JOURNAL OF AIR TRANSPORTATION
Vol. 26, No. 4, October 2018

Bayesian Network Model of Pilot Response to Collision Avoidance System Resolution Advisories

Edward H. Londner* and Robert J. Moss†
MIT Lincoln Laboratory, Lexington, Massachusetts 02421

DOI: 10.2514/1.D0100

The effectiveness of an airborne collision avoidance system is influenced by the manner in which pilots respond to the system's advisories. Current pilot response models used in collision avoidance system modeling and simulation are agnostic to properties of individual encounters affecting pilot response, such as encounter geometry and system alerting behavior. Therefore, the pilot response behavior described by these models does not vary between encounters. Simulations using these models can lead to inaccurate results, potentially including the underestimation of collision risk in encounters where pilot response probability is low. This paper proposes a method to construct parametric pilot response models in which pilot response to collision avoidance system advisories varies based on individual encounter properties. A model was constructed from radar recordings of Traffic Alert and Collision Avoidance System (TCAS) encounters. The encounter properties with the strongest influence on pilot response were identified using a Bayesian network. The identified properties were used to predict the probability that pilots would comply with TCAS resolution advisories in individual encounters. The model was then employed in simulation of safety-critical encounters. The same encounters were also simulated using a nonparametric model in which pilot response probability did not vary between encounters and was instead equal to the average response probability predicted by the parametric model. Results showed that the nonparametric model underestimated collision risk relative to the parametric model. This study has implications for the design and performance evaluation of separation advisory systems, including collision avoidance and detect and avoid systems.

I. Introduction

THE Traffic Alert and Collision Avoidance System (TCAS) is an airborne collision avoidance system (CAS) mandated worldwide on board all large passenger and cargo aircraft [1]. TCAS mitigates collision risk by surveilling and tracking nearby air traffic and issuing vertical avoidance instructions to pilots, such as *climb* or *descend*, when a threat is determined.‡ The effectiveness of these instructions—termed *resolution advisories* (RAs)—depends in large part on how pilots respond to them [2]. TCAS's threat logic assumes an initial response delay of 5 s and a vertical acceleration of $0.25g$ ($g \approx 32\ \mathrm{ft/s^2}$) and selects and times its advisories accordingly [3]. Any deviation from these assumptions can compromise system effectiveness, increasing collision risk [4,5]. As a result, it is important to understand how pilots respond to RAs and the circumstances that influence pilot response.

The performance of TCAS and other collision avoidance logics is evaluated primarily through fast-time simulation of aircraft encounters [5–7]. A model of pilot response to RAs is a critical component of these simulations. Historically, pilot response models have incorporated parameters such as response probability, delay, and acceleration. For example, ICAO defines a *Standard Pilot Model* whose parameters conform to TCAS logic assumptions [4]. Some models incorporate stochasticity in pilot response delay [8], whereas others incorporate a response probability based on aggregated operational data [5]. However, in all cases, these models are applied identically across all encounters, even though studies have demonstrated that pilot response is strongly influenced by the properties of individual encounters, such as altitude, vertical rate, and the RA issued [9,10]. This suggests the need for a pilot response model that is sensitive to encounter-specific variables.

This paper introduces a parametric model of pilot response to TCAS RAs in which response is a function of the properties of each encounter. This model characterizes the probability that pilots will comply with RAs and does not consider variables such as response delay and strength. Using radar data recorded in U.S. airspace, pilot response to TCAS RAs was characterized across tens of thousands of observed encounters. The response data were then analyzed alongside the geometry and RAs of the encounters in a Bayesian network [11]. From the network, the encounter parameters that most strongly influence pilot response were determined. The result is a lookup table that can estimate pilot response probability for any arbitrary encounter for which the influential parameters are known. This lookup-table-based model was employed in simulations of safety-critical encounters involving TCAS-equipped aircraft to observe its effect on probability of near mid-air collision (NMAC). The results were then compared with those of a nonparametric pilot response model in which response probability did not vary between encounters and was instead equal to the average probability predicted by the parametric model for the same encounters.

The goals of this study are to introduce a method by which parametric pilot response models can be created and to gauge the effect of such a model on estimations of collision risk, as compared with a nonparametric model. Although this is a study of pilot response to TCAS advisories, the methods introduced here can support an analysis of any separation advisory system. Currently, there is substantial ongoing work to integrate unmanned aircraft systems (UAS) into civil airspace. To facilitate this, large UAS will be required to carry detect and avoid (DAA) systems to maintain safe separation from other air traffic [12]. Although not the focus of this analysis, the way in which UAS will respond to DAA advisories is a matter of current study and may incorporate both automated and manual response. As DAA systems are developed and deployed, an understanding of actual unmanned aircraft response to DAA advisories will be critical.

This paper is organized as follows: Sec. II contains background on TCAS, the data source used in this study, and Bayesian networks; Sec. III describes the method used to create the pilot response model; Sec. IV presents the model, including a description of the encounter

Selected as a Best-in-Track paper at the 12th USA/Europe Air Traffic Management R&D Seminar, Seattle, WA, 26–30 June 2017; received 3 October 2017; revision received 5 July 2018; accepted for publication 20 August 2018; published online 28 September 2018.

*Associate Staff, Surveillance Systems; elondner@ll.mit.edu.
†Associate Staff, Surveillance Systems; robert.moss@ll.mit.edu.
‡In this document, TCAS refers to TCAS II, known internationally as ACAS II.

variables that most strongly influence pilot response, as well as a comparison of the model's accuracy in predicting pilot response probability versus that of nonparametric models; Sec. V analyzes the impact of the pilot response model on TCAS safety benefit and compares it to the nonparametric model; and Sec. VI concludes this work.

II. Background

A. Traffic Alert and Collision Avoidance System

TCAS issues advisories based on estimated time to closest point of approach (CPA). Its advisories come in two forms:

1) *Traffic alerts*, which prepare pilots for subsequent alerting and aid them in visually acquiring intruders.

2) *Resolution advisories* (RAs), which are recommended vertical maneuvers intended to maintain or achieve a desired separation. Advisories that require a change in vertical rate are known as *corrective* and are accompanied by a target rate. For example, TCAS issues *climb* and *descend* RAs that direct pilots to maintain a 1500 fpm (feet per minute) climb or descent, respectively.

In the United States, flight crews are nominally directed to comply with all TCAS RAs. However, they may choose to not respond in cases where they believe doing so would jeopardize the safety of flight or when they can ensure safe separation with definitive visual acquisition of the intruder causing the RA [13]. Studies of radar data have shown that pilot response varies widely. One such study estimated compliance with *climb* and *descend* RAs in the United States at 41% and 59%, respectively [9], whereas a study of European data estimated overall compliance with *climbs* and *descends* at 59% [14].

Operational studies have shown that when TCAS issues advisories, it is often during normal and safe procedures. For example, one analysis of U.S. radar data observed that 51% of TCAS RAs are issued when aircraft are safely separated in altitude by 500 ft: the standard procedural separation between VFR and IFR traffic [7]. The same analysis observed that 12% are issued during approaches to parallel runways. Therefore, it is important to keep in mind that noncompliance with TCAS RAs does not necessarily suggest that safety has been compromised.

B. TCAS RA Monitoring System (TRAMS)

The recorded radar data analyzed in this study come from the TCAS RA Monitoring System (TRAMS). TRAMS is a network of 22 secondary-surveillance radars[§] distributed across the contiguous United States (Fig. 1 and Table 1 outline the locations and coverage areas of the TRAMS radars) [3]. When a TCAS RA is issued within the TRAMS coverage area, RA information is downlinked by the transponders of the encountering aircraft and recorded along with the geometry of the encounter as measured by the radar. These RA downlinks may also contain data about the encountering aircraft, including aircraft category and type.

The format and content of the data recorded by TRAMS are a function of the transponder type and the version of TCAS on the aircraft receiving the RA. The format associated with version 6.04a of TCAS, the oldest version of the logic deployed in U.S. airspace, contains less information than the format associated with the subsequent TCAS versions 7 and 7.1. This legacy format comprises approximately 37% of TRAMS recordings.

TRAMS makes separate recordings for each TCAS-equipped aircraft involved in an encounter. This means, for example, that a single encounter in the airspace between two TCAS-equipped aircraft will be recorded as two unique encounters by TRAMS, assuming that an RA is issued by both TCAS units. In this instance, the first recorded encounter will contain RA information for the first aircraft and the second recorded encounter will contain RA information for the second aircraft. TRAMS recordings also include a small number of encounters involving three or more aircraft (approximately 0.3%), although this analysis considers encounters between two aircraft only.

[§]The number of TRAMS radars is increasing over time.

Fig. 1 Coverage area of TRAMS (excluding Parker, CO).

Over 700,000 RA encounters have been recorded by TRAMS since encounter monitoring began in 2008. With one exception, all TRAMS sensors are terminal radars with a coverage radius of 60 n miles and a rotation period of approximately 4.6 s, which is also the sampling period of data recorded by these sensors. The exception is the Parker, Colorado, sensor, which is an en-route radar with a 200 n miles coverage radius and a rotation period of approximately 10 s. Because of its longer sampling period, the Parker, Colorado, sensor was considered inappropriate for this analysis and thus excluded.

