Project Report
ATC-405

Risk-based Modeling to Support NextGen Concept Assessment and Validation

E. Harkleroad
A. Vela
J. Kuchar
B. Barnett
R. Merchant-Bennett

1 March 2013

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Federal Aviation Administration,
Washington, D.C. 20591

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
ATC-405		

4. Title and Subtitle	5. Report Date
Risk-based Modeling to Support NextGen Concept Assessment and Validation	1 March 2013
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
E. Harkleroad, A. Vela, and J. Kuchar, MIT Lincoln Laboratory B. Barnett and R. Merchant-Bennett, Federal Aviation Administration	ATC-405

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108	11. Contract or Grant No. FA8721-05-C-0002

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Department of Transportation Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	Project Report
	14. Sponsoring Agency Code

15. Supplementary Notes

This report is based on studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology, under Air Force Contract FA8721-05-C-0002.

16. Abstract

This report provides a brief review of major recent risk-based modeling (RBM) approaches, with particular emphasis on how these tools can be applied during initial Next Generation Air Transportation System (NextGen) concept development and how their use can be validated.

Effective safety analysis should begin as early as possible during a system's life cycle in order to have maximum impact. Ideally, safety considerations should play a role even during a new system's concept definition and development. Elements of NextGen are currently progressing through these early phases.

NextGen will increasingly rely on integrating multiple systems and information sources together to enable improved efficiency, safety, and reduced environmental impact. Ensuring that such complex interconnected systems are developed to meet safety goals requires corresponding advances in RBM and safety assessment approaches.

This report does not cover the more detailed safety analyses that must be applied to mature system concepts. Rather, the focus is on approaches for hazard identification, scoping, and coarse risk estimation for systems in the early conceptual development stage, when details on the design and operation of the system have yet to be resolved. Risk models applied in this constrained context cannot be expected to provide the same complete, quantitative results as they do for mature systems.

Following a review of prior models, this report continues with recommendations for RBM development, application, validation, and coordination between NextGen efforts. Also, a discussion on safety and concept development is provided.

17. Key Words	18. Distribution Statement
	This document is available to the public through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	69	

This page intentionally left blank.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the input received from reviewers Ben Ale, Henk Blom, Nancy Leveson, James Luxhoj, and Eric Perrin. We are also grateful to Roland Weibel for his contributions to this report.

This page intentionally left blank.

TABLE OF CONTENTS

		Page
	Acknowledgments	iii
	List of Illustrations	vii
	List of Tables	ix
1.	INTRODUCTION	1
2.	MODEL ANALYSIS	5
	2.1 Analysis Factors	5
	2.2 Review of Models	9
	2.3 Synthesis	19
3.	GENERAL RECOMMENDATIONS	27
	3.1 Conservatism and Hazard Identification	28
	3.2 Systemic Failures	30
	3.3 Uncertainty	30
	3.4 Emergent Behaviors	31
4.	RECOMMENDATIONS FOR CONCEPT VALIDATION	35
	4.1 Integrating Safety and Risk Modeling into Stakeholder Business Processes	35
	4.2 Recommended Concept Validation Methods	37
	4.3 Verification and Validation of Decision Support Algorithms	40
5.	RECOMMENDATIONS FOR MODELING AND MODEL VALIDATION	43
	5.1 Modeling Assumptions for Initial Concept Development	43
	5.2 Model Validation	44
6.	MODELING SOFTWARE RECOMMENDATIONS	47
	6.1 Software Design and Interface of Risk Models	47
	6.2 Databases and Information Sharing	48

TABLE OF CONTENTS
(Continued)

	Page
Glossary	53
References	57

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Example of NextGen System Integration—Trajectory-Based Operations Components [2]	2
2	Space of System Concepts with Potential for Safety Assessment	3
3	Models Reviewed by Model Type	6
4	Generic SADT Task [10]	21
5	En Route SADT Model [10]	21
6	Relative Importance of Quantitative and Expert Data Sources by Model	23
7	Summary of I2I Steps and Incremental RBM Steps	38
8	Summary of Concept Maturity Levels (CMLs)	39
9	Well-Posed Mapping by a Decision Support Tool	43
10	Ill-Posed Mapping by a Decision Support Tool	44

This page intentionally left blank.

LIST OF TABLES

Table No.		Page
1	Sources for Models Reviewed	10
2	Model Status and Implementation Details	24
3	Validation Methods Applied to Reviewed Models	25
4	Example of Potentially Hazardous Control Actions by the Flight Crew during In-Trail Procedures [9]	29
5	Recommended Models by Concept Maturity Level	40
6	Summary of Recommendations	49

This page intentionally left blank.

1. INTRODUCTION

This report provides a brief review of major recent risk-based modeling (RBM) approaches, with particular emphasis on how these tools can be applied during initial NextGen concept development and how their use can be validated.

Effective safety analysis should begin as early as possible during a system's life cycle in order to have maximum impact. Ideally, safety considerations should play a role even during a new system's concept definition and development [1]. Elements of the Next Generation Air Transportation System (NextGen) are currently progressing through these early phases.

NextGen will increasingly rely on integrating multiple systems and information sources together to enable improved efficiency, safety, and reduced environmental impact. Ensuring that such complex interconnected systems are developed to meet safety goals requires corresponding advances in RBM and safety assessment approaches. Homogeneous safety analysis tools used in the past—such as fault trees—for relatively self-contained systems cannot simply be expanded to cover these larger and more complex interactions. Influence-based methods—discussed in Section 2.1—may also need enhancements. Furthermore, safety analysis of systems that are not fully described can be challenging because of the potential for unforeseen interactions. Additional techniques—such as Bayesian Belief Networks (BBNs) and hierarchical control structure models—are gaining greater acceptance; they may capture additional system interactions and even predict unforeseen interactions. However, there is a need to carefully scope and organize how these tools and methods are applied and where new development is still required.

As shown in Figure 1, for example, midterm Trajectory Based Operations (TBO) will require components and interactions spanning ground automation systems, Automatic Dependent Surveillance-Broadcast (ADS-B) surveillance, cockpit flight management systems and displays, precision navigation, datacomm, new operating procedures, and communications and collaboration tools between cockpit, facilities, and airlines—all while also supporting legacy systems and procedures as the National Airspace System transitions into NextGen.

TRAJECTORY BASED OPS	ADS-B	DataComm	NNEW	NVS	SWIM	ERAM	TFMS	ASDE-X	LAAS	RNP	RNAV	AIM	TBFM	CATMT	Procedures
MID-TERM															
Delegated Responsiblitity for Separation	X					X									X
Oceanic In-Trail Climb and Descent		X								X	X				X
Automation Support for Mixed Environment						X									
Initial Conflict Resolution Advisories		X				X									
Flexible Entry Times for Oceanic tracks			X												
Point in Space Metering						X	X						X		
Flexible Airpsace Management			X	X	X	X	X								X
Increase Capacity and Efficiency Using RNAV and RNP						X				X	X			X	X
Provide Interactive Flight Planning from Anywhere		X			X	X	X					X			

Figure 1. Example of NextGen System Integration—Trajectory-Based Operations Components [2]

This report does not cover the more detailed safety analyses that must be applied to mature system concepts. Rather, the focus is on hazard identification, scoping, and coarse risk estimation for systems in the early conceptual development stage, when details on the design and operation of the system have yet to be resolved. It also does not cover tools designed to implement the Safety Management System (SMS) for mature systems. The NextGen safety assessment process, in early concept phases, should support SMS, but may take a different form.

There are several challenges for systems in early development: the available system description may be too vague or the system's complexity may exceed the capabilities of existing modeling tools. These ideas are illustrated in Figure 2. Assumptions about the future fully-implemented system will be required. To reduce complexity for modeling purposes, some details may need to be abstracted away, which could obscure potentially unsafe interactions.

Risk models applied in this constrained context cannot be expected to provide the same complete, quantitative results as they do for mature systems. Analysts must interpret model outputs appropriately; they may obtain results which are qualitative, directional only, or less precise than a model might otherwise produce. Such results provide general insights to guide further concept development.

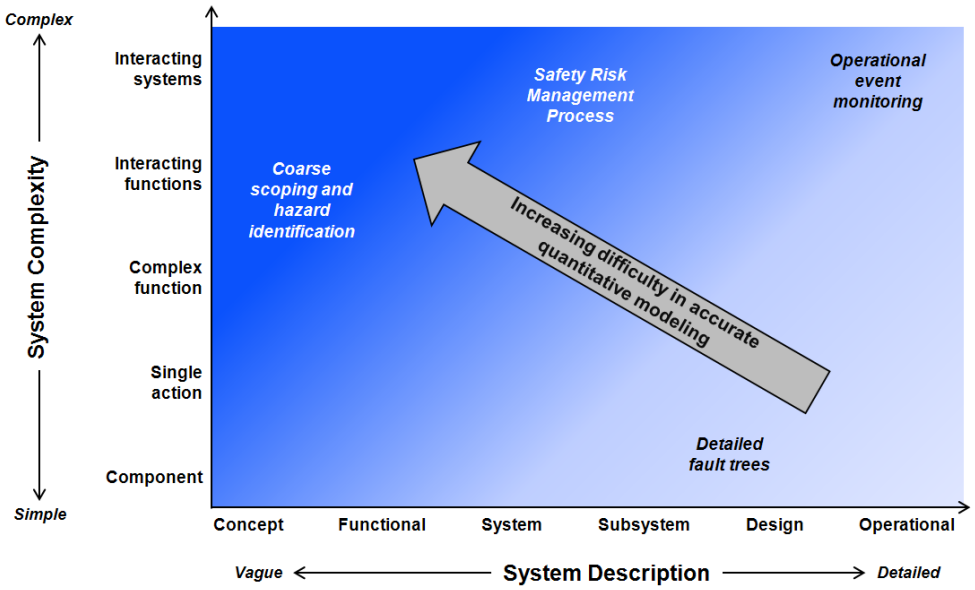

Figure 2. Space of System Concepts with Potential for Safety Assessment

This report complements other recent reviews of modeling techniques which have different objectives and emphasis. For example, a EUROCONTROL survey described several hundred risk-modeling techniques and selected subsets of these with immediate and long-term relevance to the organization's safety assessment needs [3]. That report thoroughly documents the advantages and disadvantages of individual techniques. However, safety assessment for NextGen will likely require integrated models based on a variety of techniques and data sources; therefore this report will review integrated models rather than individual modeling techniques. Similarly, the FAA Risk Assessment & Risk Management (RARM) effort seeks to review, prioritize, and recommend risk modeling techniques but does not appear to consider applicability to initial concept development [4].

Following the review of prior models, this report continues with a few key recommendations for RBM development, application, validation, and coordination between NextGen efforts. Also a discussion on safety and concept development is provided.

This page intentionally left blank.

2. MODEL ANALYSIS

To better understand risk-based models and to place them into the proper context, a framework for analysis is first provided. The analysis factors considered include the scope of the model, the data sources and methods used for construction, prior or continuing validation work, and the feasibility of applying the model to NextGen concepts. Next, descriptions of several relevant risk-based models are given using the established analysis framework.

2.1 ANALYSIS FACTORS

Due to the complex nature of NextGen concepts and the variety of RBM approaches that might be applied, it is helpful to first list attributes or dimensions of a taxonomy in which modeling and analysis can be organized. The set of all possible combinations of values for these attributes defines a multi-dimensional space of potential models.

Dimensions for organizing models and data considered in this report include:

Scope of Systems

Models may cover the influences and behavior of equipment, software, human operators, procedures, management, and regulatory oversight. As with the other dimensions, these influences may interact in various ways.

Range of Events (Phases of flight and accident types)

Example phases of flight include taxi, take-off, climb, en route, approach, and landing, which may be grouped in different ways. Some reviewed models treat flight phases separately since they are separated in time and space; however, they may be correlated when events during one phase affect future phases.

Accident types include controlled flight into terrain (CFIT), runway collision, and mid-air collision, among others. Causes may include aircraft flight control system failure, flight crew spatial disorientation, and fires onboard aircraft. Models reviewed cover various additional accident types and categorization methods. Some models assume that different accident types are somewhat independent; however, this approach may not be ideal for accident types that are correlated due to common causes.

