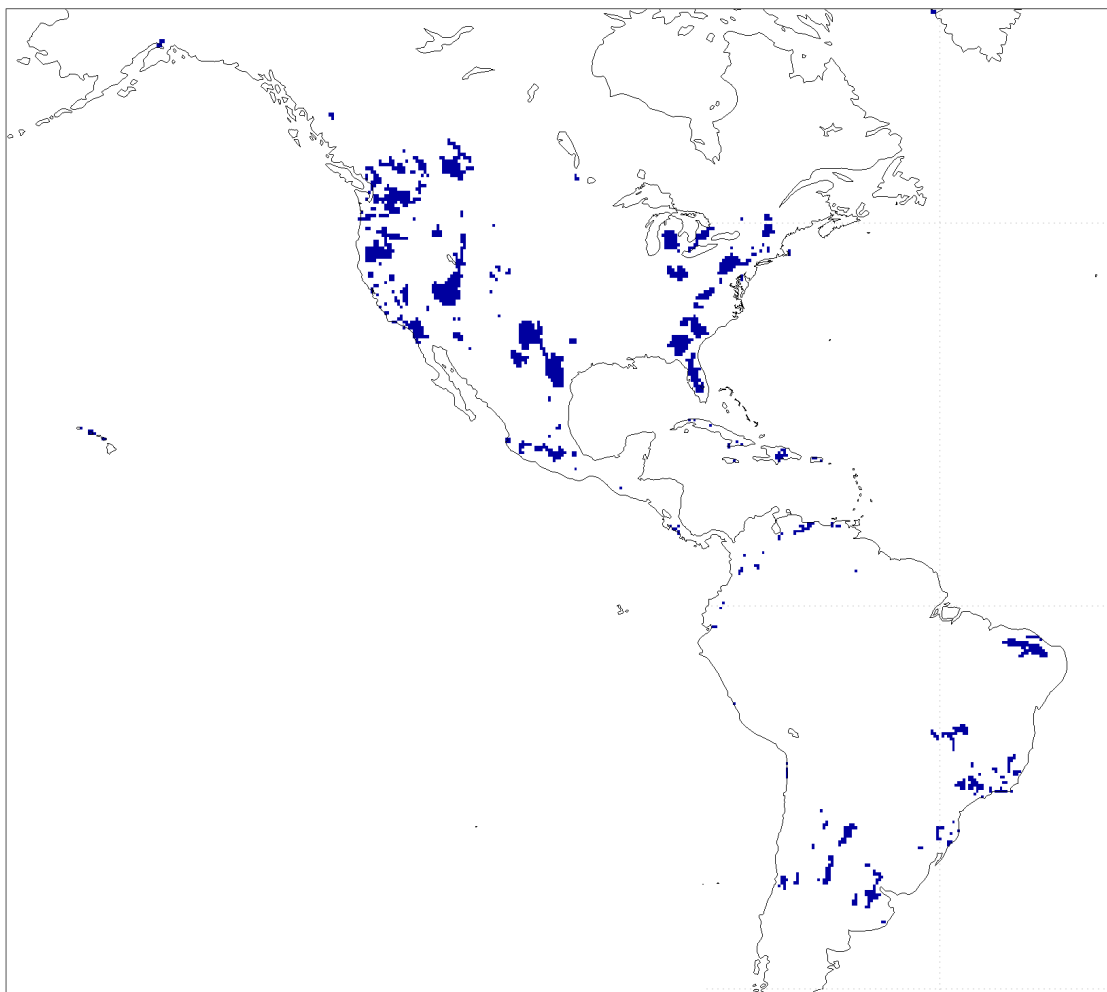


Figure D-1. Western Hemisphere track location summary.