The primary advantage of TRAMS is its large number of recorded TCAS encounters. However, TRAMS also has limitations that must be acknowledged. One limitation of TRAMS is its aforementioned sampling period of 4.6 s, which is long enough to obscure rapid yet relevant instances of aircraft movement in some encounters. TRAMS recordings also sometimes contain outliers and other inaccuracies that must be detected and removed. In this analysis, steps were taken to mitigate the impact of these limitations, as described in Sec. III.C. TRAMS is also limited by its coverage area, which is concentrated around large airports and therefore emphasizes terminal airspace over en-route airspace. And finally, TRAMS is limited in that it does not include some data that are potentially relevant to this analysis. For example, information pertaining to traffic alerts, communication with air traffic control, and visual identification of intruders could all be expected to impact pilot response to RAs. But these variables do not enter into this analysis because TRAMS does not include them.

C. Bayesian Networks

A Bayesian network is a compact graphical representation of a joint probability distribution [11]. Bayesian networks consist of *nodes* connected acyclically by arrows. Each node represents a random variable that can be discrete or continuous. Arrows point from *parent* to *child* nodes and indicate probabilistic dependencies between the nodes. Associated with each node is a conditional probability distribution that is a function of the values of the node's parents.

Table 1 Locations of the TRAMS radars

Radar	Location	Radar	Location
PHL	Philadelphia, PA	ATL	Atlanta, GA
LAXN	Los Angeles, CA	HPN	White Plains, NY
JFK	New York City, NY	SEA	Seattle, WA
DFW	Dallas-Ft. Worth, TX	ORDA	Chicago, IL
ONT	Ontario, CA	FLL	Ft. Lauderdale, FL
LGB	Long Beach, CA	LAS	Las Vegas, NV
OAK	Oakland, CA	PDX	Portland, OR
SDF	Louisville, KY	EWR	Newark, NJ
STL	St. Louis, MO	DEN	Denver, CO
PHX	Phoenix, AZ	MOD	Bedford, MA
ACY	Atlantic City, NJ	QPK	Parker, CO

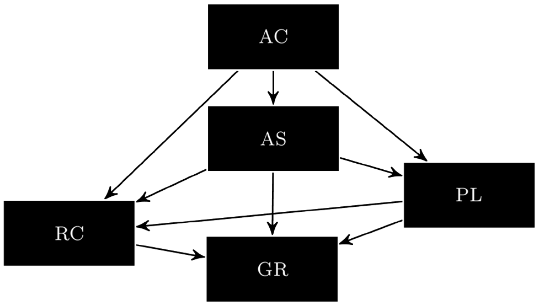


Fig. 2 Example Bayesian network.

An example Bayesian network is depicted in Fig. 2. The nodes of this sample network pertain to encounters involving a TCAS-equipped aircraft and are a subset of the pilot response network created in this study, shown in Fig. 4. Abbreviated definitions follow (complete definitions are in Sec. II.B):

1) *AC* represents the category of the TCAS aircraft.

2) *RC* represents the relative course between the TCAS aircraft and the intruder.

3) *AS* represents the airspace type in which the encounter took place (Class A, Class B, etc.).

4) *PL* represents whether the encounter took place during an approach to parallel runways.

5) *GR* represents the ground range (i.e., the horizontal range) between the aircraft when the RA was issued on the TCAS aircraft.

In this example network, the *GR* node has three parents: *RC*, *AS*, and *PL*, each of which is a child of *AC*. Arrows capture the dependencies among the nodes. Because of these dependencies, we can make inferences about the state of a node given knowledge of the states of the nodes that it depends on. In this example network, these dependencies include one between *AC* and *GR*, which exists through the connections of the nodes between them and despite the fact that they are not directly connected by arrows themselves. However, if we have knowledge of the states of the *RC*, *AS*, and *PL* nodes, then by virtue of the network structure, any inferences for the state of the *GR* node become independent of the state of the *AC* node. We say that *GR* is *conditionally independent* of *AC* given knowledge of the parents of *GR*, and therefore knowledge of these parents fully defines the probability distribution of *GR*. This notion of conditional independence is an important element of Bayesian networks and this analysis.

The objective of this work is to build a Bayesian network that characterizes the probability of pilot response to TCAS RAs based on encounter parameters. Using the principles of Bayesian network structure learning and conditional independence, this analysis will determine the encounter parameters that pilot response probability conditionally depends on (i.e., its parent nodes) and then define pilot response probability as a function of these parameters. This relationship is captured in a lookup table where each row contains a unique combination of parent node values and a corresponding value for pilot response probability.

Bayesian networks are a powerful statistical tool with precedent in aviation research. For example, Bayesian networks were used to construct the *Lincoln Laboratory Correlated Encounter Model* (LLCEM), which used U.S. radar data to model aircraft trajectories in encounters [15]. Additional applications include decision-making systems for forced UAS landings [16] and threat evaluation in air defense scenarios [17].

III. Method

The first task in the construction of the Bayesian network was to define *pilot response* in the context of this analysis. Next, the network nodes were selected. Afterward, data were collected for each node from the recorded TRAMS data. Finally, the arrows between the nodes were drawn (i.e., the network structure was *learned*) based on the gathered data. This section describes these steps.

A. Pilot Response

In this analysis, the scope and definition of pilot response were constrained by the data source. TRAMS data are sampled at approximately 4.6 s intervals and TRAMS altitude data, which are acquired from aircraft transponder replies, are quantized to either 25- or 100-foot bins. Additionally, TRAMS data downlinked in the legacy format (see Sec. II.B) do not distinguish between certain RAs, including the various types of *adjust vertical speed* or *level off* advisories issued by TCAS versions 7 and 7.1.

Because of these limitations, this analysis studies pilot response to *climb* and *descend* RAs only. Furthermore, the *climb* or *descend* RA must have been the first RA issued by the corresponding TCAS unit during the encounter (RAs issued by the intruder are not considered). This second condition was included to eliminate the influence of any previously issued RAs on pilot response. Approximately 31% of TRAMS encounters met both of these conditions and were therefore eligible for inclusion in the Bayesian network. Any conclusions drawn from this analysis must bear this in mind.

The definition of pilot response employed in this study considers RA compliance only, where a pilot is said to have responded to (i.e., complied with) the *climb* or *descend* RA if the aircraft achieved a vertical rate of at least 400 fpm in the appropriate direction within 15 s after the RA was issued (note that initial *climb* and *descend* RAs advise a rate of 1500 fpm). Response delay and vertical acceleration were not considered, as they would require a data source with finer resolution in time and altitude than TRAMS. This definition of pilot response has precedent in previous studies of TCAS operational data [9].

B. Node Selection

Nodes were selected based on subject matter expert perception of the factors influencing pilot response. For example, experience suggests that pilots may be more likely to comply with RAs that do not conflict with their current vertical rate: an intuition supported by research into compliance with *climb* RAs [10]. In addition, nodes were constrained to data that could be ascertained from TRAMS recordings. This excluded any potential effects of traffic alerts, for example, as they are not recorded by TRAMS.

A summary of the selected nodes follows. All selected nodes represent discrete quantities; the binning of the nodes is summarized in Table 2.

1) *Aircraft category (AC)*: Category of the TCAS-equipped aircraft, including major air carrier, regional air carrier, business jet, helicopter, other (typically piston engine general aviation), and unknown.

2) *Airspace (AS)*: Airspace of the encounter. Potential values include Classes A, B, C, D, and E/G, or special use.

3) *TCAS sensitivity level (SL)*: Sensitivity level of the TCAS unit issuing the RA. Potential values range between 3 and 7, with higher levels corresponding to more sensitive alerting and higher altitudes [3]. Sensitivity level served as a surrogate for aircraft altitude in this analysis.

4) *Intruder beacon category (VFR)*: Boolean variable that is true when the intruder is squawking 1200 (VFR) and false otherwise. This node is based on the assumption that intruders squawking 1200 are less likely to be receiving separation services from air traffic control, with potential effects on RA compliance by the TCAS aircraft. (Note: this node is intended to serve as a partial surrogate for the aircraft category of the intruder, which may also be relevant but could not be included in the Bayesian network due to a lack of availability for all intruders.)

5) *Parallel (PL)*: Boolean variable that is true if the encountering aircraft are on approach to parallel runways, which was determined

Table 2 Discretized encounter variables

Variable	Discretization	Units
RC	$0, 45, 90, \ldots, 315$	deg
RH	$0, 400, 800, \ldots, 1600, \geq 2000$	ft
VR	$0, 500, 1000, \ldots, 2000, \geq 2500$	fpm
GR	$0, 1, 2, \ldots, 5, \geq 6$	n miles
VMD	$0, 250, 500, \ldots, 1000, \geq 1250$	ft
HMD	$0, \frac{1}{4}, \frac{1}{2}, \ldots, 2, \geq 2\frac{1}{4}$	n miles

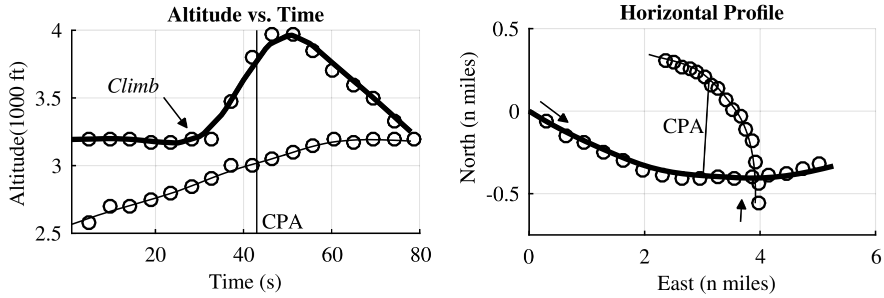


Fig. 3 Example TRAMS encounter between a TCAS aircraft (thick lines) and an intruder without TCAS.

based on algorithms incorporating aircraft course, horizontal range to intruder, and proximity to an appropriate airport.