Model Approach and Methods

A risk-based model's approach and methods describe how the model uses inputs such as data sources and assumptions to derive outputs describing the risk picture. Risk-based models can generally be classified into two categories:

1. Event-based: risk is modeled as a result of various possible event sequences or transitions between discrete system states.

2. Influence-based: risk is modeled as a result of system-wide factors such as management oversight, operator training, and maintenance practices, all of which influence the likelihood of discrete events.

These categories are not mutually exclusive—models reviewed here include elements of both (Figure 3). Combining the two techniques can improve flexibility by allowing modeling of interacting influences on events leading up to accidents. Events represent discrete occurrences that can be assigned a probability of occurrence. Influences are processes or measures that affect the likelihood of events occurring, where a probability assignment indicates a degree of belief.

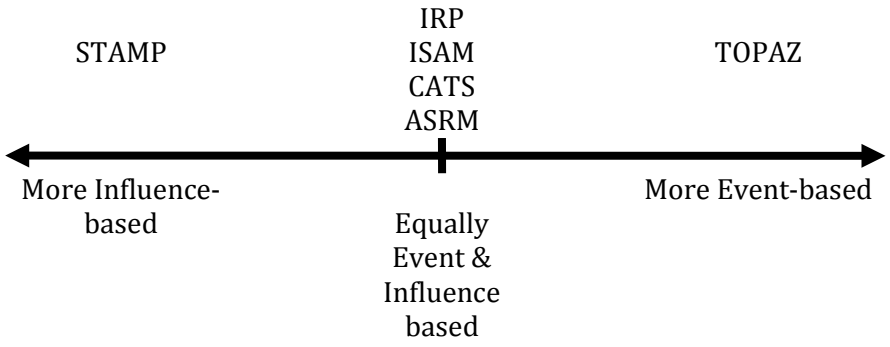

Figure 3. Models Reviewed by Model Type

Methods like STAMP are not void of event consideration. In fact, STAMP can be described as modeling the influences or causes of events. Similarly, TOPAZ considers sociotechnical interactions, which occur through influences.

Various quantitative modeling approaches are possible, including probabilistic/analytic models and fast-time Monte Carlo models [3,5]. A single model may combine different approaches to optimize their respective advantages across different modeling dimensions and modeling goals.

Data Sources

All models reviewed seek a basis in real-world numerical data. Sources include archived records of historical operations, accident reports, human-in-the-loop experiments, and equipment vendors. Data-related challenges noted by model developers include lack of data, insufficient data, and insufficient

access to subject-matter experts (SMEs) [6]. In cases where numerical data are not available, expert opinion is a possible substitute; this input may be used to judge the relative importance of different influences on the risk picture.

Outputs

Model outputs typically include a risk picture providing the risk of accidents by type, cause, or flight phase and the influence of different factors on the risk picture. Predicted accident rates may be compared to a target level of safety or to historical accident rates in order to judge whether a new system is safe.

Applicability to NextGen

The feasibility of models for application to NextGen may be judged on several criteria. The criteria considered in this report include:

1. Modeling of NextGen Operational Improvements (OIs). Models should represent the effect of each solution set and OI on the NAS risk picture. Models which explicitly include OI influences will require less additional development to produce the desired NextGen risk picture.

2. Scope of system and event coverage. NextGen OIs will potentially affect all flight phases and all human operators, such as flight crew, air traffic control (ATC), and maintenance. Hence the most applicable models will comprehensively cover a wide variety of risks. If complete coverage of all NextGen OIs is not feasible, OIs could be prioritized by planned implementation date, expected benefit, or some other metric.

3. Model status and implementation. All other metrics being equal, fully implemented and validated models are preferable to models still in development or requiring additional validation. Ideally, models should be realized in software and documented for users.

4. Modularity. For models without the complete coverage required to produce a NextGen risk picture, a potential remedy is to combine elements from different models. Certain model structures, such as BBNs, are modular, which facilitates this approach.

Model Validation

Model validation is important as one dimension for understanding and comparing risk models. Models are tools used for concept assessment and validation—a more general process described in a FAA concept development report [7].

With each modeling approach, a corresponding validation process is required. Here, validation refers to the process of determining and ensuring that model components, both individually and in

composition, correspond to reality. The methods for validating risk models differ depending upon the framework used to compose the model, as well as the level of validation required. Generally the approaches can first be classified into two categories: qualitative and quantitative validation.

Qualitative validation utilizes subjective judgment to determine in a general sense that system behavior occurs as modeled. It does not provide an analytical degree of belief in results, but relies on independent expert judgment to assess the validity of a model. In general, qualitative validation can address the following questions:

- Does the model cover all relevant events?
- Do modeled event sequences follow reality?
- Do influences have appropriate direction and magnitude?
- Do resulting outputs have appropriate direction and magnitude?

Qualitative validation requires a fundamental correspondence between the assessed situation and the model. This requires that model events are at the same level of abstraction as those observed, or can be aggregated or deconstructed as appropriate. If risk-based models are highly detailed, it may not be possible to validate the results qualitatively based on observing events at a more abstract level. For example, observing a controller separating traffic would not inform the validity of a detailed model of software failure contributing to false conflict alerts.

There are several potential methods for qualitative validation. Auditing is one process by which independent experts review and potentially revise a model before giving approval. Simulations and demonstrations can also provide qualitative demonstrations. Real-time part-task or full-task simulations of a relevant operation or system can provide validation of event sequences, structure, and completeness. However, without a long period of observation or multiple trials, statistically significant error rates or failure rates are not likely to be achieved when probabilities are low.

Quantitative validation is the process by which the likelihood of events in a risk model is compared to relevant quantitative data to assess the model's validity. Quantitative validation can be performed against experimental, simulated, or historical data; data used for proper validation should be independent of the data used to build and calibrate the model. For this process, there is still a qualitative aspect of determining if the experience from a similar situation transfers to the defined event in the risk-based model.

One challenge in validation relates to influence models. Where influence models are used, they typically reflect the effect of policies, management practices, or an organization's safety culture. For these aspects of models, effects are confounded by multiple factors—such as varying influences at different management levels—making validation through observation problematic.

When validation efforts reveal that a model does not match reality exactly, as almost by definition, it then becomes important for model designers and end-users to understand any potential limitations. Particularly with regards to safety modeling, simplifying assumptions may mask the potential for unsafe scenarios. Therefore it is vital to understand how limitations in modeling may affect any safety assessments.

Verification is closely related to validation, but they have different goals. Despite a similar emphasis on quality assurance, validation ensures that a model corresponds to reality (correct design), while verification ensures that it produces results consistent with its design requirements (correct implementation) [8]. In other words, validation ensures designers have "built the right model" while verification assures they have "built the model right." A model may satisfy one or both standards. For example, a model may be verified to produce the designer's intended results, yet these results may not match reality—for example, due to incomplete design requirements. On the other hand, a model may produce results its designer did not intend—for example, due to incorrect implementation in software or misinterpreted requirements. Evidence from space systems suggests that more safety-related software errors stem from requirements issues than from implementation issues [1].

Assessment

All safety assessment tools have their limitations. Failure to take these limitations into account can lead to inaccurate results. These limitations must be clearly understood and accounted for in any assessment plan; hence, model limitations will be discussed.

2.2 REVIEW OF MODELS

This report focuses on models with comprehensive coverage of the National Airspace System (NAS) on most or all of the dimensions in Section 2.1. Thus, simpler models covering only one flight phase or one type of maneuver are not reviewed here because they cannot individually provide a NAS-wide NextGen risk picture.

Each model is assessed using the analysis factors and validation factors discussed above. A table format of each model's description is applied in this report; it is similar to the template used in the EUROCONTROL survey of risk modeling techniques [3]. Some models may have more recent versions than those reviewed; the date of the version reviewed is listed for each model.

U.S. sources for models reviewed include FAA organizations and NASA. Other models are the result of international collaborations between FAA and EUROCONTROL or purely international efforts. Several models are collaborations across government, academia, and industry. (See Table 1.)

In addition to air traffic management (ATM), safety assessment is important in domains where safety is a concern, including healthcare, pharmaceuticals, energy and others. Best practices from other domains may be leveraged to improve the efficiency of modeling efforts and resulting risk picture. Some reviewed models have elements from non-aviation domains; one example is CATS, whose human performance modeling approach comes from the nuclear power industry [6]. Another example is STAMP/STPA, which has been applied to a variety of non-aviation safety questions [1,9].

Table 1

Sources for Models Reviewed

	Source			
Model	**FAA**	**Other US Agency**	**International ANSPs**	**Other Domains**
CATS			X	X
IRP	X		X	
ISAM	X		X	
STAMP/ STPA		X		X
ASRM	X	X		
TOPAZ			X	

Model 1: Causal Model for Air Transport Safety (CATS) [6]
Source: Netherlands Ministry of Transport, Public Works and Water Management (VenW); Netherlands National Aerospace Laboratory (NLR); Delft University of Technology (TU Delft); et al.
Date of Version Reviewed: 2 March 2009
Contact: Ben Ale, TU Delft (b.j.m.ale@tudelft.nl)
Operating System: Windows
Application: Uninet and suite of related tools from TU Delft Department of Mathematics

Analysis Factors

Scope of systems	CATS is designed to model all functions of an air transportation system.
Range of events	CATS covers all flight phases, which are categorized into three groups: taxi/take-off/climb, en route, and approach/landing/taxi; it also covers 33 accident classes grouped by cause. These accident classes include human error and technical failures. Accident consequences, including third-party risk, are also modeled.
Approach and Methods	Causal models are developed as event sequence diagrams (ESDs) and fault trees (FTs) for each accident category and flight phase combination. FTs describe key events in the ESDs; all ESDs and FTs are integrated in a single Bayesian Belief Network (BBN). CATS models performance of human operators, including flight crew, controllers, and maintenance; it also models management influence on safeguards and mitigations. These influences are customizable to different environmental factors such as weather, time of day, and delays.
Data sources and data quality	Data sources include ICAO Accident Data Reporting System (ADREP), Line Operations Safety Audit (LOSA), Airclaims, Aviation Safety Network, Flight Safety Foundation, and individual accident and incident reports. Six human performance subject-matter experts are polled, including one maintenance expert. Most data sources cover years 1990–2003, but Airclaims data cover 1985–2005.
Outputs	Outputs include accident probabilities by severity, accident type, and cause, including third-party accident risk.
Feasibility for NextGen	"CATS enables quantitative risk assessments of existing and new operations" [6]. The BBN-based structure can model interactions between different accident causes and influences. Implementation is complete and includes tools with a graphical user interface (GUI). The GUI can be configured with one of two views: one for the general user to perform risk modeling and one for the expert user to provide input on influences. Further development is ongoing.

Validation

The CATS development team has completed independent internal peer reviews and submitted several papers for external peer review; full validation is still pending due to the challenge of obtaining independent data. Some case studies are complete. Validation seems limited due to the challenge of obtaining independent data; most available data have been used for model construction and calibration.

Assessment

CATS is among the more well-developed and well-documented models reviewed. One limitation is incomplete validation, though it compares favorably to other models on this factor—none of the models reviewed has complete quantitative validation. Because CATS is not explicitly designed to model NextGen OIs, additional development may be required. Some Event Sequence Diagrams (ESDs) and Fault Trees from CATS are being adapted for use by the Integrated Safety Assessment Model (ISAM) to model changes in safety risks related to NextGen OIs.