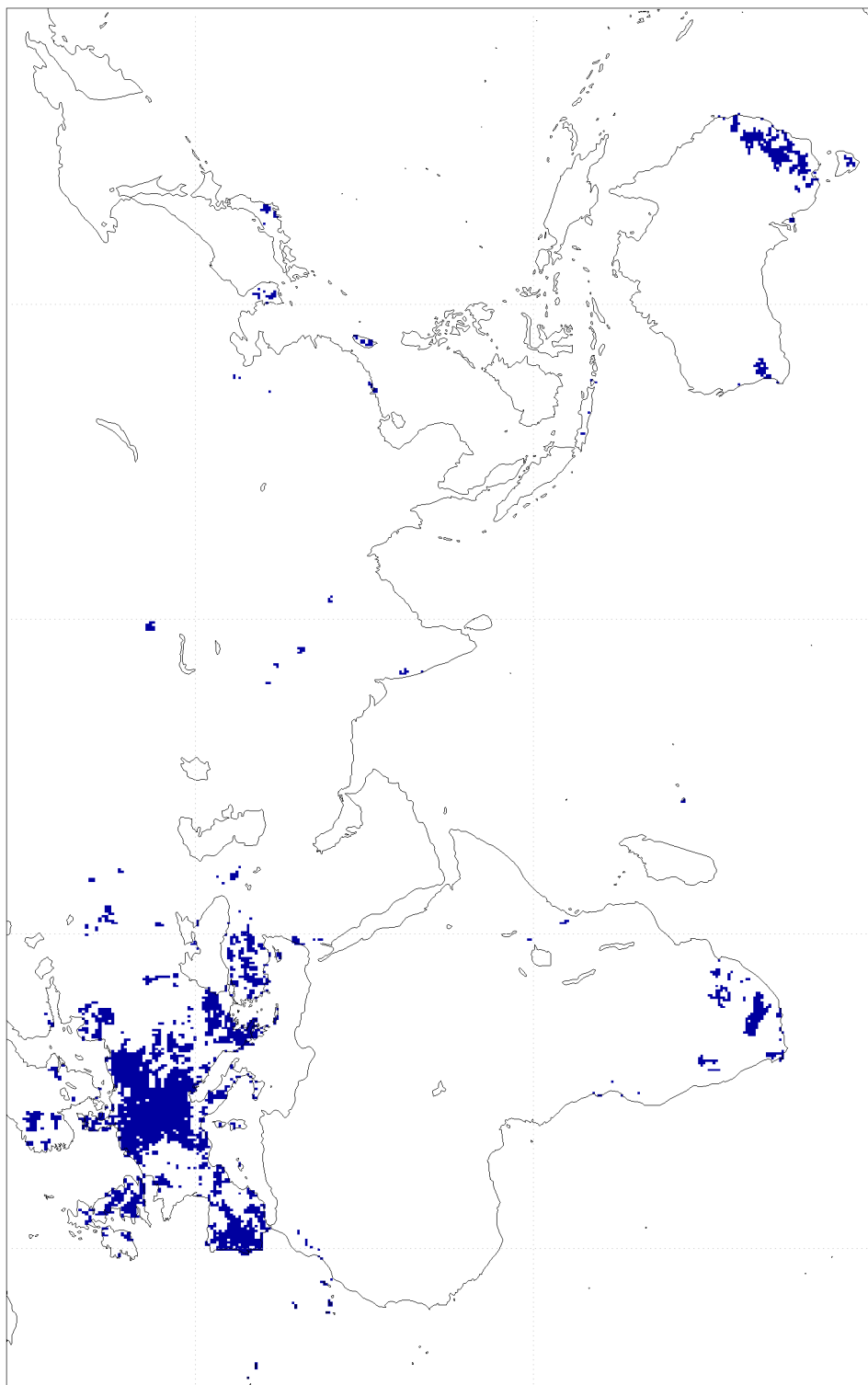


Figure D-2. Eastern Hemisphere track location summary.

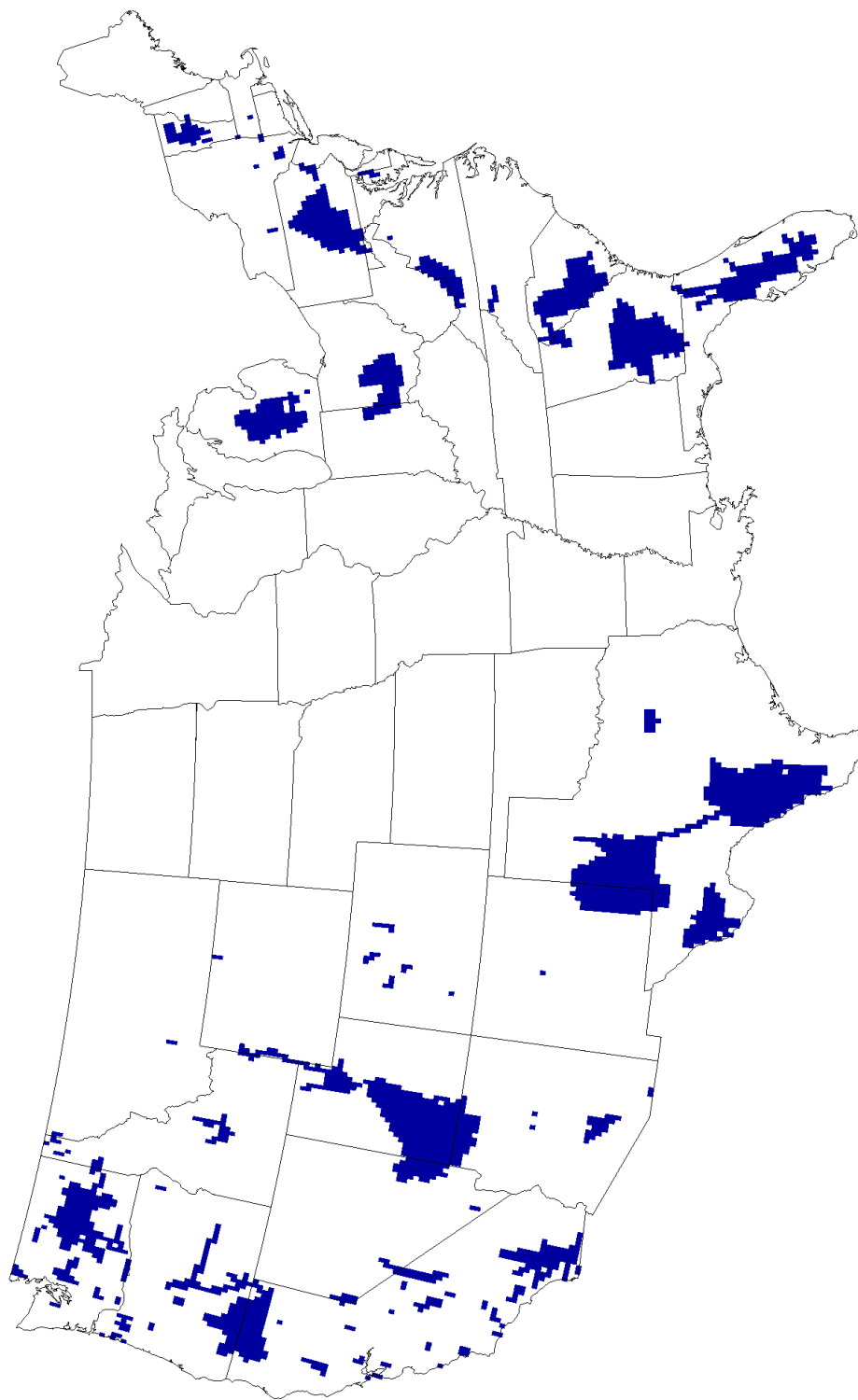


Figure D-3. United States track location summary.

This page intentionally left blank.

APPENDIX E

BAYESIAN NETWORKS

This appendix briefly reviews Bayesian networks. Further discussion of Bayesian networks may be found elsewhere [22–24].