6) *Relative course (RC)*: Difference in course between the ownship and intruder. A value of 0° corresponds to parallel courses, whereas a value of 90° corresponds to an intersection from the right.

7) *Relative altitude (RH)*: Unsigned altitude difference between the ownship and intruder at alerting time.

8) *Vertical rate (VR)*: Unsigned vertical rate of the ownship at alerting time.

9) *Rate reversal (RR)*: Boolean variable set to true if the RA commands a vertical rate in the opposite direction of the aircraft's current vertical rate, which must be in excess of 400 fpm. Note that rate reversals are distinct from RA reversals, which occur, for example, when a *climb* RA transitions to a *descend* RA.

10) *Ground range (GR)*: Horizontal range between the ownship and intruder at alerting time.

12) *Climb/descend (CD)*: Boolean variable set to true if the RA is a *climb* and false if it is a *descend*.

13) *Pilot response* (ρ): Boolean variable set to true if the aircraft complied with the *climb* or *descend* RA according to the definition outlined previously.

14) *Vertical miss distance (VMD)*: Unsigned vertical distance at time of minimum horizontal separation.

15) *Horizontal miss distance (HMD)*: Horizontal distance at time of minimum horizontal separation.

C. Data Collection

Data were collected from a subset of TRAMS encounters recorded between 2008 and 2016. Recorded position and altitude data were smoothed and interpolated to 1 s intervals using a collision avoidance simulation tool developed at Lincoln Laboratory that incorporates a dynamic model of aircraft motion. Geometric values such as relative course and ground range were calculated based on this smoothed data. Figure 3 shows an example encounter comparing aircraft trajectories before and after smoothing. In this figure, open circles represent the original radar recording and solid lines represent smoothed trajectories (note the *climb* RA downlinked by the TCAS aircraft at $t \approx 28$ and the subsequent response). Geometric parameters computed at alerting time (see previous subsection) were calculated at the time that the RA was first indicated by the TRAMS sensor—the *time of first downlink*—minus 5 s. This is because for any given encounter, the time an RA was issued was actually between the time of the corresponding downlink and the previous radar sweep approximately 4.6 s earlier. If parameters such as vertical rate were calculated at the time of first downlink without the 5 s offset, then the results would be affected by any pilot response to the RA occurring between radar sweeps. Consider that at a vertical acceleration of $0.25g$, the standard acceleration assumed by TCAS logic, vertical rate will change by approximately 2400 fpm in only 5 s. Calculating these geometric parameters 5 s before the time of first downlink eliminates this potential biasing.

TRAMS encounters were filtered for validity and appropriateness to this analysis. An encounter was included only if it met the following criteria:

1) First RA was a *climb* or *descend* advisory
2) Not a formation or military flight
3) Longer than two downlinks (approximately 10 s)
4) Contained no RA reversals (e.g., a *climb* advisory transitioning to a *descend* advisory)

Steps were also taken to eliminate duplicate encounters caused by overlapping radar coverage. The resulting dataset after applying these criteria consisted of 80,955 encounters.

D. Structure Learning

The final step in the construction of the Bayesian network was to determine the connections between nodes: the network structure. Known as *structure learning*, this step was supported by the GeNIe software environment created by the Decision Systems Laboratory at the University of Pittsburgh [18].

Several candidate networks were created using a variety of commonly used structure learning algorithms appropriate for this application. For each candidate, the nodes were organized into four *temporal layers*. Temporal layers enforce causality between nodes by imposing the constraint that the children of any particular node must be in the same or a lower temporal layer. The temporal layers of this analysis were chosen to capture the causal relationships among encounter parameters and are outlined in Table 3.

Network candidates were judged based on several criteria, including a metric known as the *Bayesian score*. A network's Bayesian score measures how well its structure probabilistically represents the data used to build it [11]. It is represented logarithmically, with higher scores corresponding to more representative structures. Other judging criteria included network structure simplicity, with simpler networks preferred, and the ease by which the network could be implemented in simulations of aircraft encounters.

The candidate networks are summarized in Table 4 along with their Bayesian scores and the algorithm used to create them.[¶] The selected network is marked in bold. This table also includes a *Naive Bayes* network, which has the *pilot response* node as the direct parent to all of the other nodes and no connections between the other nodes. It also includes a *fully disconnected* network, which has no connections between any of the nodes, including *pilot response*, and thus represents the extreme case in which all of the encounter parameters are probabilistically independent of one another. The Bayesian scores of these two networks, whose structures were not informed by the TRAMS data or structure learning algorithms, are meant to serve as baselines for comparison to the other networks.

Table 4 also includes two model performance metrics. These metrics represent the accuracy of the pilot response probability predictions made by the candidate Bayesian networks as compared with the actual response outcomes of the TRAMS encounter set. Both metrics make use of a *prediction error*, represented by ϵ and defined in Eq. (1). In this equation, ρ represents pilot response probability as predicted by one of the Bayesian networks and the subscript i represents an individual encounter.

The first metric in Table 4, ϵ_{mean}, is the arithmetic mean of the prediction error, as defined in Eq. (2). In this equation, N equals 80,955: the number of encounters in the TRAMS encounter set. Nonzero values for this metric represent a systematic bias in the

[¶]Multiple configurations of the *Greedy Thick Thinning* (GTT) [19] and *Bayesian Search* (BS) [20] algorithms were employed. Certain configurations require a maximum number of parents k for each node. Among the GTT algorithms, *K2* and *BDeu* refer to specific search strategies [21,22].

Table 3 Temporal layers

Layer	Description	Nodes
1	Aircraft parameters and encounter geometry	AC, AS, SL, VFR, RC, RH, VR, GR, PL
2	RA-related parameters	CD, RR
3	Pilot response	ρ
4	Encounter outcome	VMD, HMD

Table 4 Root mean squared error and log-scaled Bayesian scores of candidate networks

Algorithm	Bayesian score (log scale)	ϵ_{mean}	ϵ_{RMS}
$\text{GTT}^{K2}_{k=8}$	-1.090×10^6	-1.57×10^{-3}	0.3887
$\mathbf{GTT^{K2}_{K=5}}$	$\mathbf{-1.092\times10^6}$	$\mathbf{-8.36\times10^{-4}}$	**0.3915**
GTT^{BDeu}	-1.097×10^6	-1.44×10^{-6}	0.3914
$\text{BS}_{k=5}$	-1.192×10^6	-2.18×10^{-4}	0.4016
$\text{BS}_{k=8}$	-1.193×10^6	-2.18×10^{-4}	0.4016
Naive Bayes	-1.207×10^6	——	——
Fully disconnected	-1.318×10^6	——	——

predictions of the corresponding Bayesian network, and so values close to zero are desired.

The second metric in Table 4, ϵ_{RMS}, is the root mean squared value of the prediction error, as defined in Eq. (3) [N has the same value in this equation as in Eq. (2)]. Whereas the previous metric captures the presence of any systematic bias in prediction error, this metric, which is always positive, represents the average magnitude of the prediction error. As in the previous metric, smaller values are desired.

$$\epsilon_i = \begin{cases} \rho_i - 1, & \text{if pilot responded} \\ \rho_i, & \text{otherwise} \end{cases} \tag{1}$$

$$\epsilon_{\text{mean}} = \frac{1}{N}\sum_{i=1}^{N} \epsilon_i \tag{2}$$

$$\epsilon_{\text{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \epsilon_i^2} \tag{3}$$

The five candidate networks outlined in Table 4 have small values for ϵ_{mean}, indicating a lack of significant systematic bias. The networks also have similar values for ϵ_{RMS}, indicating that no network is substantially better at predicting pilot response than any other. Note that there are no ϵ_{mean} or ϵ_{RMS} values for the *Naive Bayes* and *fully disconnected* networks. Because of their network structures, these networks cannot predict pilot response probability based on parent node values and were only included as baselines for the Bayesian score.

IV. Selected Bayesian Network

A. Influence of Encounter Parameters on Pilot Response

The selected network optimally balanced the judging criteria outlined above and is depicted in Fig. 4. It was generated using the *Greedy Thick Thinning* structure learning algorithm with the constraint that an individual node could have no more than five parents, included to limit network complexity. In Fig. 4, the parents of the pilot response node ρ are enclosed in the dashed box and the child of ρ is underlined; shading and superscripts of each node indicate temporal layer; and black arrows indicate the links between ρ and its parents, with arrow thickness correlating to strength of influence.