Model 2: Integrated Risk Picture (IRP) – Accident-Incident Model (AIM) [10,11,12,13,14]
Source: European Organization for the Safety of Air Navigation (EUROCONTROL, EUROCONTROL Experimental Centre – Safety Analysis & Scientific); Federal Aviation Administration (FAA)
Date of Version Reviewed: March 2010 for IRP and June 2012 for AIM
Contacts: Eric Perrin, EUROCONTROL Experimental Centre (eric.perrin@eurocontrol.int) and Andrew Kilner, EUROCONTROL Experimental Centre (andrew.kilner@eurocontrol.int)
Operating System: Windows/MacOS
Application: For IRP: Microsoft Excel, with add-in @Risk and Visual Basic for Applications (VBA); for AIM: Isograph Fault Tree+ and Microsoft Access

Analysis Factors

Scope of systems	IRP/AIM covers the major functions of an air transportation system. IRP has been designed to model and demonstrate the safety of ATM in both the present and future.
Range of events	IRP/AIM covers all flight phases (taxi, take-off, climb, en route, approach, landing, taxi) while considering five distinct accident classes (mid-air collision, runway collision, taxiway collision, CFIT, and wake turbulence accidents). Mid-air collision modeling is specialized to different environments. Accident causes considered include direct causes (e.g., acts of commission, equipment failures), prevention failures (e.g., acts of omission), prevention opportunities (e.g., where enhanced performance of ATC might have prevented an accident), and indirect influences (e.g., poor controller performance influencing pilot errors). Accidents independent of future advances in ATM are not considered.
Approach and Methods	Causal models are developed as event sequence diagrams (ESDs) and fault trees for each accident category. Fault trees describe key events in ESDs. All ESDs and fault trees are integrated with an influence model to control probabilities of base events. The influence model represents factors that cannot be expressed quantitatively; examples include human operator and management performance, equipment, airspace complexity, and airport layouts. Where possible, the influence model is quantified from data and reports rather than expert opinion.

	The layers of protection provided by ATM are represented in the IRP/AIM as sequences of barriers. The barriers operate in a rough time sequence. Each barrier removes a fraction of risk. Events between barriers increase in severity from routine exposure to accident. Fault trees model each barrier's causes of failure.
Data sources and data quality	For the baseline IRP/AIM modeling current ATM, historical data has been used where possible. Uncertainty is addressed with confidence limits obtained from statistical models. In predictive mode, IRP/AIM uses SME judgment in addition to data. IRP uses three reports covering aircraft incidents/accidents; the publishers and associated data sets are a Boeing report of Western commercial jet accidents (1990–2005); a United Kingdom Civil Aviation Authority report (1990–2005); and a Stiftelsen Det Norske Veritas (DNV) report (1998–2004). A total of 137 accidents and incidents contribute to the construction and quantification of the model.
Outputs	Outputs include fatal and nonfatal accident and incident frequencies, and the associated causal contributions, importance factors, performance of the ATM layer, and risk curves that show how the risks will evolve as changes to the ATM System are implemented. Monte Carlo simulation helps to quantify uncertainty.
Feasibility for NextGen	IRP/AIM enables quantitative risk assessments of existing and new operations. Output allows determination of sensitive parameters or operational improvements that increase safety. FAA and EUROCONTROL currently cooperate on risk modeling and model development. IRP/AIM help to predict how risks will evolve as ATM system changes are implemented—a capability needed for modeling NextGen concepts. The model allows adjustments to the changes' effectiveness and implementation sequence to help ensure safety targets are met throughout the transition process.

Validation

IRP model results have been validated to the extent possible through several methods: (1) Calibration to ATM changes that occurred between 1990 and 2005; (2) empirical validation against independent estimates; (3) convergent validity against available statistics; and (4) face validity through stakeholder review. Ten European air navigation service providers (ANSPs) reviewed the qualitative model structure. The IRP development team has published a number of externally peer-reviewed papers and reports.

Assessment

Some elements of IRP are adopted by the Integrated Safety Assessment Model (ISAM), which uses them to model changes in safety risks due to NextGen OIs. For the purpose of comparison, unlike the CATS model, the influence modeling in the 2006 EUROCONTROL IRP reports does not use Bayesian Belief Networks to propagate probability distributions [10,11]. Instead the influence model acts on base events by directly scaling their likelihood of occurrence. Later development for use in the U.S. was begun, with the intent of replacing the influence model with BBNs [15].

Model 3: Integrated Safety Assessment Model (ISAM) [16]
Source: Federal Aviation Administration (FAA) – SMS Safety Risk Management Group; Saab Sensis Corporation
Date of Version Reviewed: 7 December 2011
Contact: Stojan Trajkov, Saab Sensis Corporation
Operating System: No specific requirement
Application: Web browser, Adobe Flash

Analysis Factors

Scope of systems	ISAM models all functions of the NAS affected by NextGen OIs.
Range of events	ISAM covers all flight phases: taxi, take-off, climb, en route, approach, and landing. It also covers 6 accident categories: wake, CFIT, taxiway collision, runway collision, and midair collision. ISAM groups accident causes into the same categories as IRP: direct causes, prevention failures, prevention opportunities, and indirect influences.
Approach and Methods	ISAM combines elements of two earlier models—CATS and IRP—along with additional elements. Its influence model modifies baseline fault trees depending on which combination of NextGen OIs is implemented. ISAM assesses changes in the risk picture due to OI implementation and can model partial or full implementation of individual OIs. An example of partial OI implementation might be partial fleet capability for Flight Deck-based Interval Management (FIM) or differing versions of ATC automation at different ATC facilities. ISAM can also model several types of constraints on OI hierarchical dependencies.
Data sources and data quality	Refer to data sources for CATS and EUROCONTROL IRP, which include expert judgment of influences. Inputs also include assumptions about the NextGen implementation schedule, OI hierarchical dependencies, and future traffic volumes.
Outputs	ISAM predicts accident rates by category before, during and after the NextGen implementation timeline; results are compared to a user-specified target level of safety. Within each accident category, ISAM determines a breakdown of accident risk by cause—specifically, the contribution from each type of pilot task, ATC task, and base event or root cause.
Feasibility for NextGen	ISAM is explicitly designed to model incremental changes in safety risks due to NextGen OI implementation. Model implementation is in progress and a GUI-based tool is under development. The GUI can be configured with one of two views: one for the general user to perform risk modeling and one for the expert user to provide input on influences.

Validation

Calibration of the model to data from the NAS is pending.

Assessment

ISAM is still under development but appears more directly relevant to NextGen safety modeling than other models reviewed; it may require fewer updates and less additional development effort to produce the desired NextGen risk picture. However, it may not yet cover all accident categories—for example, loss of control does not appear to be covered.

Model 4: System-Theoretic Accident Model and Processes (STAMP), System-Theoretic Process Analysis (STPA) [1,9]
Source: National Aeronautics & Space Administration (NASA); National Science Foundation (NSF); Massachusetts Institute of Technology (MIT) – Department of Aeronautics and Astronautics/Complex Systems Research Laboratory (CSRL); Safeware Corporation
Date of Version Reviewed: 13 January 2012
Contact: Nancy Leveson, MIT (leveson@mit.edu)
Operating System: No specific requirement
Application: Optional SpecTRM (Safeware Corporation) assists with implementation in Windows

Analysis Factors

Scope of systems	STAMP can be used to model socio-technical systems that enforce safety constraints.
Range of events	STAMP considers all factors leading to an accident including immediate prior events and influences across different levels of a system's hierarchical control and feedback structure. A typical structure includes legislators, regulators, company or agency management, supervisors, and system operators—both human and machine.
Approach and Methods	STAMP considers risks during both development and operation of a system; it models the system's hierarchical control structure, including organizational and cultural factors.
Data sources and data quality	Sources include accident and incident reports or statistics along with details of hierarchical control and feedback structures.
Outputs	Outputs include identification of causal mechanisms, unsafe system interactions, and unsafe cultural or organizational influences. This includes hazards such as design shortcomings, flawed requirements, software flaws, and operational errors.
Feasibility for NextGen	NextGen is a "system of systems," so a systems approach is appropriate. A software tool suite to support the approach is in development. STAMP/STPA has previously been applied to aviation-related research problems including predictive hazard identification, forensic accident investigation, and software error analysis.

Validation

STAMP-based analysis of accidents in aerospace, healthcare and other domains has been published and peer reviewed.

Assessment

STAMP is a general approach to RBM rather than a model itself; it has not yet been applied to a complete NextGen model—only to individual concepts. Additional development effort would be required to realize the approach in a complete model. Unlike the other reviewed models, it is less mathematically focused; this difference may be an advantage for modeling NextGen OIs, whose behavior and interactions are not yet fully described. STAMP's focus on identifying unanticipated causal mechanisms and interactions may make it more appropriate for NextGen than other modeling approaches. STPA is a hazard analysis and identification approach which can support STAMP or other RBM methods. It is intended to provide guidance to system designers and operators at any point in a system's lifecycle—even before a design is available—so that safety can be designed into the system. It is also intended to identify hazards such as design shortcomings, software flaws, unsafe interactions, and other hazards which may be missed by other hazard identification methods [1]. Section 3.1 includes an example of applying STPA to predictive hazard identification [9].

Model 5: Aviation Safety Risk Model (ASRM) [17,18]
Source: Federal Aviation Administration (FAA); National Aeronautics & Space Administration (NASA); Rutgers University – Department of Industrial & Systems Engineering
Date of Version Reviewed: 26 January 2006
Contact: James T. Luxhøj, Rutgers University (jluxhoj@rci.rutgers.edu)
Operating System: No specific requirement
Application: Hugin BBN tool [17]

Analysis Factors

Scope of systems	ASRM models functions of an air transportation system relevant to certain accident types and safety technologies.
Range of events	ASRM has been implemented for six accident types: CFIT, loss of control, runway incursion, engine failure, maintenance, and general aviation.
Approach and Methods	ASRM uses Bayesian Belief networks (BBNs) to model probabilities of different causal relationships; it also models interactions between accident causes. Human operator performance and management performance are modeled using the Human Factors Analysis and Classification System (HFACS). Operational and equipment-related factors are also modeled. Expert judgment is used to estimate the impact of system changes on causal factors in the BBNs. "The BBN modeling approach enables an assessment of single or multiple technologies impacting either single or multiple causal factors" [17].

Data sources and data quality	Accident and incident reports are used to identify relationships between accidents and their causes. To cover cases without available data, 30 subject matter experts are polled.
Outputs	ASRM provides accident causes, predicted changes in risk due to system changes, and prioritization of mitigation strategies.
Feasibility for NextGen	ASRM "assesses the impact of new technology insertions or products designed to mitigate the likelihood or consequence of aviation accidents" [17] and hence could be used to model changes in safety risks due to NextGen OI implementation. ASRM has previously served as a decision support tool to evaluate system changes & new technologies, including NASA's Aviation Safety and Security Program (AvSSP).

Validation

The ASRM development team has published their approach and results of case studies for peer review [17].

Assessment

Because ASRM is not explicitly designed to model NextGen concepts, some additional development to cover them would be required. ASRM may not yet cover all accident categories—for example, wake vortex encounters do not appear to be covered as of 2006; however, more recent model versions seek to model additional accident types.

Recent model updates include additional, improved BBN types to increase model fidelity and flexibility. The model has also been applied to additional systems and accident types such as unmanned aerial systems and wake vortex encounters. These updates are described in more recent (2012) publications from Luxhøj and Sarlo [19] and Luxhøj and Topuz [20] which are not yet included in this review.

Model 6: Traffic Organization and Perturbation Analyzer (TOPAZ) [5,21,22,23]
Source: Netherlands National Aerospace Laboratory (NLR) – Air Transport Division
Date of Version Reviewed: 2001
Contact: H. A. P. Blom, NLR (blom@nlr.nl)
Operating System: No specific requirement
Application: Specialized software

Analysis Factors

Scope of systems	TOPAZ models functions of an air transportation system relevant to certain accident types and safety technologies.
Range of events	TOPAZ models collision risk for various phases of flight and several accident categories, including terrain impact and wake vortex encounters.
Approach and Methods	TOPAZ includes analytic and fast-time Monte Carlo simulation elements. It models human operator performance and the likelihood of rare anomalies. It is implemented as an agent based model using stochastic differential equations (SDE) and Petri networks.
Data sources and data quality	Required inputs include data to build the mathematical models, distributions of human operator response times (conditioned on environment and workload), and preliminary hazard analysis. Also, the model requires a statistical description of the ATM situations and environments to be analyzed; this description might include, for example, distributions of parameters describing traffic flows, weather conditions, and avionics equipage rates.
Outputs	TOPAZ outputs frequencies of non-nominal events and conditional collision probabilities for various event sequences—both before and after system changes.
Feasibility for NextGen	TOPAZ has been applied in a variety of ATM modeling efforts, including safety assessment of satellite-based communication/surveillance, simultaneous converging approaches, and data communications. The development team writes that "with the help of TOPAZ, it is, in principle, possible to evaluate the safety characteristics of any new operational ATM concept under consideration" [5].