E.1 DEFINITION

A Bayesian network is a graphical representation of a multivariate probability distribution over variables $\mathbf{X} = X_1, \dots, X_n$. In particular, a Bayesian network is a directed acyclic graph G whose nodes correspond to variables and edges correspond to probabilistic dependencies between them. Associated with each variable X_i is a conditional probability distribution $P(x_i \mid \boldsymbol{\pi}_i)$, where $\boldsymbol{\pi}_i$ denotes an instantiation of the parents of X_i in the graph. The probability of an instantiation of the variables is specified directly by the conditional probability distributions in the Bayesian network:

$$P(\mathbf{x}) = P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid \boldsymbol{\pi}_i). \quad (\text{E-1})$$

E.2 SAMPLING

It is rather straightforward to sample from a multivariate distribution represented by a Bayesian network. The first step is to produce a topological sort of the nodes in the network. A topological sort orders the nodes in a Bayesian network such that if a node X_i comes before X_j there does not exist a directed path from X_j to X_i . Every Bayesian network has at least one topological sort, but there may be many. Efficient algorithms exist for finding a valid topological sort [25].

To produce a sample from the joint distribution represented by a Bayesian network, we simply iterate through a topologically sorted sequence of the variables and sample from their conditional probability distributions. The topological sort ensures that when sampling from each conditional probability distribution the necessary parents have been instantiated.

E.3 PARAMETER LEARNING

The parameters $\boldsymbol{\theta}$ of a Bayesian network determine the associated conditional probability distributions. Given some fixed network structure G , we can learn these parameters from data. In this appendix, we assume that the variables are discrete.

Before discussing how to learn the parameters of a Bayesian network, it is necessary to introduce some notation. Let r_i represent the number of instantiations of X_i and q_i represent the number of instantiations of the parents of X_i . If X_i has no parents, then $q_i = 1$. The j th instantiation of the parents of X_i is denoted $\boldsymbol{\pi}_{ij}$.

There are $\sum_{i=1}^n r_i q_i$ parameters in a Bayesian network. Each parameter is written θ_{ijk} and determines $P(X_i = k \mid \boldsymbol{\pi}_{ij})$, i.e.,

$$P(X_i = k \mid \boldsymbol{\pi}_{ij}) = \theta_{ijk}.$$

Although there are $\sum_{i=1}^n r_i q_i$ parameters, only $\sum_{i=1}^n (r_i - 1) q_i$ are independent.

Computing the posterior $p(\boldsymbol{\theta} \mid D, G)$ involves specifying a prior $p(\boldsymbol{\theta} \mid G)$ and applying Bayes' rule

$$p(\boldsymbol{\theta} \mid D, G) = \frac{P(D \mid \boldsymbol{\theta}, G) p(\boldsymbol{\theta} \mid G)}{P(D \mid G)} = \frac{P(D \mid \boldsymbol{\theta}, G) p(\boldsymbol{\theta} \mid G)}{\int P(D \mid \boldsymbol{\theta}, G) p(\boldsymbol{\theta} \mid G) d\boldsymbol{\theta}}. \quad (\text{E-2})$$

If N_{ijk} is the count of $X_i = k$ given $\boldsymbol{\pi}_{ij}$ in the data D , then the probability of the data given the parameters $\boldsymbol{\theta}$ is

$$P(D \mid \boldsymbol{\theta}) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}}. \quad (\text{E-3})$$

Let $\boldsymbol{\theta}_{ij} = (\theta_{ij1}, \dots, \theta_{ijr_i})$. Since $\boldsymbol{\theta}_{ij}$ is independent of $\boldsymbol{\theta}_{i'j'}$ when $ij \neq i'j'$, the prior probability of the parameters assuming a fixed structure G is

$$p(\boldsymbol{\theta} \mid G) = \prod_{i=1}^n \prod_{j=1}^{q_i} p(\boldsymbol{\theta}_{ij} \mid G). \quad (\text{E-4})$$

The density $p(\boldsymbol{\theta}_{ij} \mid G)$ is a distribution over relative frequencies. Under some very weak assumptions, it is possible to prove that $p(\boldsymbol{\theta}_{ij} \mid G)$ is Dirichlet (see [24], Section 6.2.3). Hence,

$$p(\boldsymbol{\theta}_{ij} \mid G) = \begin{cases} \frac{\Gamma(\alpha_{ij0})}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1} & \text{if } 0 \leq \theta_{ijk} \leq 1 \text{ and } \sum_{k=1}^{r_i} \theta_{ijk} = 1 \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha_{ij1}, \dots, \alpha_{ijr_i}$ are the parameters of the Dirichlet distribution and $\alpha_{ij0} = \sum_{k=1}^{r_i} \alpha_{ijk}$. For the prior to be objective (or noninformative), the parameters α_{ijk} must be identical for all k . Different objective priors have been used in the literature. Cooper and Herskovits [26] use $\alpha_{ijk} = 1$. Heckerman, Geiger, and Chickering [27] use and justify $\alpha_{ijk} = 1/(r_i q_i)$.

It is possible to show that $p(\boldsymbol{\theta}_{ij} \mid D, G)$ is Dirichlet with parameters $\alpha_{ijk} + N_{ijk}, \dots, \alpha_{ijk} + N_{ijk}$. Hence,

$$p(\boldsymbol{\theta}_{ij} \mid D, G) = \begin{cases} \frac{\Gamma(\alpha_{ij0} + N_{ij})}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk} + N_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk} + N_{ijk} - 1} & \text{if } 0 \leq \theta_{ijk} \leq 1 \text{ and } \sum_{k=1}^{r_i} \theta_{ijk} = 1 \\ 0 & \text{otherwise} \end{cases},$$

where $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

Sampling from a Bayesian network with G known, $\boldsymbol{\theta}$ unknown, and D observed involves assigning k to X_i with probability

$$P(X_i = k \mid \boldsymbol{\pi}_{ij}, D, G) = \int \theta_{ijk} p(\theta_{ijk} \mid D, G) d\theta_{ijk} = \frac{\alpha_{ijk} + N_{ijk}}{\sum_{k'=1}^{r_i} (\alpha_{ijk'} + N_{ijk'})}. \quad (\text{E-5})$$