In this network, pilot response probability is fully defined by the values of its five parents: *rate reversal*, *parallel approach*, *climb/descend*, *ground range*, and *vertical rate*. A *strength of influence* analysis, conducted using techniques described in the literature [23], showed that among the five parents, rate reversals and parallel approach encounters have the strongest influence on pilot response. The relative strengths of influence for each parent node are summarized in Table 5, which also contains results for the other candidate networks (the selected network is marked in bold).

Table 5 Normalized strength of influence on pilot response

Algorithm	RR	PL	CD	GR	VR	VFR
$\text{GTT}^{K2}_{k=8}$	0.24	0.27	0.12	0.13	0.12	0.13
$\mathbf{GTT^{K2}_{K=5}}$	**0.31**	**0.27**	**0.15**	**0.14**	**0.12**	——
GTT^{BDeu}	0.32	0.26	0.16	0.14	0.13	——
$\text{BS}_{k=5}$	0.67	——	——	0.18	0.16	——
$\text{BS}_{k=8}$	0.67	——	——	0.18	0.16	——

The overall pilot response probability for this dataset is 56%. Considering nonparallel approach encounters only, response probability becomes 62% overall, 58% for *climb* RAs, and 69% for *descend* RAs. These results are close to the response probabilities reported by the previously referenced studies of radar data, suggesting consistency in the encounter datasets and definitions of pilot response.

As mentioned in Sec. II.C, the numerical relationship between pilot response probability and its five parent nodes is captured in a lookup table. Table 6 contains a sample of this lookup table and includes those parent value combinations that are the most represented in the TRAMS encounter set, as indicated by the last column.

Figure 5 shows the probability distributions of *pilot response* and its five parent nodes in the dataset. In this figure, the y axis represents the proportion of the dataset corresponding to each node value. Discussions of each of the five parents follows. Some of these discussions reference parts of Table 7, which outlines pilot response probability for various subsets of the dataset. The values in this table correspond to the probability that each of the included nodes is *true* within the corresponding subset (when RR is true, it indicates a rate reversal; when PL is true, it indicates a parallel approach encounter; when CD is true, it indicates a *climb* RA; and when ρ is true, it indicates a response to the RA—the GR and VR nodes are not included). Values in bold indicate which subset of the encounter set is being examined. For example, if RR is set to 1, then the values of the other nodes correspond to the subset of encounters that contain a rate reversal. Part A of the table corresponds to the complete dataset and represents a baseline. The rightmost column indicates the size of each subset represented as a percentage of the complete dataset. A tabular breakdown of the dataset containing values for all nodes is included in the Appendix.

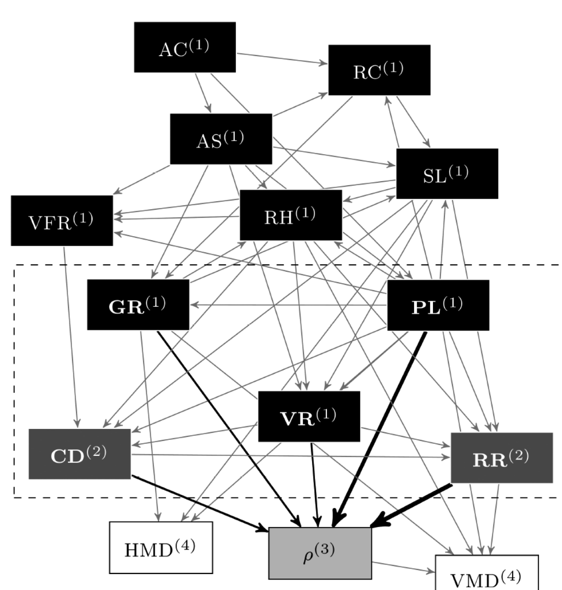


Fig. 4 The selected Bayesian network, constructed using the *Greedy Thick Thinning* algorithm.

Table 6 The eight rows of the lookup table that are the most represented in the TRAMS encounter set

Rate reversal	Climb/descend	Parallel	Ground range	Vertical rate	ρ	TRAMS
Boolean	*Binary*	*Boolean*	*n miles*	$\lvert fpm \rvert$	*Probability*	*Percentage*
False	Climb	False	[1, 2)	[0, 500)	0.5717	11.4
True	Climb	True	<1	[500, 1000)	0.0440	9.0
False	Climb	False	[2, 3)	[0, 500)	0.6999	7.1
False	Descend	False	[1, 2)	[0, 500)	0.5912	5.1
False	Descend	True	<1	[500, 1000)	0.9998	5.0
False	Climb	False	<1	[0, 500)	0.3887	4.6
False	Descend	True	<1	[0, 500)	0.7359	3.9
True	Climb	True	<1	[1000, 2000)	0.0523	3.8

1) *Rate reversal (RR)*: The data support the notion that pilots are less likely to respond to RAs that oppose their current flight path. As Part B of Table 7 shows, rate reversals are associated with *climb* RAs and a lower probability of pilot response. These associations remain true for both the parallel approach and nonparallel approach subsets of the dataset.

2) *Parallel (PL)*: Part C of Table 7 shows that parallel approaches are associated with a lower probability of pilot response: 45%. A study of pilot response to TCAS RAs in simulated parallel approaches found a comparable result of 40% [24]. The third line of Part C shows that, in 92% of the parallel approach encounters where the pilot did not respond, the RA was a *climb* advisory, which would notionally necessitate a go-around. Considering the potential disruption caused by go-arounds, TCAS's propensity to alert unnecessarily against parallel approach intruders, and the fact that pilots oftentimes have these intruders in sight, it is reasonable that pilot response rate would be relatively low for these operations.

3) *Climb/descend (CD)*: In keeping with the discussion so far, Part D of Table 7 shows that pilots are less likely to respond to *climb* RAs than *descend* RAs. This result agrees with other studies of pilot response to TCAS RAs [25] and is true even when considering only the nonparallel approach subset of the dataset.

4) *Ground range (GR)*: Although not outlined in Table 7, the data show that the probability of response is lower for RAs issued at smaller ground ranges. This is due in part to the strong correlation between ground range and parallel approaches, though it is also true when considering only nonparallel approach encounters. One plausible explanation for this observation is that lower ground ranges correlate to slower airspeeds (TCAS issues advisories based on time to CPA) and lower altitudes where visual acquisition of intruders is more likely.

Table 7 Probability of response for various parent node values

Part	RR	PL	CD	ρ	Subset, %
A	0.35	0.32	0.63	0.56	100
B	**1**	0.44	0.88	0.29	35.2
	1	**0**	0.78	0.47	19.5
	1	**1**	0.99	0.07	15.7
C	0.29	**0**	0.68	0.62	68.0
	0.49	**1**	0.54	0.45	32.0
	0.84	**1**	0.92	**0**	17.6
D	0.49	0.27	**1**	0.44	63.5
	0.12	0.40	**0**	0.77	36.5
	0.34	**0**	**1**	0.58	46.2
	0.19	**0**	**0**	0.69	21.8

5) *Vertical rate (VR)*: Compared with the other parent nodes, the correlation between vertical rate and pilot response is relatively weak. An examination of the data suggests that the relationship between the two nodes is potentially a consequence of the definition of pilot response used in this study.

Given these parent nodes, it is possible to calculate the combinations of node values that result in the highest and lowest probabilities of pilot response. These combinations are outlined in Table 8 for both the complete dataset and nonparallel approach encounters.

B. Influence of Pilot Response on Encounter Outcomes

While the previous subsection discussed the influence of encounter parameters on pilot response, this subsection discusses the

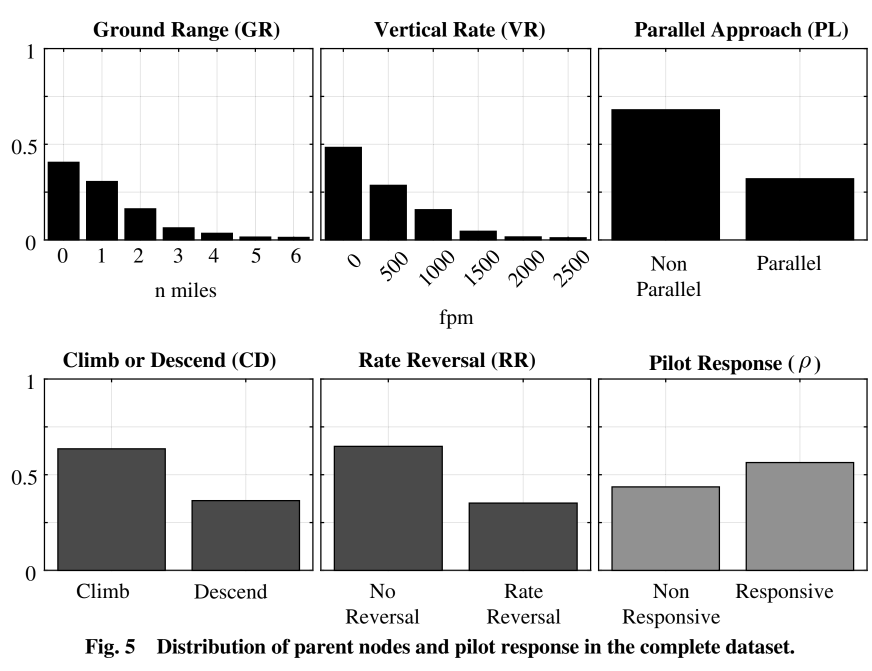


Fig. 5 Distribution of parent nodes and pilot response in the complete dataset.