Validation

Software implementation of TOPAZ has been verified for correctness using the model's mathematical properties. The TOPAZ development team has published the approach, methods, and results for peer review [21]. Quantitative results from the model have been verified consistent with qualitative hazard assessments through discussion with subject-matter experts.

Assessment

The TOPAZ architecture appears more mathematically sophisticated than others reviewed and hence may require greater depth of expertise from developers and users. TOPAZ does not currently model all or the majority of NextGen OIs, so additional development to cover them would be required. The major benefits of TOPAZ are derived from the agent-based modeling approach. Agent based modeling

enables the discovery of emergent behaviors. And furthermore, as agent-based models improve, their use can be extended to later stages of concept development. TOPAZ provides a strong modeling framework at both early and later stages of concept development. And unlike STAMP/STPA, which is another approach well suited to early states, TOPAZ is able to provide numerical results which can be validated against known data sources while considering complex interactions, much like the influence models in CATS or ISAM.

2.3 SYNTHESIS

The reviewed models seek to build a risk picture of the NAS using a variety of approaches. Model developers recognize that composite modeling techniques and multiple data sources are needed for a complete risk picture. No model reviewed relies completely on a single method or data source. In order to understand the models holistically and to set the stage for future recommendations, it is useful to provide a comparative analysis of the models based on the dimensions identified earlier: Model and Data descriptions, Applicability to NextGen, and Validation.

2.3.1 Model and Data Descriptions

All reviewed models cover different accident types and causes, but do so using different categorization approaches which are not necessarily comparable. Furthermore, even the very definition of an accident varies. For example, categories may be based on the type of accident (e.g., CFIT or mid-air collision) or based on events leading to the accident (e.g., single engine failure, flight crew member spatially disorientated). Human operator roles may be grouped into categories, such as direct causes or prevention failures, or more specific categories, such as flight crew decision error or spatial disorientation. Some models use multiple methods to define categories. For example, IRP, ISAM, ASRM, and CATS all use accident categories defined partly by physical mechanism and partly by operator role; however, CATS categories are defined using more specific operator roles. When accident categories and even accidents themselves are defined in different ways across models, direct comparison is not meaningful.

Considering dependencies between different accident types and causes can be more challenging than assuming independence between accidents. However, it can also yield more complete results– including more realistic modeling of common causes for different accident types. For example, CATS approximately models interactions between accident causes in a BBN structure. However, due to limited source data, it does not include a full probabilistic model with details of all possible interactions. For example, there exist some cases where it does not model the possibility of an aircraft entering an event sequence in a degraded state following an earlier incident during a previous phase of flight [6]; often this is a function of limited data on such types of event. ASRM also uses a BBN-based approach to model interactions between accident types and causes.

In addition to predicting accident rates and likelihoods, a risk model may also predict the distribution of severity or consequences for each accident type; these may be grouped into categories such

as none, minimal, minor, major, and catastrophic. CATS and ISAM are designed to predict both consequences and likelihood of accidents. Also, IRP is capable of outputting the frequency of fatal accidents, ICAO-defined accident frequencies, and precursor incident frequencies.

Similar to accident categories, flight phases are grouped differently by different models. However, the reviewed models use frameworks general enough to provide comprehensive coverage of all flight phases.

Models reviewed in this report were selected for their wide coverage of a variety of ATM-related systems; hence, they have similar scope of systems coverage. Some models, such as TOPAZ, IRP, ISAM, and ASRM, have been applied to specific accident types and as such their coverage is focused on systems relevant to these accident types. However, the structures behind each of the models are general enough that they can be extended to cover additional systems as needed. For example, even though CATS does not consider NextGen components or potential future concepts of operation—such as closely spaced parallel operations in instrument conditions or 3 NM en route separation—it does not prevent their inclusion in the model. These concepts could be modeled by applying the same model construction methods to the new concepts.

It is worthwhile to take note of methods for constructing the models. The designers of IRP use the Structured Analysis and Design Technique (SADT) to develop the ESDs, FTs, and influence models of the complete model. SADT provides a fixed procedure to relate resources and tasks to inputs, outputs, and constraints. A general representation of a SADT task is shown in Figure 4, where each task can be identified through a complete ATM system model. Starting from a EUROCONTROL Operational Concept document, an example en route SADT model is shown in Figure 5. One benefit of SADT is that interdependencies between tasks and resources are easily identified. Furthermore, a structured procedure for identifying model elements implies that the complete risk model is less likely to miss valuable elements unless there is a fundamental shortcoming to the procedure. In short, creation of IRP relies on a consistent and reproducible procedure which should minimize the number of overlooked hazards. Like IRP, STAMP/STPA also provides a framework for systematically identifying hazards and operator actions causing them [1,9].

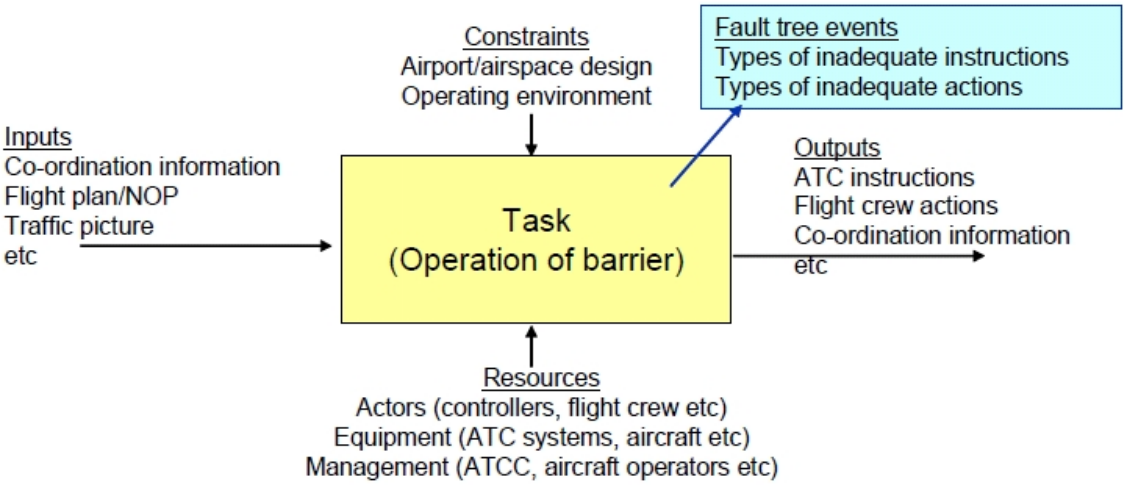

Figure 4. Generic SADT Task [10]

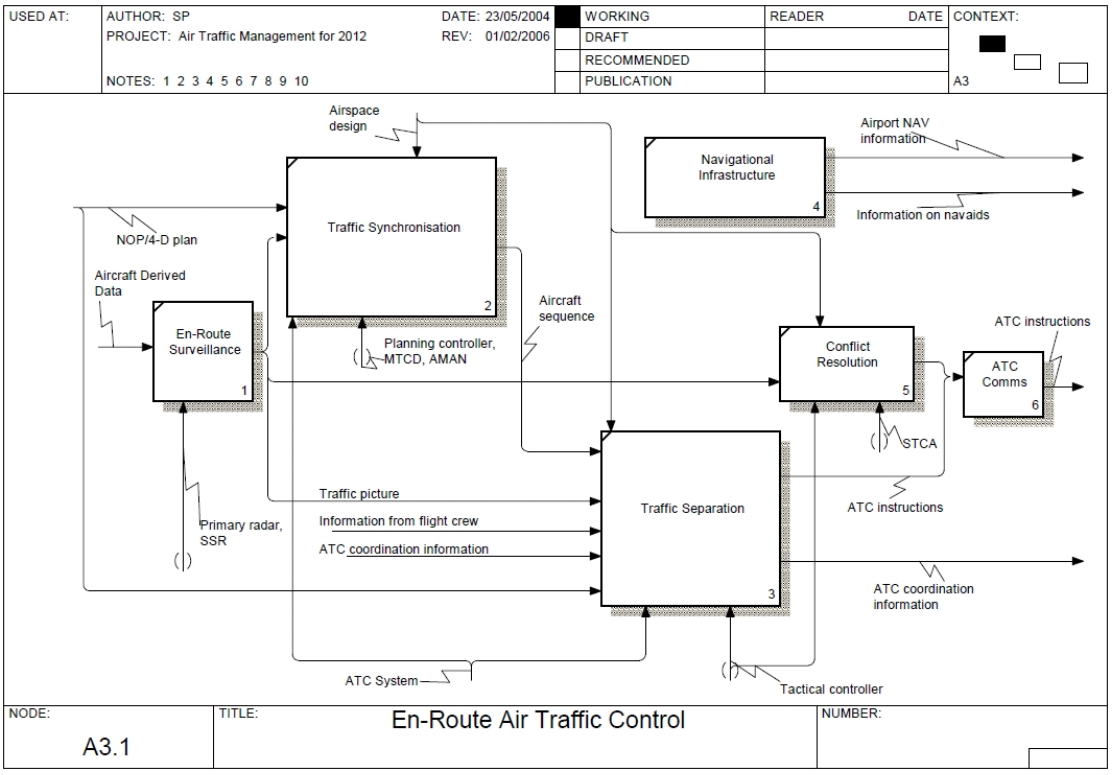

Figure 5. En Route SADT Model [10]

IRP, CATS, and ISAM share a model structure based on ESDs, FTs, and an influence model. It is within the influence model that many NextGen (or SESAR) operational improvements are considered. Among the influence-based models reviewed, there are two major approaches for representing influences: a BBN-based method (found in CATS) and a direct influence method (found in IRP). The influence model in IRP scales the probabilities of root events or initial hazards in the ESD. Scaling factors are established by estimating the expected benefit or harm of an influence. In this case, each influence is measured along a performance scale from poor to perfect. Perfect performance implies that the frequency of failure events cannot be decreased by improvements to any influencing factors (e.g., improved pilot training would not reduce the number of conflicts). In contrast, BBNs can be considered more sophisticated because they can represent complex relationships and propagate influences among interrelated conditional probability distributions. One benefit of influence models is that common cause failures can be naturally included.

The influence model in IRP does not rely on a BBN like CATS and other related models. Instead, the frequency of each bottom-event (i.e., root event in the ESD, or more colloquially the initial event leading to an accident) is scaled according to the influences. Unlike in CATS, probability distributions are not propagated. Despite the potential benefits of using BBNs to represent more complex relationships at the influence layer, there are some drawbacks. An analysis performed by EUROCONTROL concluded that development of BBNs requires more data for quantification, and they may expose a model's end-users to unnecessary mathematical complexity [12]. Furthermore, risk models which are text-based or require configuration files may overwhelm some users—particularly policy makers or those with very specific interests; these users may lack the understanding to properly interpret parameter settings. Recommendations for controlling a model's level of complexity are included in Section 5.1.

Input sources also vary across models. Figure 6 groups models reviewed by relative reliance on expert opinion and numerical data. At one end is STAMP, which is mostly driven by expert assessment of relationships in a system's hierarchical control structure. Model developers identify these relationships through interviews with system operators and management and review of system designs and accident reports. At the other end are TOPAZ and IRP. TOPAZ is heavily driven by statistical data and mathematical structures describing the system; like IRP, it relies less on expert opinion. However, IRP is not as data-driven as TOPAZ. Despite these observations, none of the models relies completely on one type of input. For example, a STAMP approach could consider numerical data describing system behavior, while TOPAZ requires input from ATM experts to identify hazards, design the model's mathematical structures, and validate results. The other models fall in the middle with somewhat equal reliance on expert opinion and numerical data.