E.4 STRUCTURE LEARNING

Finding the most likely structure G that generated a set of data D . The objective is to find the most likely graph given data. By Bayes' rule,

$$P(G | D) \propto P(G)P(D | G) = P(G) \int P(D | \boldsymbol{\theta}, G)p(\boldsymbol{\theta} | G) d\boldsymbol{\theta}. \quad (\text{E-6})$$

The previous section explains how to compute the likelihood $P(D | \boldsymbol{\theta}, G)$ and the prior $p(\boldsymbol{\theta} | G)$. Cooper and Herskovits [26] show how to evaluate the integral above, resulting in

$$P(G | D) = P(G) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij0})}{\Gamma(\alpha_{ij0} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \quad (\text{E-7})$$

where $N_{ij} = \sum_{k=1}^r N_{ijk}$. Heckerman, Geiger, and Chickering [27] suggest priors over graphs, but it is not uncommon in the literature to assume a uniform prior. For numerical convenience, most Bayesian network learning packages calculate and report $\log P(G | D) + K$, where K is a constant independent of G . This quantity is often called the *Bayesian score* and may be used for structure comparison and search.

The scoring of the network structures for the conventional model assumed a uniform prior—i.e., in the absence of data, all incorporated network structures are equally likely. For the unconventional models, we incorporated an extra term into the Bayesian score to penalize complex structures when there is little data. We used a form of the minimal description length (MDL) principle, resulting in Equation E-8, where B is the total number of parameters in the network and N is the number of data points [28],

$$\text{Score}(G | D) = \log P(G | D) - \frac{B \log N}{2}. \quad (\text{E-8})$$

This page intentionally left blank.

APPENDIX F

VEHICLE MODELS AND NETWORK CANDIDATES

This section summarizes a selection of network structures along with their MDL score for each aircraft type, number of edges, and number of parameters. Bayesian model selection balances the complexity of the model according to the amount of data available. The bin cutpoints and overall bounds for each model are also presented. We used a separate model for flexible (FAI1) and rigid (FAI5) hang gliders. Using the log-ratio comparison scheme used for analysis of seasonal fluctuations in the uncorrelated model, we compared the two hang glider models and found them to be significantly different.

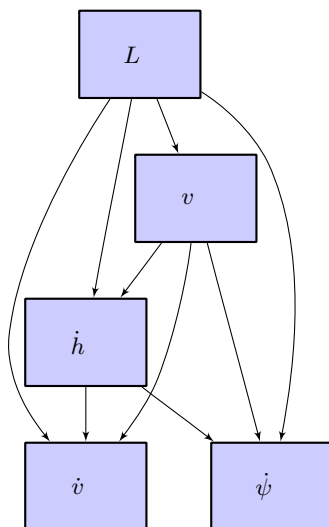
F.1 NETWORK CANDIDATES

Table F-1 summarizes the graphical networks for each model. These correspond to the highest scoring models for each aircraft type. The graphical structures were arbitrarily numbered. Note that graphical structures 9, 10 and 11 are balloon structures. The best score for each vehicle type is presented in bold.

TABLE F-1
Summary of model graphical networks

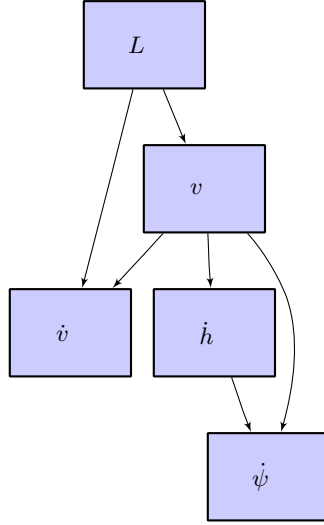
Aircraft Type	Initial Network	Transition Network
Airship	8	8
Flexible Hang Glider	6	3
Glider	6	2
Paraglider	6	2
Paramotor	5	7
Rigid Hang Glider	6	3
Skydiver	5	7
Hot Air Balloon	9	10
Weather Balloon	9	9

F.1.1 Initial Network Candidates



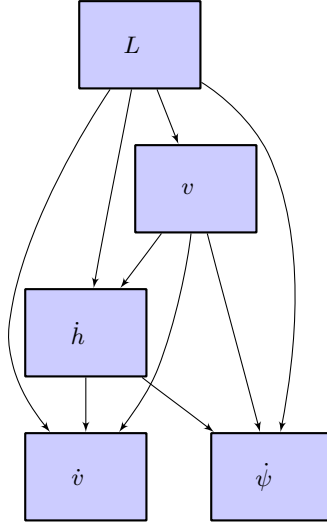
(1) Edges: 9, Parameters: 2948

Aircraft Type	Score	Aircraft Type	Score
Airship	-127442	Flexible Hang Glider	-288561215
Glider	-746171018	Paraglider	-2050764042
Paramotor	-2267859	Rigid Hang Glider	-169005203
Skydiver	-2771974		



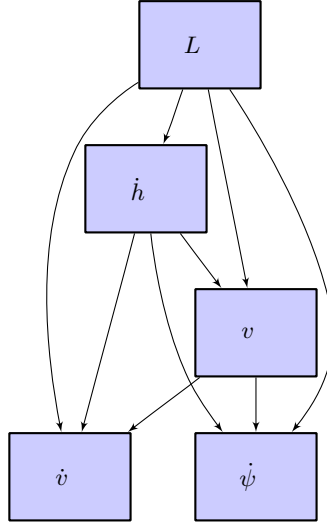
(2) Edges: 6, Parameters: 644

Aircraft Type	Score	Aircraft Type	Score
Airship	-117370	Flexible Hang Glider	-290989903
Glider	-766024848	Paraglider	-2065428858
Paramotor	-2271239	Rigid Hang Glider	-170307819
Skydiver	-2880320		