Table 8 Parent node values for maximum and minimum probability of response

	Rate reversal	Climb/descend	Parallel	Ground range	Vertical rate	ρ	TRAMS
	Boolean	*Binary*	*Boolean*	*n miles*	*fpm*[a]	*Probability*	*Percentage*[b]
Complete dataset	False	Descend	True	<1	[−500, −1000)	0.9998	5.0
	True	Climb	True	<1	[−500, −1000)	0.0440	9.0
Nonparallels	False	Descend	False	[1, 2)	[−500, −1000)	0.9991	1.3
	True	Climb	False	<1	[−2000, −3000)	0.1875	0.06

[a]The sign of the vertical rate is inferred based on the RA type and whether there was a rate reversal.
[b]The percentage indicates the proportion of the TRAMS encounters that were used to create each row.

influence of pilot response on encounter outcomes. Note that the following results pertain to nonparallel approach encounters only.

As Fig. 4 shows, VMD is a direct descendant of the pilot response node, meaning that there is a direct probabilistic dependency between pilot response probability and VMD. And as Fig. 6a shows, pilot response correlates with higher values of VMD (note the spikes in VMD at the procedural vertical separations of 500 and 1000 ft for the nonresponsive distribution). There is no such correlation between pilot response and HMD shown in Fig. 6b, and in accordance with this, HMD is not a descendant of pilot response in the selected Bayesian network. This is expected, as TCAS RAs mitigate collision risk by increasing vertical separation and notionally have no effect on horizontal separation.

As mentioned previously, this study assigned a threshold of 400 fpm to pilot compliance with *climb* or *descend* RAs. The actual maximum vertical rates achieved for both compliance and noncompliance are shown in Figs. 7a and 7b, respectively. The mean values of the results in Fig. 7a are close to 1500 fpm: the rate advised by TCAS for *climb* and *descend* RAs. The shapes of these distributions and the differences between them support the validity of the pilot response definition used in this study by demonstrating that it is able to distinguish pilot compliance from noncompliance. If, for example, the mean values shown in Fig. 7a were farther away from 1500 fpm, it would suggest that the response definition might instead be confounding compliance with noncompliance.

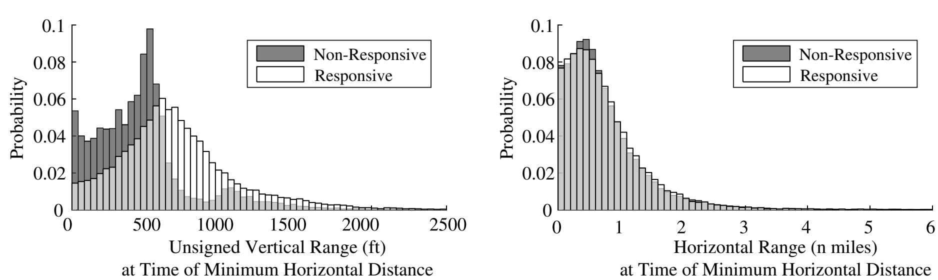
C. Performance of Selected Network Versus Nonparametric Response Models

The next section of this document fulfills one of its primary purposes: to gauge the effect on collision risk estimation of including encounter parameters in a response model. This is accomplished by comparing the TCAS safety benefit estimated using the Bayesian model developed for this study to that of a nonparametric model. Before addressing this topic, however, it is important to compare the *accuracy* of the Bayesian model to that of nonparametric models.

This comparison is captured in Fig. 8, which contains two graphs. To understand the x axis of each graph, consider that whereas parametric pilot response models are sensitive to the parameters of individual encounters, nonparametric models are not. Assuming a lack of stochasticity, these nonparametric models "predict" the same response probability in every encounter. The values of the x axis of each graph represent potential values of this fixed response probability. An x-axis value of 0.75, for example, corresponds to the nonparametric model in which the pilot response probability is 0.75 for every encounter. Because each x axis spans the full range of possible pilot response probabilities, each graph represents the full range of potential nonparametric models of the type described above.

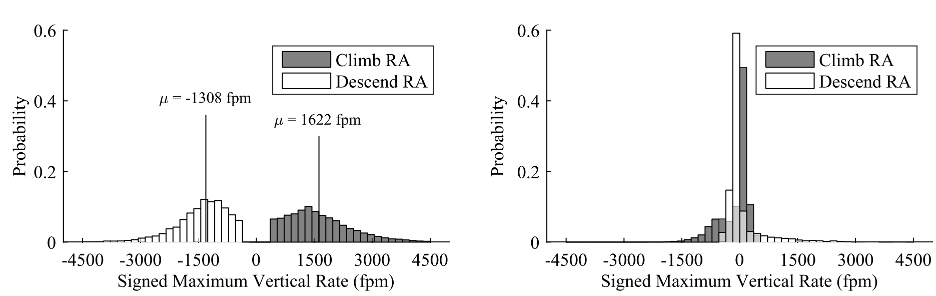
The y-axis values of each graph represent the model performance metrics outlined in Eqs. (1–3) (see Sec. III.D). The top graph depicts ϵ_{mean}, whereas the bottom graph depicts ϵ_{RMS}. In both cases, the calculations were based on the TRAMS encounters used to develop the Bayesian response model. Each graph also includes a dashed line, which depicts the respective value of each metric for the Bayesian model. The dashed lines are included for comparison and are not sensitive to the values of the x axis.

The dashed line on the top graph is close to 0, indicating that there is no substantial systematic bias in the Bayesian model's predictions. The solid line on this graph crosses the x axis at 0.56, indicating that the nonparametric model with the smallest systematic bias is the one with a fixed response probability of 56%. This value equals the mean



a) Vertical miss distance distributions **b) Horizontal miss distance distributions**

Fig. 6 Miss distance distributions as a function of pilot responsiveness.



a) Maximum vertical rate when pilot is responsive **b) Maximum vertical rate when pilot is non-responsive**

Fig. 7 Maximum vertical rate achieved during the RA, split by RA type and pilot responsiveness.

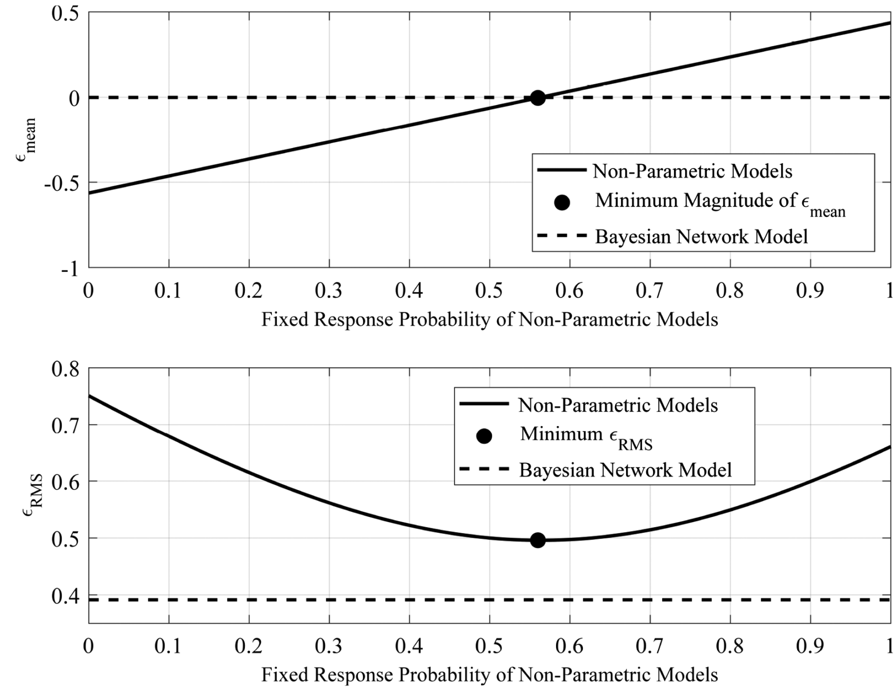


Fig. 8 Values of $\epsilon_{\mathbf{mean}}$ and $\epsilon_{\mathbf{RMS}}$ for nonparametric models and Bayesian model.

pilot response probability predicted by the Bayesian model, averaged across the TRAMS encounters (this value was first mentioned in Sec. IV.A). On the bottom graph, the curve representing the root mean squared error of the nonparametric models has a minimum at the same x-axis value of 0.56. Note that this curve exceeds the root mean squared error of the Bayesian model (the dashed line) for every fixed response probability. This indicates that, for the TRAMS encounter set, the predicted pilot response probability averaged across all encounters is more accurate using the Bayesian model than all possible nonparametric models.

The discussion above addresses the nonparametric model with a fixed response probability equal to the mean probability predicted by the Bayesian model. This nonparametric model is a special case for comparison to the Bayesian model. This is because the only substantive difference between the two models is how their predicted response probabilities, which average to the same number, are distributed among individual encounters. Any differences in the collision risk estimates based on the two models will, therefore, be due solely to the impact of including encounter parameters in the Bayesian model: a primary area of investigation of this study. This being so, the next section of this document will make heavy use of a nonparametric model similar to the one described above.