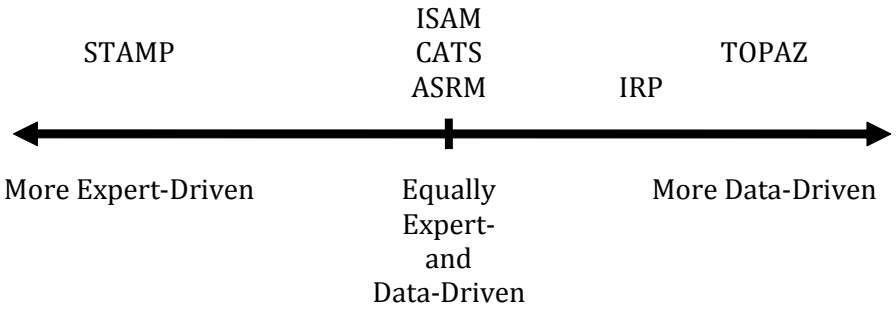

Figure 6. Relative Importance of Quantitative and Expert Data Sources by Model

2.3.2 Applicability to NextGen

It is also useful to assess each model on its applicability to NextGen. For example, ISAM appears more directly relevant to NextGen risk modeling than other models reviewed since it explicitly seeks to model the influences of each OI on the NAS risk picture. However, other models are also capable of modeling OI influences, but they will require further development to achieve this goal.

Model status and implementation will influence the choice of appropriate models for NextGen risk modeling. Table 2 summarizes these factors, including development of model architecture, calibration to data from the NAS, and implementation in software. This table is only a rough guide because model development is typically an ongoing process. For example, development of some models, including CATS and ISAM, continues even after initial results and demonstration versions. As discussed above, some models, including ASRM and TOPAZ, are only implemented for certain accident types and do not yet comprehensively cover the NAS and NextGen OIs. In the "Implementation" section of Table 2, the "Software" column indicates whether a software implementation of the model has been documented and the "GUI" column indicates whether the implementation includes a graphical user interface. No entries for STAMP are shown because it constitutes an approach or procedure for analyzing and understanding safety; it is not intended to serve as a quantitative risk-based model.

Table 2

Model Status and Implementation Details

	Status		Implementation	
Model	**Initial Development Complete**	**Calibration Complete**	**Software**	**GUI**
CATS	X	X	X	X
IRP	X		X	
ISAM	X		X	X
STAMP/ STPA				
ASRM	X	X	X	X
TOPAZ	X	X	X	

Model developers recognize the advantage of reusing existing model elements and data sources; for example, ISAM's developers are leveraging elements from CATS and IRP rather than recreating them independently. Modularity of model structures makes their development and reuse easier; most models reviewed have such a structure.

2.3.3 Validation

Reviewed models have been validated using a variety of methods. Table 3 summarizes the methods applied to each. In addition to the general methods tabulated here, developers have implemented a variety of more specific methods within each category; these methods are described in the references. Among models reviewed, CATS has the most thoroughly documented validation methods [6]. No reviewed models have documented complete quantitative validation against independent datasets; however, developers of CATS, ASRM, and TOPAZ have verified internal consistency of mathematical model structures. Again, quantitative validation of STAMP is not applicable because the approach has not yet been implemented in a model for the ATM domain, nor is STAMP intended to provide numerical results; rather, it serves as a hazard analysis tool.

It is important to highlight one particular caveat: NextGen risk-based models cannot be completely validated until NextGen and its components are deployed to the NAS. Until then, only partial validations are possible through HITL testing, which in many cases only considers individual elements.

Table 3

Validation Methods Applied to Reviewed Models

	Qualitative		Quantitative	
Model	**Expert Audit**	**Peer-Reviewed Publications**	**Internal Consistency**	**External Independent Dataset**
CATS	X	X	X	
IRP		X		
ISAM	X			
STAMP/ STPA		X		
ASRM	X	X	X	
TOPAZ	X	X	X	

This page intentionally left blank.

3. GENERAL RECOMMENDATIONS

This section begins the recommendations portion of the report. It discusses general recommendations and guidelines for NextGen concept validation and risk modeling. It covers best practices such as (1) conservative and credible worst-case assumptions; (2) consideration of rare or "black swan" events, including surveillance outages and degraded ATC capabilities; (3) understanding uncertainty related to the NextGen concepts under assessment, and (4) investigation of emergent behaviors in the NAS—an increasingly complex sociotechnical system.

Sections 4 through 6 cover additional, more specific recommendations. Finally, Table 6 summarizes all recommendations from this section and Sections 4 through 6.

In the early stages of design, limiting the scope and level of detail of risk modeling will help to address the challenges of limited data and system descriptions. These challenges may initially prevent full quantitative results and quantitative validation; however, such detailed results are not necessarily helpful in the early design stages of NextGen elements. Instead, NextGen risk modeling efforts could focus exclusively on coarse or high-level results; such results could include rough order-of-magnitude accident risk estimates, identification of previously unknown hazards, or prioritization of known hazards. These results would provide feedback to system designers, who could then modify NextGen system designs or concept of operations (ConOps) to mitigate or prevent any newly-identified hazards. In particular, modeling and simulation should be used proactively to determine whether new requirements are needed to ensure safety of NextGen concepts. For example, analysts studying the safety of potential future ADS-B-only, non-radar operations may find that aircraft will require a backup navigation method apart from GPS. Backup navigation would prevent aircraft navigating with GPS from losing navigation altogether during a GPS outage—a doubly hazardous situation due to simultaneous loss of GPS-dependent ADS-B surveillance. Mandating backup navigation or other new requirements may be a necessary preventive measure to ensure the safety of future ADS-B-only operations. In general, the results of safety studies during NextGen concept development should be used to ensure that NextGen concepts are designed with safety in mind.

NextGen risk modeling should seek to identify each concept's hazards as early as possible so they can be eliminated, reduced, or mitigated, hence increasing safety. Some of these hazards may be initially unknown due to the new types of interactions enabled by the concept. Ideally, risk modeling results should be used not only to describe a system's level of safety, but also to identify opportunities for safety improvements. Seeking these opportunities helps risk modeling avoid the limitations of "cosmetic system safety" and "compliance-only exercises," which make safety arguments but do not actually help to improve safety.

In later stages, as detailed system descriptions and results of test and evaluation become available, this greater information availability will allow removal of simplifying and conservative assumptions and ultimately improve risk model fidelity. Eventually, more detailed Safety Management System (SMS) processes could be applied to move each concept toward approval and implementation in the NAS.

3.1 CONSERVATISM AND HAZARD IDENTIFICATION

In general, NextGen risk modeling must strive to avoid false negative results—that is, declaring a concept safe when in reality it is not. Such results can occur when models overlook hazards or underestimate risk. The risk modeling methods adopted for NextGen must include protections against overlooking hazards, such as systematic hazard identification procedures, expert feedback, and peer review; they must also include appropriate assumptions or other methods to prevent underestimation of risk. Conservative assumptions are particularly important when modeling concepts under development because their behavior is not fully defined. Assuming the worst credible outcome in all situations is a best practice, especially for initial lower-fidelity modeling efforts; it ensures that potential risks are properly accounted for even if they are not explicitly modeled [24]. These assumptions must be clearly documented for users and other stakeholders. For example, model users must be informed what NAS environment is assumed and whether a model focuses only on certain data sources, accident types or hazards.

Regardless of the modeling approach adopted, an important priority should be to identify any new hazards introduced by NextGen concepts; one example hazard is simultaneous loss of surveillance and navigation due to GPS outage in an ADS-B-only environment. Hazard identification benefits from systematic approaches such as Hazard and Operability Study (HAZOP) and STPA, which facilitate structured brainstorming [1,3]; either of these techniques would complement any of the risk models reviewed in Section 2.2. Leveraging a variety of hazard assessment techniques and including opinions from a variety of independent experts will also help to minimize the possibility that important hazards are overlooked.

The JPDO Capability Safety Assessment (CapSA) for TBO provides an example of the hazard identification and mitigation process; it adopts consensus definitions and assumptions about future TBO capabilities and uses SME judgment to assess hazards and their mitigations [25]. SMEs in this study prioritize hazards using combined judgments of their significance, likelihood, and strength of potential mitigations; they judge hazards with well-defined, effective mitigations as less hazardous than their significance and likelihood would otherwise suggest. Risk modelers should apply a similar approach to other NextGen concepts to estimate which hazards are most important.

Reports from the STAMP/STPA development team at MIT provide a variety of hazard identification examples from different domains, including air traffic management, space and missile systems, robotics, water quality, food safety, and pharmaceuticals [1,9]. These examples show how the STPA method provides a systematic framework SMEs can use to identify the possible ways each operator control action could cause an unsafe situation. One example covers the In-Trail Procedure (ITP), which

permits aircraft to temporarily operate at reduced separation while changing flight levels en route; this capability improves airspace flexibility and efficiency by permitting more frequent flight level changes [9]. Table 4 lists a subset of potential hazards of ITP; they are identified for each operator control action—such as "execute ITP." The four categories (1) "not providing," (2) "providing," (3) "wrong timing/order," and (4) "stopped too soon/applied too long" summarize the possible ways each control action could produce a hazard; SMEs begin with the table blank and determine which combinations produce a hazard. Not all combinations are hazardous; for example, the combination "not providing" and "execute ITP" is left blank, indicating no hazard occurs when the flight crew chooses not to execute ITP and instead conforms to standard separation minima. This table-based method helps SMEs to exhaustively list the possible consequences of each operator control action and helps to prevent hazards from being overlooked.

STPA helps identify hazards and losses caused by complex system interactions—such as unexpected software behavior—in addition to physical component failures [1]. It has been demonstrated to be capable of identifying hazards missed by other methods, even when system concepts are not fully described. NextGen safety studies should leverage STPA or similar approaches as much as possible even if other methods are also used.

Table 4

Example of Potentially Hazardous Control Actions by the Flight Crew during In-Trail Procedures [9]

Control Action	Not Providing Causes Hazard	Providing Causes Hazard	Wrong Timing / Order Causes Hazard	Stopped Too Soon / Applied Too Long
Execute ITP		ITP executed when not approved ITP executed when ITP criteria are not satisfied ITP executed with incorrect climb rate, final altitude, etc	ITP executed too soon before approval ITP executed too late after reassessment	ITP aircraft levels off above requested FL ITP aircraft levels off below requested FL
Abnormal Termination of ITP	FC continues with maneuver in dangerous situation	FC aborts unnecessarily FC does not follow regional contingency procedures while aborting		

3.2 SYSTEMIC FAILURES

Reviewed models do not generally address how ATM systems are expected to respond during systemic failures affecting regions of the NAS. In a typical RBM approach, components are characterized by their failure rates, typically measured in expected failures per performance cycle, per aircraft, or per flight hour. In nominal operating conditions without widespread system failure, such an approach is reasonable—especially because system-wide failures, such as communication outages, are rare, and so the likelihood of multiple simultaneous or consecutive failures is almost nonexistent during nominal conditions.

However, because NextGen includes a number of system-wide technological changes to enable desired operational improvements, it is vital to ask: how does the risk picture change with the failure of an entire system or subsystem, as opposed to only a singular failure event? While systemic failures are not commonplace, their magnitude and the duration of their effects can be large and long-lasting. Examples include

1. A 2009 lightning strike of the Hartsfield-Jackson Atlanta International Airport control tower that forced evacuation, coupled with a power outage to the airport [26].

2. A 2007 communication failure at the Memphis Air Route Traffic Control Center [27].

3. An August 2010 failure of ADS-B reporting in the Gulf of Mexico when a central network hub failed without any back-up systems; FAA's ability to monitor ADS-B reporting through the Surveillance and Broadcast Services (SBS) Monitor was also compromised during the outage [28].

These examples describe severely degraded operating conditions. With the inclusion of new technologies and high-density traffic patterns, similar system failures in the future cannot be ruled out. There is a need to then address how the probability of accidents changes in these degraded environments, particularly when heavy traffic conditions are present just prior to the failure. Would a facility be able to revert to prior well-established traffic management techniques or alternative solutions, or would high traffic volumes prevent safe management of aircraft?

The risks of events affecting large, critical aspects of the NAS should be considered in NextGen risk modeling due to the variety of NAS-wide OIs planned.