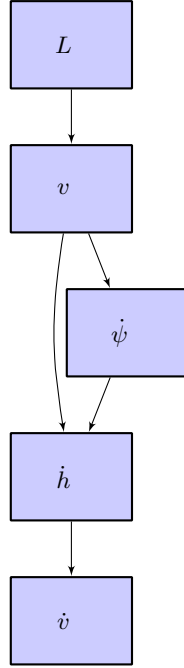
(3) Edges: 9, Parameters: 2948

Aircraft Type	Score	Aircraft Type	Score
Airship	-127442	Flexible Hang Glider	-288561215
Glider	-746171018	Paraglider	-2050764042
Paramotor	-2267859	Rigid Hang Glider	-169005203
Skydiver	-2771974		



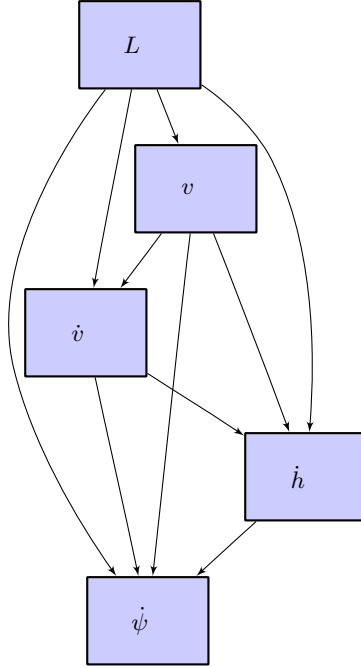
(4) Edges: 9, Parameters: 2944

Aircraft Type	Score	Aircraft Type	Score
Airship	-127337	Flexible Hang Glider	-288561205
Glider	-746171096	Paraglider	-2050764020
Paramotor	-2267664	Rigid Hang Glider	-169005220
Skydiver	-2771715		



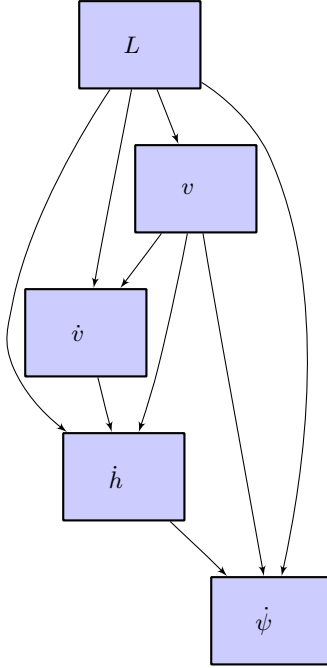
(5) Edges: 5, Parameters: 519

Aircraft Type	Score	Aircraft Type	Score
Airship	-118059	Flexible Hang Glider	-291175949
Glider	-756959320	Paraglider	-2066015497
Paramotor	-2261417	Rigid Hang Glider	-170310651
Skydiver	-2466258		



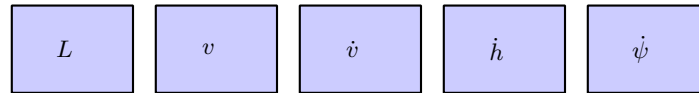
(6) Edges: 10, Parameters: 9156

Aircraft Type	Score	Aircraft Type	Score
Airship	-157260	Flexible Hang Glider	-286972934
Glider	-741023302	Paraglider	-2038659176
Paramotor	-2269845	Rigid Hang Glider	-168136948
Skydiver	-2804623		



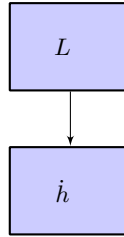
(7) Edges: 9, Parameters: 2884

Aircraft Type	Score	Aircraft Type	Score
Airship	-127688	Flexible Hang Glider	-288561019
Glider	-746170985	Paraglider	-2050764428
Paramotor	-2268296	Rigid Hang Glider	-169005146
Skydiver	-2772689		



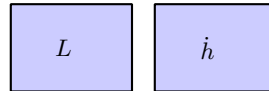
(8) Edges: 0, Parameters: 31

Aircraft Type	Score	Aircraft Type	Score
Airship	-117069	Flexible Hang Glider	-310654712
Glider	-826170679	Paraglider	-2171362734
Paramotor	-2311021	Rigid Hang Glider	-182772361
Skydiver	-3951366		



(9) Edges: 1, Parameters: Hot Air - 32, Weather - 48

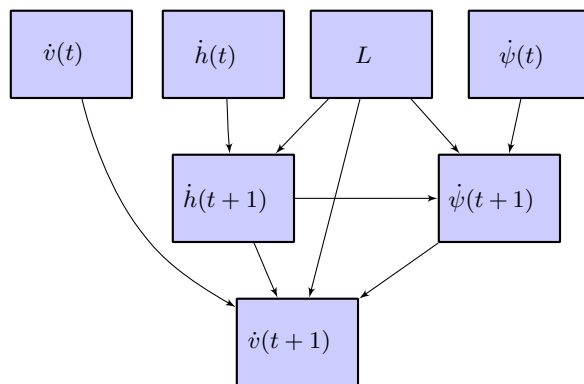
Aircraft Type	Score	Aircraft Type	Score
Hot Air Balloon	-67786	Weather Balloon	-20563438



(10) Edges: 0, Parameters: Hot Air - 11, Weather - 13

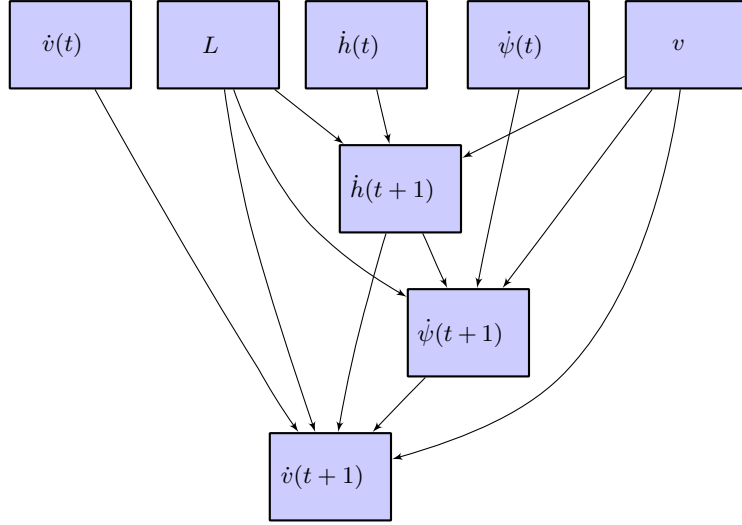
Aircraft Type	Score	Aircraft Type	Score
Hot Air Balloon	-71556	Weather Balloon	-21882514