V. Safety Impact

A. Fast-Time Simulation Setup

The final step in this analysis was to assess the impact of the Bayesian pilot response model on the calculation of the safety benefit provided by TCAS. To accomplish this, the model was employed in simulations of safety-critical encounters where one or both aircraft were equipped with TCAS. Safety benefit was gauged using the *risk ratio* metric, which measures the effect of collision avoidance advisories on probability of NMAC.[**] Risk ratio is defined in Eq. (4).

$$\text{Risk Ratio} = \frac{P(\text{NMAC})\ \text{with CAS}}{P(\text{NMAC})\ \text{without CAS}} \tag{4}$$

The lower the risk ratio, the greater the safety benefit of the CAS. A risk ratio less than 1 indicates a net safety benefit, a risk ratio of 1 indicates no net effect on safety, and a risk ratio greater than 1 indicates a net safety detriment. Furthermore, TCAS was the sole means by which pilots could acquire and avoid intruders in these simulations. This means that any differences between the numerator and denominator of risk ratio could only be caused by the simulated pilot's response to TCAS RAs.

The simulated encounter set consisted of 3,976,080 two-aircraft encounters drawn from the LLCEM. As mentioned, the LLCEM models encountering aircraft trajectories based on U.S. radar data. The LLCEM encounter set used in this study was designed to incorporate a large number of encounters where CAS intervention is necessary to avert an NMAC. This makes the encounter set ideal for assessing the safety benefit of CAS advisories. This is contrasted with the TRAMS encounters, which do not include a significant number of encounters where an NMAC would have occurred without CAS intervention, as these encounters are rare in the airspace. This makes TRAMS unsuitable to assessing CAS safety benefit.

Despite their differences, both TRAMS recordings and LLCEM encounters represent operations in U.S. airspace. This begs the question: how can the LLCEM represent U.S. airspace if it contains a large number of safety-critical encounters? The answer is that the LLCEM assigns a likelihood-based weight to each encounter, with relatively high weights assigned to those encounters possessing a relatively high likelihood of occurring [15]. Safety-critical encounters, though they occur frequently in the encounter set, are assigned relatively low weights, reflecting their rarity in the airspace. In this analysis, aggregated statistics—including probability of NMAC and risk ratio—were calculated as weighted averages using these likelihood-based weights.

The encounter set was simulated in two ways: as encounters between two TCAS-equipped aircraft and as encounters between one TCAS-equipped aircraft and an intruder equipped with a Mode S transponder only. Version 7.1 of the logic was used for the TCAS aircraft. Surveillance noise conforming to standard error models was included, and both aircraft reported altitude with 25 ft quantization. In addition, when responsive to RAs, TCAS-equipped aircraft responded according to the standard model: for initial RAs, with 5 s of delay and $0.25g$ vertical acceleration; for subsequent RAs, with 2.5 s of delay and $0.35g$ vertical acceleration. Finally, a standard altimetry error model was employed in the calculation of NMAC probability [4]. With this model, NMAC probability is calculated from a distribution over vertical range at CPA, with higher probabilities corresponding to smaller vertical ranges.

**An NMAC occurs when encountering aircraft come within 500 ft horizontally and 100 ft vertically.

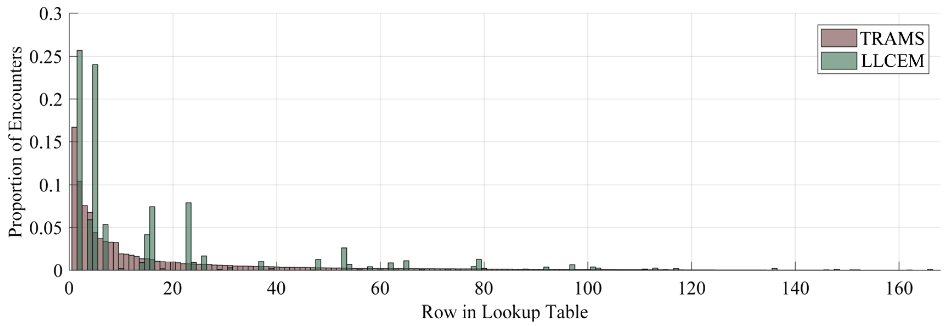


Fig. 9 Distribution of rows in the Bayesian model lookup table used by each encounter set.

The Bayesian network pilot response model was considered valid only for those encounters where the first corrective advisory issued by TCAS was a *climb* or a *descend* RA. Encounters beginning with *level off* RAs, for example, were not considered valid. In the simulated encounter set, the pilot response model was valid in approximately 20% of the encounters, when weighted. However, these encounters represent 88% of the nominal NMAC risk found in the encounter set (i.e., the risk without mitigation from TCAS). Given these factors, the following analysis considers only those encounters for which the pilot response model was considered valid.

For each valid encounter, pilot response probabilities for each TCAS-equipped aircraft were obtained from the lookup table based on their respective values for the *vertical rate*, *ground range*, *rate reversal*, and *climb/descend* nodes. These response probabilities were incorporated in the calculation of NMAC probability. Note that this calculation assumed that aircraft responded to either all or none of the RAs they received in individual encounters. Also note that because the LLCEM does not model parallel approaches, the *parallel approach* node was fixed at *nonparallel* for these calculations.

Two nonparametric pilot response models were also simulated and their resulting risk ratios compared with those of the Bayesian model, meaning that three separate pilot response models were evaluated in total. They are described below.

1) *Nonparametric 100%*: Response probability was 100% for all encounters.

2) *Nonparametric 66%*: Response probability was 66% for all encounters. This is equal to the mean response probability obtained from the Bayesian network model for the LLCEM encounter set. Section IV.C describes why the mean response probability is an important case for comparison to the Bayesian model. In that section, the mean response probability was 56%. The reason why a different value is used here is that 56% was the mean response probability for the TRAMS encounter set, while 66% is the mean response probability for the LLCEM encounter set.

3) *Bayesian Network Lookup*: A separate response probability was obtained from the lookup table for each encounter. *Bayesian Network Lookup* is the full implementation of the pilot response model developed in this study.

The progression of the pilot response models outlined above is from lower to higher sophistication and sensitivity to encounter parameters: the first model is completely naive to pilot noncompliance with RAs, assuming 100% response probability; the second model applies a constant, nonperfect pilot response probability identically among all encounters; and the third and most sophisticated model incorporates all of the relevant encounter parameters outlined in the Bayesian network of this study.

Note that while the average pilot response probability for the LLCEM encounter set was 66%, the corresponding value for nonparallel encounters in the TRAMS encounter set was 62% (this value was mentioned previously in Sec. IV.A). The closeness of these numbers suggests that the parent nodes of pilot response probability —with the exception of *parallel approach*—are distributed similarly between the two encounter sets, on the aggregate, which one would expect if both encounter sets are representative of U.S. airspace. It also supports the application of the TRAMS-based pilot response model to the LLCEM encounters. This topic is expanded on in the next subsection.

B. Comparison of TRAMS and LLCEM Encounter Sets

It is important to consider that, in this analysis, the Bayesian model was employed on an encounter set that it was not trained on. This was necessary: as mentioned, TRAMS recordings do not include enough safety-critical encounters for a collision risk analysis. But although both TRAMS and LLCEM are representative of U.S. airspace, there are differences between the two encounter sets that could affect estimates of pilot response probability and therefore collision risk.

One difference between the two encounter sets is the way in which their encounters, as categorized by the parent node values of the Bayesian model, are distributed. This is captured in Fig. 9. The values on the x axis of this histogram represent individual rows of the Bayesian model lookup table (see Sec. IV.A).[††] Recall that each row of this table corresponds to a unique combination of parent node values. In the figure, these rows have been sorted from high to low based on the number of TRAMS encounters corresponding to each combination. This number is represented on the y axis in normalized fashion (the bar values sum to 1). Rows with larger numbers of corresponding TRAMS encounters are data-rich and therefore their estimates of pilot response probability will have less uncertainty, whereas the opposite is true for rows with smaller numbers of corresponding encounters.

This figure also shows the distribution of lookup table rows for a critical subset of the LLCEM encounters (note that the values of Fig. 9 have been weighted based on the likelihood-based weights mentioned earlier). The subset is those LLCEM encounters for which an NMAC would occur without TCAS intervention, chosen because these encounters have the largest effect on collision risk estimation. From this figure, we can make several observations:

1) As with the TRAMS distribution, for the LLCEM distribution, higher encounter counts occur on the left and lower encounter counts occur on the right. This indicates that parent node combinations that occur frequently in TRAMS also occur frequently in the LLCEM, with the inverse also being true.

2) The LLCEM encounters are concentrated into relatively few lookup table rows when compared with the TRAMS encounter counts.

3) Some lookup table rows with higher numbers of corresponding TRAMS encounters have few or no corresponding LLCEM encounters (this observation is a consequence of the previous one).