3.3 UNCERTAINTY

When concepts are under development, their performance is uncertain because their behavior may be only partly defined or understood. Quantifying this uncertainty allows stakeholders to make better decisions on how to proceed with a concept's development. Such understanding may be gained through more precisely defining requirements, through modeling and simulation, or through operational evaluation.

Risk model developers and users should understand the uncertainty in a model's results; all parties should be knowledgeable of the relationships between uncertainty in a model's outputs and uncertainty in its inputs. For quantitative models such as BBNs, algorithms exist to determine such relationships [29]. If uncertainties are too large, model results may be inconclusive or meaningless. For example, if the behavior of a particular set of concepts is not defined precisely enough, there may not be enough evidence to show these concepts maintain, improve, or degrade the level of safety in the NAS.

Uncertainty relationships can be used in the model design process to identify key areas needing more focused attention. By designing the risk model to bracket parameters between upper and lower bounds, those with large impact or uncertainty can be identified. High-impact or uncertain parameters should be prioritized for greater scrutiny; further refinements may permit narrowing of their bounds or confirm their importance. In other words, if a parameter is uncertain, and adjusting it causes large changes in the risk picture, then additional research to ascertain its true value is needed. Hence the design process of an integrated risk model also drives future research in related areas.

Similar to testing upper and lower bounds, exploring the propagation of distributions is also useful. BBNs, like those contained in CATS, are useful for such explorations as they enable parameters and influences to be manifested not only as singular values, but also as probability distributions. When probabilistic distributions are not directly applicable to a risk model, a perturbation analysis can be performed.

3.4 EMERGENT BEHAVIORS

It is worth discussing the value of an integrated risk model for NextGen, with regards to validation efforts. First, the modeling process should be classified as either *consolidative modeling* or *exploratory modeling*. The goal of consolidative modeling is to codify laws, behaviors, parameters, etc., into a representative model that serves as an approximation, simplification, or surrogate for an actual system. Two key benefits of surrogate consolidative models include running simulation studies in a safe environment which real-world studies would preclude, and rapidly running simulations over a wide range of scenarios (e.g., Monte Carlo simulations) in a reasonable time-frame. Often, exploratory modeling is an application process used when system behaviors are not well understood or undefined. The associated uncertainties relating to the system behaviors then require a number of assumptions, approximations, and/or guesses to bridge knowledge gaps in the model. Thus, when making use of an exploratory model there should only be limited expectation that results accurately reflect the real-world system. Instead, the value of such a model is that the modeling process and experimentation can be used "to explore the implications of varying assumptions and hypotheses" [30].

An integrated risk model for NextGen essentially attempts to model system changes (i.e., operational improvements) and their effects on the NAS. The options for potential operational improvements are and will always be at various phases of development, from rough concept to near implementation. And yet, regardless of their stage of development, until actual implementation, there will never exist data to completely support complete model validation of future concepts within a NAS model.

A shortcoming of many modeling procedures appears when extending models based on current concepts of understanding. Traditionally, many models, like some listed in Section 2.2, make use of historical data to tune parameters and the overall model to reflect past reality. Use of the model is then extended future concepts under the guise that system relationships remain static as improvements are brought online. That is, the underlying model structure (influence model and/or fault trees) remains the same despite changes in operations. Certainly, such an approach in itself does not imply failure of a model but rather limits potential usage. Extensions of a model to system changes are possible when the said changes are within the range and scope of the model. However, one must be careful with radical changes, which often times result in structural changes that are not necessarily identifiable in advance. A commonly understood example of where extending a model falls flat can be found in physics. Newtonian mechanics are able to accurately model dynamics in nature from planetary motion to the path of a baseball, however, these models quickly fall apart at the atomic level (e.g., understanding friction or van der Waals forces)—hence the need for Quantum mechanics. Along these lines the models reviewed here may have shortcomings as exploratory models.

One often-proposed purpose of exploratory models is to identify and describe emergent behaviors, that is, more complicated behaviors that arise from simple interactions. In systems engineering emergence can be described as "weak emergence" or "strong emergence." Weak emergence describes new behaviors (sometimes unexpected or not explicitly designed for) that are generated as a result of dynamics at the elemental level. Typically, weak emergence is thought to be traceable. Conversely, strong emergence may be more difficult to trace, as it is related to complicated behaviors resulting from the interaction of sub-systems.

With regards to risk-based models like CATS, ISAM, and IRP, there can be no new identifiable emergent behaviors. As reviewed, this class of risk-based NextGen models maintains similar structures and static relationships as a model describing current operations. And because NextGen OIs manifest themselves through an influence model with static relationships (and limit changes to the output by scaling the same set of parameters), there is nothing uniquely different between a NextGen model and a model representing current operations. More so, even structural changes that are directly inserted into the models infer planning and expectation of certain behaviors. Thus, with the underlying model remaining the same despite NextGen improvements, there should be no expectation to expose emergent properties.

If the goal is to expose emergent properties, then the modeling structure must fundamentally change from the lowest levels. Instead of simply providing a metric or means for an operational improvement to influence risk, the functionality and laws governing of the behaviors of the concept must be directly incorporated into the model. In this sense, the TOPAZ model is an exception, as individual agent dynamics are described, as well as their interactions. So in many ways, TOPAZ is able to expose both weak and strong emergent behaviors.

To date, airspace service providers, researchers, and engineers have systemically worked to remove known hazards in air transportation. So much so, that is it rare that two catastrophic accidents are the

result of the same sequence of events. And while human error is often listed as a root cause, causal factors that lead to such an error vary significantly.

With NextGen whole new sets of safety hazards will arise. However, when creating safety models based on current operations or our present-day view of future operations, the result is likely a model that has been structured to fit within a framework of well-understood behaviors and potential hazards. Likewise, there is always the concern that the introduction of concepts or operational improvements might lead to changes in agent (e.g., pilots, air traffic controllers) behaviors that are not predicted in advance. Future accidents that are emergent properties of NextGen concepts might be unforeseeable from straightforward examination of system diagrams. Future accidents related to NextGen will be considered "black swans" – once thought to be impossible, yet easily explained after the fact [31]. One best practice to prevent the appearance of "unforeseen" accidents is to build detailed exploratory models that test and adjust the functional behavior each NextGen concept—not only in stand-alone operation, but also in conjunction with other operational improvements. And in these cases the exploratory model should not be expected to calculate risk probabilities, but rather identify potential hazards. Therefore, as part of any safety analysis that seeks to expose emergent behaviors, an agent-based approach similar to TOPAZ is recommended. HITL testing might also be able to expose how agents use and interact with new technologies and concepts when they are provided with leeway.

This page intentionally left blank.

4. RECOMMENDATIONS FOR CONCEPT VALIDATION

This section provides recommendations related to concept validation—the process of assessing ideas for improvements in the NAS to ensure they have a net positive impact on the NAS and that they enhance or maintain safety.

The section discusses the need to ensure safety is properly included in wider research and development efforts, business processes, and organizational culture—using the FAA I2I process as an example. It also discusses a FAA concept development framework which defines Concept Maturity Levels (CMLs). Both of these frameworks are useful for planning safety efforts and ensuring they have appropriate scope and focus.

There is some overlap between concept validation and risk model validation, but key differences limit the applicability of one to the other. Concept validation seeks to answer questions about a concept's operation and its expected benefits and costs. More specific recommendations for risk model validation are later covered in Section 5.

4.1 INTEGRATING SAFETY AND RISK MODELING INTO STAKEHOLDER BUSINESS PROCESSES

Commitment to safety by NextGen stakeholders—such as JPDO partner agencies, airlines, airport operators, and equipment vendors—is vital to the success of risk modeling efforts and the prevention of accidents. The best way to ensure safety is a priority is to integrate it into wider business processes, both tactical and strategic: "safety engineering is effective when it participates in and provides input to the design process" [1]. For example, risk modeling and safety concerns should be part of stakeholders' research & development pipelines and innovation initiatives. This integration ensures that safety concerns are addressed during the process of implementing new ideas and operational concepts. FAA's Ideas to In-Service process (I2I) is one example of an innovation initiative which considers safety but could be enhanced by more explicit safety emphasis and integration with risk modeling [32].

I2I begins with initial ideas from a variety of sources, ushers these ideas through increasingly detailed evaluations, and solicits input from an expanding group of stakeholders. Once feasible solutions are developed, they are prioritized, implemented in the NAS, and maintained throughout their lifecycles.

I2I does include safety concerns in several sub-steps but does not appear to emphasize safety in particular. A possible remedy would be to include appropriate validation steps and specify safety constraints which the I2I process must satisfy; it is also important to involve staff with explicit responsibility for safety and delegate this responsibility clearly.

Because I2I is an incremental process which builds on itself, it should accommodate incremental risk modeling which starts simple and becomes more sophisticated. The risk modeling process can evolve in parallel with the concepts modeled as more developed concepts enable more detailed risk modeling.

Daniel Murray's report on flight safety analysis argues that risk modeling may begin with only a roughly defined concept and a variety of conservative assumptions that overestimate risk [24]. Over time, more information about the concept's effects on NAS operations may enable more detailed and sophisticated analysis; this improved analysis may permit relaxation of assumptions and improve confidence in the results. The most detailed types of risk modeling permit relaxation of many starting assumptions. Analysts may choose to model different hazards or subsystems at different levels of fidelity; these differences may be driven by data quality, data availability, and modeling priorities.

Figure 7 provides a summary of the steps in I2I and incremental risk-based modeling. In both cases, ideas evolve from simple to complex, approximate to precise, and general to specific. Approximate correspondence between risk model evolution and I2I innovation steps suggests these processes can evolve in parallel and support each other from a concept's origin through operational implementation in the NAS.

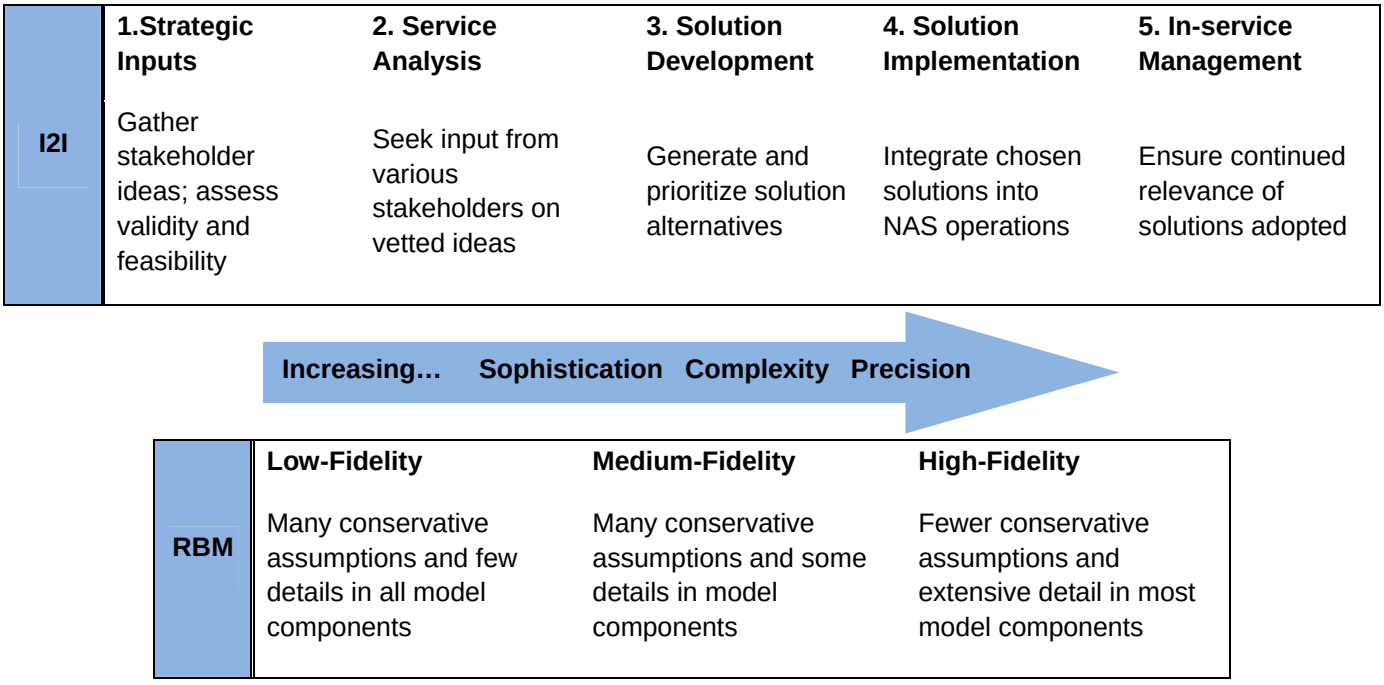

Figure 7. Summary of I2I Steps and Incremental RBM Steps

4.2 RECOMMENDED CONCEPT VALIDATION METHODS

Because NextGen concepts are expected to influence the level of safety in the NAS, these new operational concepts and their risk models will require thorough validation. To maximize cost-effectiveness of the concept validation process, it should begin when concepts are early in their development [33].