F.1.2 Transition Network Candidates



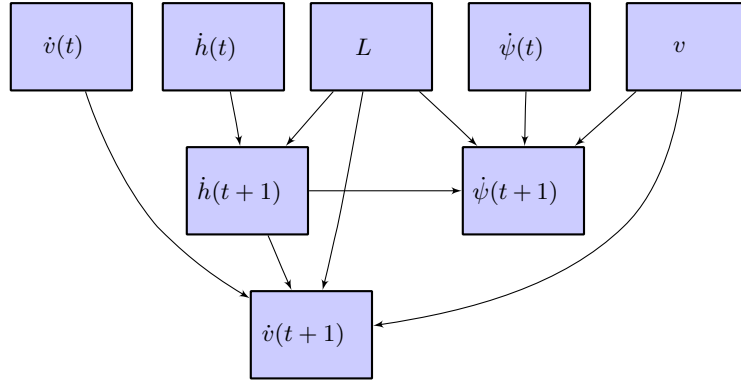
(1) Edges: 9, Parameters: 6468

Aircraft Type	Score	Aircraft Type	Score
Airship	-45278	Flexible Hang Glider	-27262343
Glider	-83891230	Paraglider	-220983956
Paramotor	-285194	Rigid Hang Glider	-15583808
Skydiver	-172723		



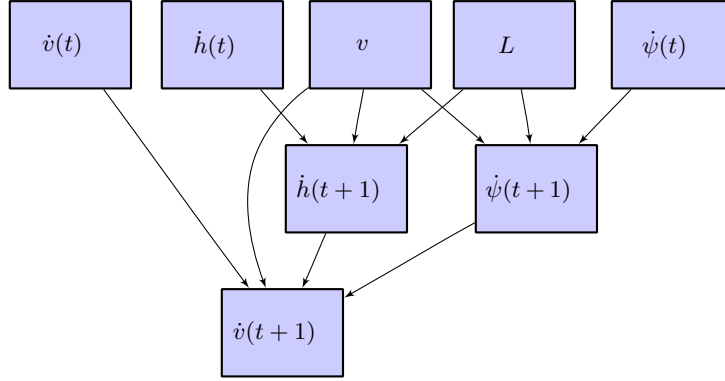
(2) Edges: 12, Parameters: 51744

Aircraft Type	Score	Aircraft Type	Score
Airship	-274019	Flexible Hang Glider	-27129715
Glider	-82456518	Paraglider	-214820107
Paramotor	-594864	Rigid Hang Glider	-15640301
Skydiver	-430370		



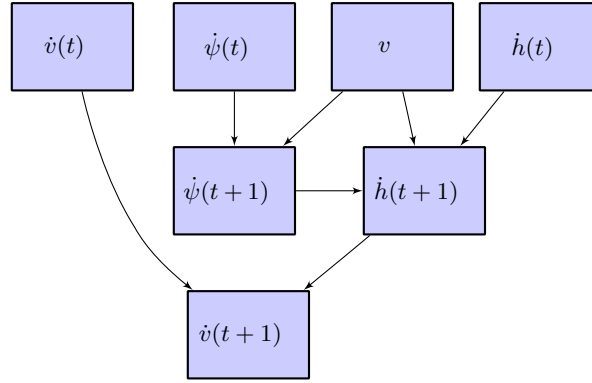
(3) Edges: 10, Parameters: 16772

Aircraft Type	Score	Aircraft Type	Score
Airship	-98071	Flexible Hang Glider	-26886062
Glider	-82993315	Paraglider	-215882995
Paramotor	-356711	Rigid Hang Glider	-15362604
Skydiver	-306750		



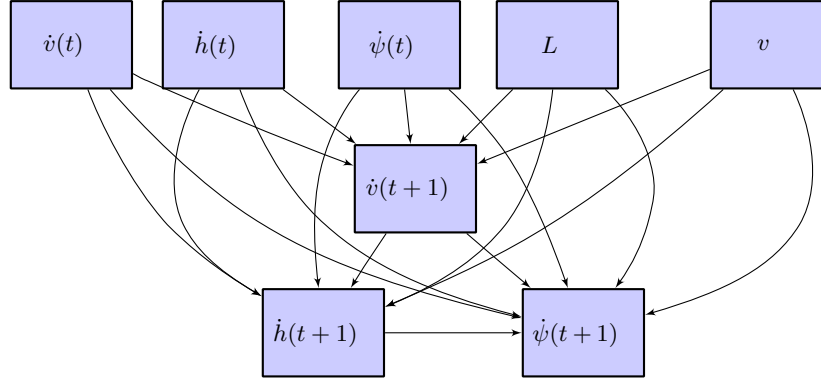
(4) Edges: 10, Parameters: 12936

Aircraft Type	Score	Aircraft Type	Score
Airship	-78980	Flexible Hang Glider	-27129433
Glider	-82998849	Paraglider	-217211818
Paramotor	-333793	Rigid Hang Glider	-15457317
Skydiver	-207748		