4) Some lookup table rows with few or no corresponding TRAMS encounters have significant numbers of corresponding LLCEM encounters.[‡‡]

Based on these observations, there are no obvious differences between the two encounter sets significant enough to suggest that the Bayesian model should not be applied to LLCEM encounters. However, it is also important to acknowledge potential differences among variables not captured by this analysis. For example, as previously mentioned, the visual acquisition of intruders by pilots can be expected to strongly influence pilot response to RAs. Visual

[††]In this figure, 168 table rows are represented: half of the number of rows in the full lookup table. Only those rows for which the *parallel approach* node is set to *nonparallel* have been included, as the purpose of this figure is for comparison to the LLCEM, which contains no parallel approaches.

[‡‡]For the Bayesian model, pilot response probability was estimated to be 0.5 for those rows of the lookup table with no corresponding TRAMS encounters, as 0.5 was the Bayesian prior used to construct the model.

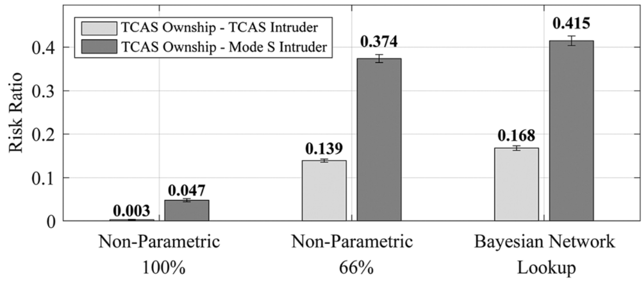


Fig. 10 Risk ratio evaluated for the simulated pilot response models.

acquisition is not an accessible variable among the TRAMS encounters or the LLCEM encounters, nor does it factor into the safety simulations. By not including such variables, there is the potential for qualitative differences among TRAMS and LLCEM encounters belonging to the same lookup table row, impacting the validity of the Bayesian model's predictions. This topic is addressed further in the following subsection.

C. Results and Discussion

The risk ratio results for these encounters are shown in Fig. 10. These results are accompanied by error bars representing confidence intervals of approximately 95%, calculated using a bootstrap approach [26]. From a comparison of *Nonparametric 100%* and *Nonparametric 66%*, one can see that lower pilot response probabilities correspond to higher risk ratios, as one would expect. However, the critical comparisons are among the risk ratios calculated using the *Nonparametric 66%* and *Bayesian Network Lookup* models. Recall that 66% is the result when the pilot response probabilities obtained from the Bayesian network are averaged across all of the simulated encounters. This means that the *Nonparametric 66%* and *Bayesian Network Lookup* models have the same response probability, on average—the only practical difference between the two is that one applies 66% response probability to all encounters and the other applies distinct, encounter-specific response probabilities that average out to 66%. In other words, one is sensitive to the parameters of individual encounters and the other is not. This being so, if the risk ratio results obtained from these models were identical, it would suggest that it makes no aggregate difference whether or not a pilot response model is sensitive to the parameters of individual encounters. But as one can see in Fig. 10, this is not so: risk ratio is higher for the *Bayesian Network Lookup* model than for the *Nonparametric 66%* model: 21% higher for the TCAS-TCAS case and 11% higher for the TCAS-Mode S case. This suggests a critical result: using encounter-agnostic pilot response models can result in an underestimation of collision risk.

The reason that there is a difference between the *Nonparametric 66%* and *Bayesian Network Lookup* results is explored in the following figures. Figure 11 compares probability of pilot response to benefit of pilot response. In this analysis, *benefit* is defined as the reduction in probability of NMAC that is brought about by the pilot following TCAS RAs. For example, if in some encounter there is a nominal (i.e., without TCAS intervention) NMAC probability of 0.5 and responding to TCAS RAs reduces it to 0.1, then the RA response benefit for that encounter would be 0.5–0.1 = 0.4. As this example suggests, response benefit is a function of both nominal collision risk and TCAS RA effectiveness. Figure 11 also depicts the fixed response probability of the *Nonparametric 66%* model as a dashed line.[§§] Figure 12 depicts the relative likelihood of each value of response benefit in the simulated encounter set, when weighted (note the logarithmic scale on the y axis). From this figure, we can see how encounters with relatively small response benefits are much more likely than those with relatively large response benefits. This is because the simulated encounter set is representative of the NAS, where the collision risk of most encounters is small even without TCAS intervention, meaning that the benefit of responding to TCAS RAs in these encounters is naturally limited. Note that in Figs. 11 and 12, benefit was binned in 0.001 increments and the mean y-axis value is plotted for each bin.

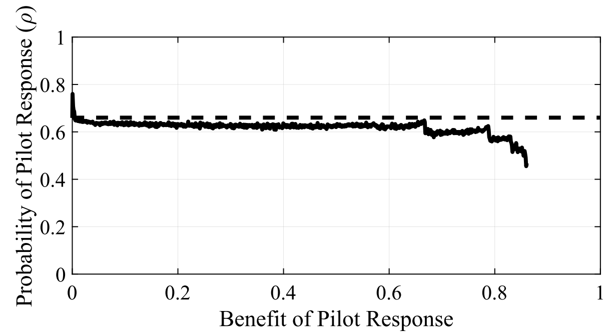


Fig. 11 Probability of pilot response relative to benefit of pilot response for two TCAS-equipped aircraft.

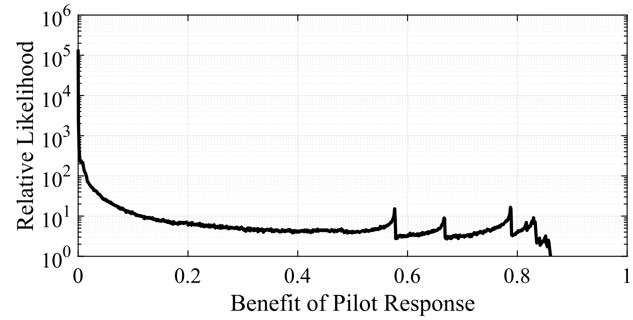


Fig. 12 Relative encounter likelihood as a function of pilot response benefit.

In Fig. 11, there is a general trend suggesting that as response benefit increases, response probability decreases. This is especially noticeable for response benefits close to 0 and greater than 0.66.[¶¶] This trend is the reason why estimated collision risk was lower for the *Nonparametric 66%* model than for the *Bayesian Network Lookup* model: the *66%* model assigned higher response probabilities to those encounters where response has a greater benefit, decreasing the overall estimate of collision risk relative to the *Lookup* model. This is another critical result. It suggests that whenever there is a correlation between the probability and benefit of RA response, simulations using an averaged pilot response probability will either underestimate collision risk (for negative correlations) or overestimate collision risk (for positive correlations). Furthermore, the extent of these over- or under-estimations will be a function of the strength of the correlation, the distribution of nominal collision risk in the encounter set, and the effectiveness of the RAs issued by the CAS or DAA system.

The drop-off in response probability for response benefits greater than 0.66 bears further discussion. The simple explanation for this drop-off is that the simulated LLCEM encounters with these response benefits correspond to rows of the lookup table with relatively low response probabilities. That being said, further insight can be gained by comparing these LLCEM encounters to the TRAMS encounters that were used to build the same lookup table rows. In the LLCEM encounters, the two encountering aircraft are typically flying head-on, level trajectories, and in all cases, RA response is required to avoid a potential collision. The TRAMS encounters, on the other hand, are more diverse. In some cases, these encounters took place during structured airspace procedures in which real pilots would likely have the intruder in sight and therefore might safely ignore TCAS RAs.[***] Therefore, for these rows of the lookup table, there is a

[§§]The maximum value of response benefit is approximately 0.86. This corresponds to the highest possible value of NMAC probability when calculated using the altimetry error model mentioned in Sec. V.A.

[¶¶]It is a coincidence that this value is 0.66 while the average pilot response probability is 66%.

[***]In many of these encounters, the aircraft are on parallel trajectories. These encounters do not constitute approaches to parallel runways (they may instead be, for example, parallel departures) and were therefore not captured by the PL node of the Bayesian network (see Sec. III.B).

discrepancy in the benefit of RA response between the encounters used to build the model (TRAMS) and the encounters the model is being applied to (LLCEM). This illustrates an important aspect of the method outlined in this study: if a pilot response model built from operational encounters is to be used in simulations with safety-critical encounters, then the qualitative differences between the two encounter sets may affect estimates of collision risk and therefore must be understood, even if both encounter sets are representative of the same airspace.

Despite this discrepancy, the results demonstrate that not including encounter parameters in pilot response models can result in an underestimation of collision risk. Underestimating collision risk has many consequences, one of which is the masking of undesired system behavior. For example, given the choice between a *climb* and a *descend* RA in some encounter, a CAS or DAA logic may choose *climb* because by some standard response model it results in a safer outcome. However, a higher-fidelity model may reveal that pilots are more likely to respond to the *descend*, making it ultimately safer. If millions of encounters such as this one are incorporated into the development and evaluation of a collision avoidance system, then incorporating a higher-fidelity pilot response model could result in a safer system. In addition, lower-fidelity models may create unfair comparisons between multiple CAS or DAA logics simulated on the same encounter set. Consider a pair of logics, one of which issues advisories such that response probability and benefit are negatively correlated (as above) and the other of which does not. Simulations using a simple pilot response model such as *Nonparametric 66%* would unfairly favor the first logic, confounding the results.