The FAA Concept Development and Validation (CD&V) Guidelines report describes best practices for development and validation of NextGen concepts [7]. It also provides a scale of concept maturity levels (CMLs) which should be used to identify appropriate validation techniques for concepts at different levels of maturity. For example, some techniques are only applicable to more mature concepts because they require more detail than is available for less mature concepts. A summary of concept maturity levels is reproduced in Figure 8.

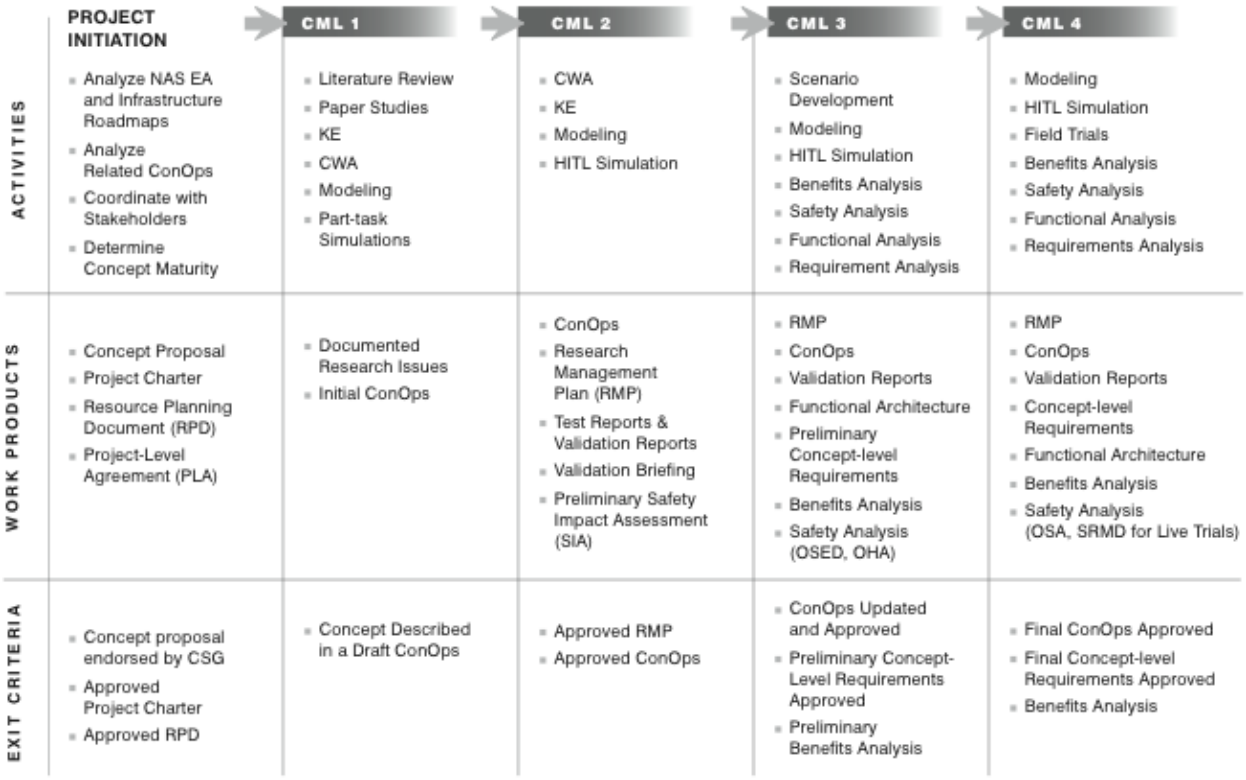

Figure 8. Summary of Concept Maturity Levels (CMLs)

NextGen concept validation from a safety perspective will likely focus on concepts satisfying exit criteria for CML 1; these concepts have a preliminary ConOps identifying their functions along with any necessary assumptions or constraints [7]. For example, a certain concept may only be applicable in certain types of airspace, in certain operational situations, or for aircraft with particular equipage. CML 1 concepts also include a list of open questions for future research but do not yet include any risk modeling or safety assessment results.

Many NextGen concepts have not yet satisfied exit requirements for CML 2; these include a more refined ConOps defining detailed functional requirements, user interfaces, supporting infrastructure, and human factors concerns [7]. CML 2 concepts have also completed a Safety Impact Assessment (SIA) providing potential safety hazards and benefits, though their hazard severities and likelihoods are not necessarily understood yet [7].

CML 3 and CML 4 concepts are considered mature enough to be approved and transitioned to the implementing organization or service unit; approval and transition may occur at either level, depending on the particular concept.

The CD&V report also proposes a list of concept validation techniques [7]:

1. Paper studies to assess safety and risk, cost-benefit analysis, and theoretical concept of operations

2. Knowledge elicitation with stakeholders and subject matter experts (SMEs)

3. Cognitive walk-throughs of operational improvements to develop deep understanding of new procedures or functions and their potential consequences

4. Modeling to examine how a system works or could be optimized

5. Human performance studies to estimate potential error rates, response times, workloads, and other human factors issues

6. Fast-time simulation studies to assess capacity, efficiency, and potential safety issues

7. Real-time human-in-the-loop (HITL) simulations to study the system's impact on humans and overall system performance

Among this list of validation techniques, the first four—paper studies, knowledge elicitation, walk-throughs, and modeling—are appropriate for NextGen concepts at CML 1. Modeling and analysis may be applied at any level of abstraction and hence can be customized to any CML. Knowledge elicitation from SMEs and walk-throughs of new procedures similarly may be performed at any CML; however, walk-throughs are more applicable to more mature CML's whose functions are defined sufficiently well for SMEs to draft detailed ConOps and procedures for them.

Human performance studies, fast-time simulations, and real-time HITL simulations most likely require more assumptions and detail than is available at CML 1; they are recommended for concepts at CML 2 and higher [7]. Simulations of a concept at CML 1 may not provide meaningful results because specific requirements and detailed assumptions about the concept's behavior are not yet available. These missing details would likely limit a simulation's level of fidelity; the simulation results would also likely have limited applicability to the concept's later, more refined forms due to the intervening evolution of its requirements and assumptions. If simulations are performed at CML 1 despite these disadvantages, the results must be interpreted carefully if used to support a more detailed safety case after the concept reaches CML 3 or 4.

As an example, TBO appears to be a CML 1 concept because its basic functions, constraints, operational environment, and open research questions are documented and generally agreed upon [25,34]. TBO is evolving toward CML 2 because initial safety studies are underway but not yet complete; also, TBO requirements are not yet defined because detailed functions and system architectures are still in development. The TBO Study Team report and Capability Safety Assessment are examples of early concept validation work for TBO [25,34].

While the current framework described in the CD&V report indicates that safety assessments are not required prior to CML 2, the lack of such studies does not imply that they are impossible to complete [7]. By the time an operational improvement concept has reached CML 1, it is already well scoped in a proposal that outlines "how the elements of the ATM systems—including personnel, technology, and procedures—will work together to meet a set of well-defined goals." Furthermore, the concept development and validation process requires a functional analysis to describe the means by which a concept's objectives are achieved. The functional analysis breaks down high-level functions in to low-level sub-functions that work towards realizing an operational concept. Regardless of the level of detail, performing a functional analysis permits and eases application of STPA for hazard and safety analysis.

Again, even at CML 1, when the details and requirements of an operational improvement are not yet defined, a safety analysis is still possible and recommended. In fact, because of the indefinite nature of the proposed operational improvements, there is still leeway to make adjustments to a concept's algorithmic and functional framework; that is, there is still the possibility of designing safety into a concept.

A benefit of using the STAMP/STPA analysis process is that as the implementation details of an operational improvement are filled in, the functional system description is expanded to include additional sub-functions of the original sub-functions. Thus the structure preserves any prior results relating to an initial safety analysis while seeking to expose potential hazards and safety mechanisms that can be designed into the newly included sub-function level. TOPAZ's agent based models are similar, in that as a system becomes better defined, the agent models can only improve.

Ultimately, application of STPA at CML 1 is suggested because it may aid in exposing potential safety hazards and lead to discovery of mitigation techniques.

Table 5 suggests reviewed models which may be appropriate to apply at each CML in support of concept validation. In general, more detailed probabilistic models will be better suited to more mature concepts with more information available. STAMP/STPA, TOPAZ, and ASRM should be well-suited to CML 1 and CML 2 concepts because these models appear to require less detailed inputs than others reviewed. Furthermore, STAMP/STPA is designed to handle early stages of system design, while TOPAZ is well suited for exploratory modeling. Models which take statistics from accident reports as inputs will need to rely more on expert judgment when applied to concepts not yet implemented in the NAS; these concepts will not yet have any related accident reports. At later stages of development, models like CATS, IRP, and ISAM become more useful, particularly at estimating system-wide risk. TOPAZ is also potentially useful at later stages in the CML timeline, as it is also capable of establishing risk.

Table 5

Recommended Models by Concept Maturity Level

	CML			
Model	**1**	**2**	**3**	**4**
CATS			X	X
IRP			X	X
ISAM			X	X
STAMP/ STPA	X			
ASRM		X		
TOPAZ	X	X	X	X

4.3 VERIFICATION AND VALIDATION OF DECISION SUPPORT ALGORITHMS

Many NextGen operational improvements such as TBO will involve increased automation or algorithms computing decision support solutions in real time. In these cases, in addition to performing STPA analysis at CML 1, an analysis of algorithmic failure points should be explored when possible. In particular, if an operational improvement has multiple well-documented solution methods (i.e., algorithmically or mathematically described), then each method should be analyzed to discover if there exist any weaknesses or failure points inherent in them. For example, in the case of heuristic algorithms like TCAS, it is important to verify that each possible encounter scenario maps to a single value in the decision space and, more importantly, that there are no indeterminate solutions. Figure 9 illustrates a well-posed solution mapping where each region of the scenario space maps to a single solution in the decision

space. A heuristic is ill-posed when two scenario spaces overlap and map to two different decisions in the solution spaces, as shown in Figure 10, or when some scenarios have no solution. In the case of a collision avoidance system, such a case would occur if an encounter with an intruder aircraft could result in both a "descend" command and an "ascend" command. Another potential problem with discrete solution decision support tools is chatter of the solution. When a scenario lies near the boundary of two scenario spaces, noise can enter the decision making process, resulting in a time-dependent, rapidly oscillating decision solution (e.g., "ascend" → "descend" → "ascend" …). Ill-posed heuristics and chattering are two examples of common algorithmic problems encountered for decision systems with discrete solutions. Undesired behaviors such as these must be considered in the safety validation of a concept.

For mathematical solution methods that are part of an operational improvement or concept of operations, common failures such as discontinuities and numerical instabilities must be explored. Some of the most challenging algorithmic failure cases involve proving a solution exists for each scenario and that numerical methods exist to solve for the solution. Each of these failure modes presents a significant safety risk if a decision-support tool is unable to calculate a solution necessary for the proper implementation of an operational improvement.

Early exploration of potential algorithmic failure points improves decisions related to the viability of a concept. If failure modes are not discovered until late in the development process, then late-term solutions may only "band-aid" problems or in the worst-case require a complete reset of the concept development process.