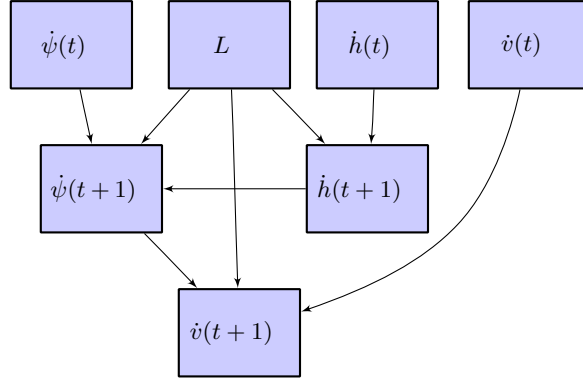
(5) Edges: 7, Parameters: 3311

Aircraft Type	Score	Aircraft Type	Score
Airship	-29766	Flexible Hang Glider	-27474885
Glider	-84104001	Paraglider	-223595098
Paramotor	-266815	Rigid Hang Glider	-15653995
Skydiver	-151368		



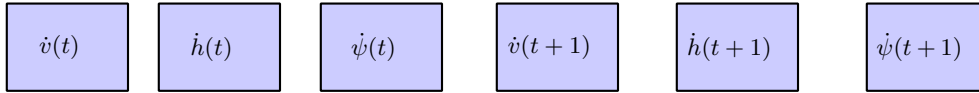
(6) dges: 18, Parameters: 2234400

Aircraft Type	Score	Aircraft Type	Score
Airship	-11145348	Flexible Hang Glider	-46578881
Glider	-100984037	Paraglider	-234877011
Paramotor	-14778111	Rigid Hang Glider	-34550374
Skydiver	-12647615		



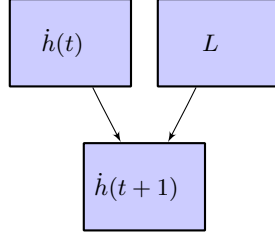
(7) Edges: 8, Parameters: 2268

Aircraft Type	Score	Aircraft Type	Score
Airship	-23701	Flexible Hang Glider	-27243027
Glider	-84029053	Paraglider	-221304636
Paramotor	-255063	Rigid Hang Glider	-15553030
Skydiver	-147131		



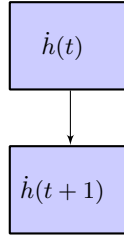
(8) Edges: 0, Parameters: 19

Aircraft Type	Score	Aircraft Type	Score
Airship	-12069	Flexible Hang Glider	-182281851
Glider	-505502679	Paraglider	-1336513258
Paramotor	-1365190	Rigid Hang Glider	-107821615
Skydiver	-1955722		



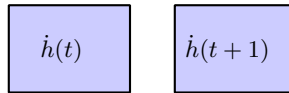
(9) Edges: 2, Parameters: Hot Air - 196, Weather - 294

Aircraft Type	Score	Aircraft Type	Score
Hot Air Balloon	-5460	Weather Balloon	-1071368



(10) Edges: 1, Parameters: Hot Air - 49, Weather - 49

Aircraft Type	Score	Aircraft Type	Score
Hot Air Balloon	-4424	Weather Balloon	-1128619



(11) Edges: 0, Parameters: Hot Air - 7, Weather - 7

Aircraft Type	Score	Aircraft Type	Score
Hot Air Balloon	-47555	Weather Balloon	-7935033

F.2 SAMPLING BOUNDARIES

Airship	
L	0, 1200, 3000, 5000, 10000
v	0, 5, 10, 15, 20, 27.5, 35, 42.5, 52.5
\dot{v}	-0.5, -0.2, -0.05, 0.05, 0.2, 0.5
\dot{h}	-700, -500, -300, -100, 100, 300, 500, 700
$\dot{\psi}$	-8, -5, -3, -1, 1, 3, 5, 8
Flexible and Rigid Hang Glider	
L	0, 1200, 3000, 5000, 10000
v	0, 5, 10, 15, 20, 28, 36, 44, 56
\dot{v}	-0.75, -0.5, -0.15, 0.15, 0.5, 0.75
\dot{h}	-1050, -750, -450, -150, 150, 450, 750, 1050
$\dot{\psi}$	-17.5, -12.5, -7.5, -2.5, 2.5, 7.5, 12.5, 17.5
Glider	
L	0, 1200, 3000, 5000, 10000
v	0, 15, 30, 45, 60, 75, 90, 105, 130
\dot{v}	-1.25, -0.5, -0.15, 0.15, 0.5, 1.25
\dot{h}	-1250, -750, -450, -150, 150, 450, 750, 1250
$\dot{\psi}$	-17.5, -12.5, -7.5, -2.5, 2.5, 7.5, 12.5, 17.5
Paraglider	
L	0, 1200, 3000, 5000, 10000
v	0, 5, 10, 15, 20, 25, 30, 35, 45
\dot{v}	-0.5, -0.2, -0.05, 0.05, 0.2, 0.5
\dot{h}	-1050, -750, -450, -150, 150, 450, 750, 1050
$\dot{\psi}$	-17.5, -12.5, -7.5, -2.5, 2.5, 7.5, 12.5, 17.5
Paramotor	
L	0, 1200, 3000, 5000, 10000
v	0, 5, 10, 15, 20, 25, 30, 35, 45
\dot{v}	-0.6, -0.2, -0.05, 0.05, 0.2, 0.6
\dot{h}	-1050, -750, -450, -150, 150, 450, 750, 1050
$\dot{\psi}$	-17.5, -12.5, -7.5, -2.5, 2.5, 7.5, 12.5, 17.5
Skydiver	
L	0, 3000, 5000, 7500, 15000
v	0, 10, 20, 30, 45, 60, 90, 120, 150
\dot{v}	-2.5, -1.5, -0.5, 0.5, 1.5, 2.5
\dot{h}	-13000, -8000, -3500, -2500, -1500, -500, 0, 500
$\dot{\psi}$	-12, -8, -4, -2, 2, 4, 8, 12
Hot Air Balloon	
L	0, 1200, 3000, 5000, 10000
\dot{h}	-800, -500, -300, -100, 100, 300, 500, 800
Weather Balloon	
L	0, 20000, 40000, 60000, 80000, 100000, 120000
\dot{h}	0, 300, 600, 900, 1200, 1500, 1800, 2100