VI. Conclusions

The purpose of this study was to demonstrate the construction and safety impact of a pilot response model that is sensitive to the parameters of individual encounters. A model was built from operational TCAS data incorporated into a Bayesian network. Within this model, pilot response to TCAS *climb* and *descend* RAs was shown to be sensitive to five encounter parameters: parallel approach, rate reversal, vertical rate, RA type (*climb* or *descend*), and ground range. The model was then employed in simulations of safety-critical encounters and compared with other pilot response models. The results demonstrated that encounter-agnostic (nonparametric) pilot response models can underestimate collision risk, potentially impacting the design and safety benefit of separation advisory systems.

Any conclusions drawn from this study must recognize its limitations. These limitations include those of the TRAMS data source, which was described in Sec. II.B. These limitations also include the definition of pilot response used in this study, its application to *climb* and *descend* RAs only, and the qualitative differences between the TRAMS and LLCEM encounter sets.

Appendix: TRAMS Probability Distribution Data

Table A1 contains the probability distributions for all nodes in the Bayesian network used in this study, based on the TRAMS encounter set.

Acknowledgments

This work is sponsored by the Federal Aviation Administration 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. The authors gratefully acknowledge Neal Suchy and the FAA TCAS Program Office for their leadership and support. We also thank Mykel Kochenderfer, Wes Olson, Rod Cole, Jim Kuchar, Shirley Phillips, Tom Teller, Robert Klaus, Ian Jessen, and Matthew Edwards for their comments and insight. GeNIe, QGeNIe, and SMILE Software are licensed from the University of Pittsburgh. This paper was approved for public release. Distribution is unlimited.

References

[1] Kuchar, J., and Drumm, A. C., “The Traffic Alert and Collision Avoidance System,” *Lincoln Laboratory Journal*, Vol. 16, No. 2, 2007, pp. 277–296.

[2] Pritchett, A., “Pilot Non-Conformance to Alerting System Commands During Closely Spaced Parallel Approaches,” Ph.D. Thesis, Massachusetts Inst. of Technology, Cambridge, MA, 1997.

[3] Federal Aviation Administration, “Introduction to TCAS II Version 7.1,” Booklet, Feb. 2011.

[4] International Civil Aviation Organization, *Annex 10–Aeronautical Telecommunications, Volume 4–Surveillance and Collision Avoidance Systems*, 5th ed., International Civil Aviation Organization, ICAO Document 8168, 2014.

[5] Londner, E. H., “Collision Avoidance System Effectiveness on Low Performance Unmanned Aircraft,” *AIAA Infotech@Aerospace*, AIAA Paper 2016-1987, 2016. doi:10.2514/6.2016-1987

[6] Espindle, L., Griffith, J., and Kuchar, J., “Safety Analysis of Upgrading to TCAS Version 7.1 Using the 2008 U.S. Correlated Encounter Model,” Tech. Rept. ATC-349, Massachusetts Inst. of Technology Lincoln Lab., Lexington, MA, 2009.

[7] Holland, J. E., Kochenderfer, M. J., and Olson, W. A., “Optimizing the Next Generation Collision Avoidance System for Safe, Suitable, and Acceptable Operational Performance,” *Air Traffic Control Quarterly*,

Table A1 Probability distributions of all nodes in the Bayesian network

AC category		Airspace		Rel. course, deg		HMD, n miles		VMD, ft	
0.08	Major air carrier	0.44	Class B	0.28	0	0.15	0	0.23	0
0.29	Regional air carrier	0.02	Class C	0.08	45	0.29	0.25	0.22	250
0.26	Business jet	0.06	Class D	0.08	90	0.20	0.50	0.26	500
0.36	Helicopter	0.01	Special Use	0.12	135	0.14	0.75	0.15	750
>0.00	Other	0.46	Class E, G	0.17	180	0.09	1.00	0.07	1000
0.01	Unknown	0.02	Class A	0.12	225	0.05	1.25	0.07	1250+
				0.08	270	0.03	1.50		
				0.08	315	0.02	1.75		
						0.01	2.00		
						0.03	2.25+		
Vertical rate, fpm		Rel. altitude, ft		Ground range, n miles		TCAS SL		Pilot response	
0.48	0	0.54	0	0.40	0	0.31	SL3	0.44	No response
0.28	500	0.37	400	0.30	1	0.29	SL4	0.56	Response
0.16	1000	0.03	800	0.16	2	0.28	SL5		
0.05	1500	0.02	1200	0.06	3	0.10	SL6		
0.02	2000	0.02	1600	0.04	4	0.02	SL7		
0.01	2500+	0.02	2000+	0.02	5				
				0.01	6+				
Beacon code		Climb/descend		Rate reversal		Parallel			
0.71	Discrete	0.63	Climb	0.65	No	0.68	Nonparallel		
0.29	1200	0.37	Descend	0.35	Yes	0.32	Parallel		

Vol. 21, No. 3, July 2013, pp. 275–297.
doi:10.2514/ATCQ.21.3.275

[8] Chryssanthacopoulos, J. P., and Kochenderfer, M. J., "Collision Avoidance System Optimization with Probabilistic Pilot Response Models," *Proceedings of the American Control Conference*, IEEE Publ., Piscataway, NJ, 2011, pp. 2765–2770.
doi:10.1109/ACC.2011.5990776

[9] Olson, W. A., and Olszta, J. E., "TCAS Operational Performance Assessment in the US National Airspace," *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference*, IEEE Publ., Piscataway, NJ, 2010, pp. 4.A.2.1–11.
doi:10.1109/DASC.2010.5655351

[10] Pritchett, A., and Fleming, E. S., "Pilot Compliance to TCAS Resolution Advisories," *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference*, IEEE Publ., Piscataway, NJ, 2013, pp. 1–13.
doi:10.1109/DASC.2013.6712618

[11] Kochenderfer, M., *Decision Making Under Uncertainty*, MIT Press, Cambridge, MA, 2015, pp. 11–55, Chap. 2.

[12] International Civil Aviation Organization, *Manual on Remotely Piloted Aircraft Systems (RPAS)*, 5th ed., International Civil Aviation Organization, ICAO Document 10019, 2015.

[13] Federal Aviation Administration, "Air Carrier Operational Approval and Use of TCAS II," Advisory Circular 120-55C Change 1, March 2013.

[14] Eurocontrol, "TCAS II Performance in European TMAs; Part 1: Analysis," Eurocontrol, Feb. 2009.

[15] Kochenderfer, M., Espindle, L., Kuchar, J., and Griffith, J. D., "Correlated Encounter Model for Cooperative Aircraft in the National Airspace System Version 1.0," Massachusetts Inst. of Technology Lincoln Lab. Tech. Rept. ATC-344, Lexington, MA, 2008.

[16] Coombes, M., Chen, W.-H., and Render, P., "Site Selection During Unmanned Aerial System Forced Landings Using Decision-Making Bayesian Networks," *Journal of Aerospace Information Systems*, Vol. 13, No. 12, 2016, pp. 491–495.
doi:10.2514/1.I010432

[17] Johansson, F., and Falkman, G., "A Bayesian Network Approach to Threat Evaluation with Application to an Air Defense Scenario," *11th International Conference on Information Fusion*, IEEE Publ., Piscataway, NJ, 2008, pp. 1–7.
doi:10.1109/ICIF.2008.4632368

[18] Druzdzel, M. J., "SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A Development Environment for Graphical Decision-Theoretic Models," *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, American Association for Artificial Intelligence, Menlo Park, CA, 1999, pp. 342–343

[19] Cheng, J., Bell, D. A., and Liu, W., "An Algorithm for Bayesian Belief Network Construction from Data," *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Soc. for Artificial Intelligence and Statistics, 1997, pp. 83–90.

[20] Heckerman, D., Geiger, D., and Chickering, D. M., "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data," *Machine Learning*, Vol. 20, No. 3, 1995, pp. 197–243.
doi:10.1023/A:1022623210503

[21] Cooper, G. F., and Herskovits, E., "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, Vol. 9, No. 4, 1992, pp. 309–347.
doi:10.1007/BF00994110

[22] Buntine, W., "Theory Refinement on Bayesian Networks," *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, Burlington, MA, 1991, pp. 52–60.
doi:10.1016/B978-1-55860-203-8.50010-3

[23] Koiter, J. R., "Visualizing Inference in Bayesian Networks," Ph.D. Thesis, Delft Univ. of Technology, 2006.

[24] Pritchett, A., Carpenter, B., Asari, K., Kuchar, J., and Hansman, R., "Issues in Airborne Systems for Closely-Spaced Parallel Runway Operations," *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference*, IEEE Publ., Piscataway, NJ, 1995, pp. 140–145.
doi:10.1109/DASC.1995.482820

[25] Pritchett, A., Haga, R., and Thakkar, D., "Pilot Responses to Traffic Events During NextGen High Traffic Density Terminal Operations," *Proceedings of the IEEE/AIAA Digital Avionics Systems Conference*, IEEE Publ., Piscataway, NJ, 2014, pp. 2C4–1.
doi:10.1109/DASC.2014.6979429

[26] Efron, B., "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, Vol. 7, No. 1, 1979, pp. 1–26.
doi:10.1214/aos/1176344552

H. R. Idris
Associate Editor