REFERENCES

- [1] MITRE, “System safety study of minimum TCAS II,” MITRE, Technical Rep. MTR-83W241 (1983).
- [2] A. Drumm, “Lincoln Laboratory evaluation of TCAS II Logic Version 6.04a,” MIT Lincoln Laboratory, Project Report ATC-240 (1996).
- [3] M.P. McLaughlin, “Safety study of the Traffic Alert and Collision Avoidance System (TCAS II),” MITRE Corporation, Technical Rep. MTR 97W32 (1997).
- [4] B. Chludzinski, “Lincoln Laboratory evaluation of TCAS II logic version 7,” MIT Lincoln Laboratory, Project Report ATC-268 (1999).
- [5] T. Arino, K. Carpenter, S. Chabert, H. Hutchinson, T. Miquel, B. Raynaud, K. Rigotti, and E. Vallauri, “Studies on the safety of ACAS II in Europe,” Eurocontrol, Technical Rep. ACASA/WP-1.8/210D (2002).
- [6] ICAO, “ACAS manual,” Eurocontrol, Technical Rep. SCRSP/1-WP/53 (2004).
- [7] T.A. Choyce and K.M. Ciaramella, “Test and evaluation of TCAS II logic version 7,” Federal Aviation Administration, Technical rep. (2000).
- [8] ICAO, “Surveillance, radar and collision avoidance,” in *ICAO Standards and Recommended Practices*, vol. IV, annex 10 (1998).
- [9] T. Miquel and K. Rigotti, “European encounter model,” CENA/Sofréavia and QinetiQ, Technical Rep. ACASA/WP1/186/D (2002).
- [10] S. Chabert, “Safety encounter model focused on issue SA01a,” CENA/Sofréavia and QinetiQ, Technical Rep. SIRE/WP2/21/D (2005).
- [11] Federal Aviation Administration, “General Aviation and Part 135 Activity Surveys,” http://www.faa.gov/data_statistics/aviation_data_statistics/general_aviation/CY2006/ (2006).
- [12] M. Kochenderfer, J. Kuchar, L. Espindle, and J. Griffith, “Uncorrelated encounter model of the National Airspace System version 1.0,” MIT Lincoln Laboratory, Project Report ATC-345 (2008).
- [13] M. Kochenderfer, L. Espindle, J. Kuchar, and J. Griffith, “Correlated encounter model of the National Airspace System version 1.0,” MIT Lincoln Laboratory, Project Report ATC-344 (2008).
- [14] Fédération Aéronautique Internationale, “The Online Contest,” (2008), <http://www.fai.org/gliding/olc>.

- [15] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, Lincoln, Massachusetts: Ganga-Jamuna Press, 2nd ed. (2006).
- [16] “Data documentation for data set 9948 (DSI-9948): Six second upper air data,” National Climatic Data Center, Asheville, NC, Technical rep. (2003).
- [17] T. Dean and K. Kanazawa, “A model for reasoning about persistence and causation,” *Computational Intelligence* 5(3), 142–150 (1989).
- [18] K. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. thesis, University of California, Berkeley (2002).
- [19] F.N. Fritsch and R.E. Carlson, “Monotone piecewise cubic interpolation,” *SIAM Journal of Numerical Analysis* 17(2), 238–246 (1980).
- [20] W.M. Bolstad, *Introduction to Bayesian Statistics*, Wiley, 2nd ed. (2007).
- [21] R. Srinivasan, *Importance Sampling: Applications in Communications and Detection*, Springer (2002).
- [22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco, CA: Morgan Kaufmann (1988).
- [23] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Springer Verlag (2001).
- [24] R.E. Neapolitan, *Learning Bayesian Networks*, Upper Saddle River, NJ: Prentice Hall (2004).
- [25] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, 2nd ed. (2001).
- [26] G.F. Cooper and E. Herskovits, “A Bayesian method for the induction of probabilistic networks from data,” *Machine Learning* 9(4), 309–347 (1992).
- [27] D. Heckerman, D. Geiger, and D.M. Chickering, “Learning Bayesian networks: The combination of knowledge and statistical data,” *Machine Learning* 20(3), 197–243 (1995).
- [28] W. Lam and F. Bacchus, “Learning Bayesian belief networks. An approach based on the MDL principle.” *Computational Intelligence* 10, 269–293 (1994).

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE 10 April 2009		2. REPORT TYPE Project Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Encounter Models for Unconventional Aircraft, Version 1.0				5a. CONTRACT NUMBER FA8721-05-C-0002	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Matthew W. Edwards, Mykel J. Kochenderfer, James K. Kuchar, and Leo P. Espindle				5d. PROJECT NUMBER 10082	
				5e. TASK NUMBER 1	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108				8. PERFORMING ORGANIZATION REPORT NUMBER PR-ATC-348	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Science and Technology Directorate Department of Homeland Security 245 Murray Lane, SW Washington DC 20528				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ESC-TR-2007-065	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Airspace encounter models, covering close encounter situations that may occur after standard separation assurance has been lost, are a critical component in the safety assessment of aviation procedures and collision avoidance systems. Of particular relevance to Unmanned Aircraft Systems (UAS) is the potential for encountering unconventional general aviation aircraft that are flying under Visual Flight Rules (VFR) and are not in contact with air traffic control. In response to the need to develop a model of these types of encounters, Lincoln Laboratory undertook an extensive data collection and modeling effort involving more than 96,000 unconventional aircraft tracks. The outcome of this effort was nine individual models encompassing ultralights, gliders, balloons, and airships. The models use Bayesian networks to represent relationships between dynamic variables and to construct random trajectories that are statistically similar to those observed in the data. The intruder trajectories can be used in fast-time Monte Carlo simulations to estimate collision risk.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as report	18. NUMBER OF PAGES 84	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

This page intentionally left blank.