After application of STPA and once CML 2 is reached, or at more progressed versions of CML 1, a specific implementation of the operational improvement is selected. At this point, the remaining safety assessments described in [7] and listed in Figure 8 can be performed. In any simulations, the ideas of conservatism discussed in Section 3.1 can be applied.

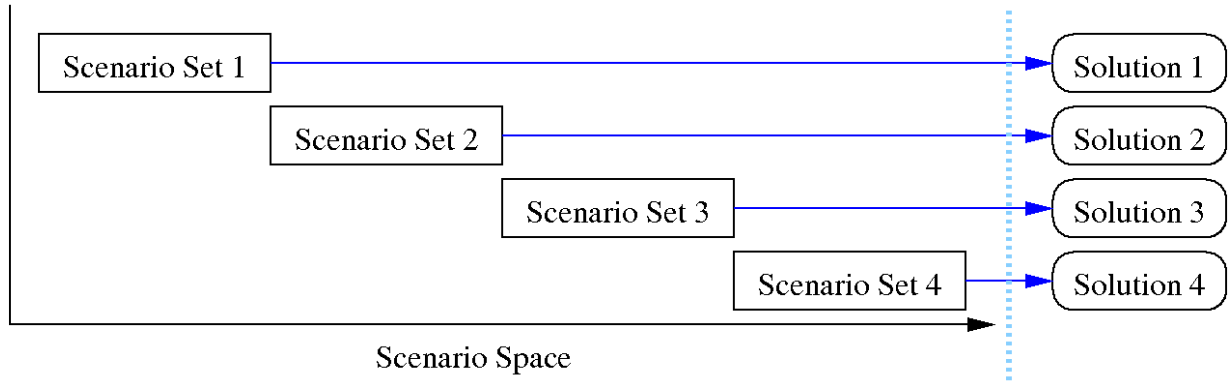

Figure 9. Well-Posed Mapping by a Decision Support Tool

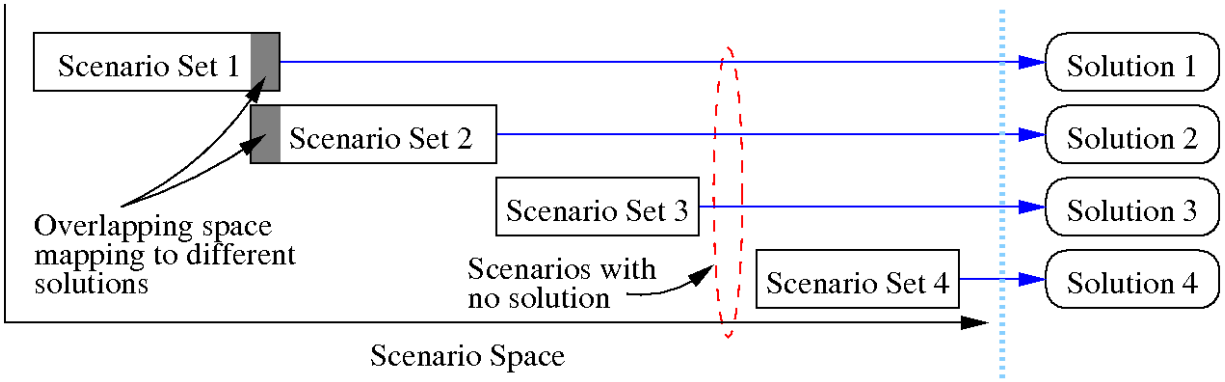

Figure 10. Ill-Posed Mapping by a Decision Support Tool

5. RECOMMENDATIONS FOR MODELING AND MODEL VALIDATION

This section covers suggested assumptions and methods for building and applying risk models—similar to those in Section 2.2—to NextGen concepts.

5.1 MODELING ASSUMPTIONS FOR INITIAL CONCEPT DEVELOPMENT

Most risk models reviewed in this report are intended for fully-developed systems with well-defined behavior. In order to apply them during initial NextGen concept development, when system behavior is not yet well-defined, some changes to model architecture and inputs will be needed. These changes will likely involve additional assumptions to constrain initial models of system and operator behavior; such assumptions would be formulated through discussion with experts and would fill in limited system descriptions or substitute for any missing data. Several types of assumptions are possible:

1. Assume values for unknown influences. Models could initially assume a particular level of OI performance before actual performance is known or agreed upon by experts.

2. Assume limited influences. Models could initially assume concepts are only partly implemented (as ISAM allows) or have relatively limited influence. Under such a partial-implementation assumption, behavior of future ATM systems should be somewhat similar to their better-understood current behavior, hence simplifying the model.

3. Assume limited or no interactions. Models could initially limit each concept's influences to a subset of ATM systems or assume influences are independent of each other.

4. Model only high-priority concepts. Models could initially consider only a subset of high-priority concepts such as "parent" concepts as defined in ISAM; these enable or enhance other concepts and hence may have greater impact. One example is integration of weather information into ATM automation systems, which will improve low-visibility landing capabilities [35]. Parent concept development may need to be prioritized due to these dependencies. Stakeholder input, predicted safety benefits, and costs may also influence the prioritization.

Another way to simplify risk modeling during initial concept development is to reduce a model's fidelity or level of detail. There are several possible methods:

1. Increase the level of abstraction. Less detailed models may require less input data and analyst effort; they also may more easily accommodate simplifying assumptions. For example, general hazard categories may require less effort to model than individual hazards, though more assumptions may be required [24].

2. Prioritize hazards rather than determining absolute risk. Determining which hazards pose the greatest relative risk may require less effort than determining each hazard's absolute risk. The TBO CapSA Team apply this approach and note that full quantitative risk modeling of TBO is not yet possible due to limited current understanding of this NextGen capability; their report also states that the process "is by and large a qualitative one" due to the assumptions required [24].

Some combination of these assumptions and approaches could simplify initial risk-based modeling while still permitting insights. When simplifying assumptions are used, they should be conservative and represent credible worst-case behavior whenever possible as discussed in Section 3.1. Care is needed when using simplifying assumptions. Each assumption—especially those that are nonconservative—should be clearly stated alongside results. Doing so will help prevent the false belief that a system is guaranteed safe when applying nonconservative assumptions.

5.2 MODEL VALIDATION

This section covers risk model validation—an activity which ensures risk models make a valid contribution to the more general goal of concept validation (Section 4).

Successful risk model validation does not guarantee that concept validation will also be successful. For example, a concept's risks may be successfully modeled with sufficient detail and realism, yet the result may show that the concept does not satisfy safety requirements. Without additional mitigations or other safety enhancements, this result would prevent concept validation and approval of the concept.

Early risk model versions for NextGen concepts are likely to be conceptual or qualitative rather than numerical or computer-based. Due to these models' limited quantitative structure, peer review will likely be the most appropriate validation method; it has proven a powerful tool for model development, verification, and validation. To enable peer review, risk modeling teams should publish details on data sources, assumptions, methods, software implementations, and results for review by the NextGen community and researchers in relevant fields. Professional journals and conferences can serve to disseminate material and attract the interest of expert reviewers new to the model. Reviewers should include management and ATM system operators such as pilots and ATC staff; all reviewers should be independent of the model development team. A formal peer-review process—commonly found with professional journals—provides a robust feedback mechanism. Ultimately, peer review provides increased confidence in model validity even in the absence of full quantitative validation. Also, a crucial part of any verification and validation process is thorough documentation outside the publication process. Maintaining up-to-date documents clearly describing the purpose, function, and structure of each concept element and model component—both individually and in relation to others—will aid in establishing credibility and ultimately acceptance or approval [36].

Risk models of concepts in early development will be driven mostly by assumptions rather than operational data or results of simulation and testing [24]. Therefore, in these cases, validation of the model mostly reduces to validation of supporting assumptions, though model developers must also validate the methods used to derive insights from these assumptions. Hypothetical "what-if" scenarios—also known as "thought experiments"—provide one possible method which expert reviewers might use; for example, they could investigate whether the model's assumptions are reasonable even under credible worst-case conditions [33].

When expert opinion is solicited, accepted best practices should be used to organize and validate it—for example, by consulting at least 25 independent SMEs in each of the specialties applicable to a model [6]. Applicable specialties might include pilot tasks, ATC tasks, maintenance tasks, or more specific specialties as needed. Model developers must take care to prevent or mitigate a variety of potential biases which can arise while collecting expert input through surveys or group meetings [37]. These effects may be subtle yet statistically significant. A short list of examples is provided below, while a more comprehensive list is included in a review by Podsakoff, MacKenzie et al. [37]:

1. Conflict of interest. The experts consulted may have a financial interest or other personal interest in approval or nonapproval of a particular model or concept.

2. Context bias. Recent events or experiences may disproportionately influence expert opinions. For example, assessment of ATC performance by ATC staff may be biased by recent operational experiences. Context effects may also arise when the order of questions within a survey influences responses.

3. Consensus-seeking. Individual experts on a panel may not voice their concerns about the group's conclusions due to a desire for consensus; this type of bias is informally known as "groupthink."

4. Question characteristic effects. Experts may be biased toward certain responses to survey questions due to survey design or other features of the system used to gather their opinions. One example is a leading question: a question whose wording suggests a particular response and hence biases results.

5. Social desirability effects. Experts may avoid certain responses in order to conform to the opinions of stakeholders or fellow experts. For example, they may underestimate or overestimate the influence of automation due to stakeholder opinions about its role in ATM systems.

Addressing these concerns will likely require careful vetting and training of the SMEs consulted and the assistance of professionals with training in psychology research methods. Likewise, it is important that expert opinion is solicited in a manner that limits sources of bias. While communities of experts might be small within specific fields, attempts should be made to solicit comment from those not

directly related to the concept being developed nor the safety study. Furthermore, this process should occur individually, as opposed to gathering experts together in order to prevent groupthink. Like data-driven methods, SME responses should be well documented and freely available for review as part of the wider validation of the model building process. Regardless of the method and procedures, a structured process for eliciting expert knowledge is required, and should be documented. Two such methods that are well described and documented are provided by Cooke and Goossens [38].

The techniques for concept validation—discussed earlier in this section—are also applicable to model validation. In particular, a NAS-wide risk-based model's individual components—such as a TCAS sub-model—will undergo a similar validation process. However, complete model validation may be impossible using the same procedures due to a lack of available data. For example, while NAS-wide HITL studies are theoretically possible for a sequenced OI implementation schedule, the amount of effort needed to gather sufficient data may make such an effort infeasible.

Users may choose to combine components from different models to take advantage of a greater variety of capabilities. Additional validation concerns arise in this case since models may have different developers and assumptions. When combining components, it is necessary to verify compatibility at the interfaces—including consistency of assumptions and data sources. Per standard systems engineering practice, it is also necessary to validate the entire integrated model as a whole after validating individual components [8].

6. MODELING SOFTWARE RECOMMENDATIONS

This section covers suggestions for the software and information technologies used to implement risk models. These guidelines can be applied regardless of the particular model or models chosen.

6.1 SOFTWARE DESIGN AND INTERFACE OF RISK MODELS

Early versions of integrated NextGen risk models should not be expected to produce exact results. However, they should be designed to be agile and adjustable to required uses and available data; they will likely evolve along with the modeled elements of NextGen.

Several key software features should be included in any risk model software implementations used to model safety effects of NextGen concepts. These features will enhance model utility and user experience:

1. Open architecture and transparency. Open access to model designs, source code, and documentation, at least within the NextGen risk modeling community, will allow model developers to more quickly improve features and eliminate errors. This approach has been used successfully in the scientific community, the open-source software community, and DoD acquisition processes. There is increasing acceptance of open architecture in technology development projects [39]. Furthermore, developers should actively request review of the approach, methods, and individual model components by experts in their relevant fields. Researchers should be able to easily edit, add, and remove components for personal use. Model development can leverage standard project management practices such as common programming languages, application programming interfaces (APIs), and software versioning and revision control systems (e.g., Subversion and Git). Finally, it may be beneficial to create a mechanism for integrating changes proposed by outside researchers into the model.

2. Flexible import of data and expert opinion. Input data and expert opinion may be provided in a variety of formats which must be consistently imported to the model. The ability to recognize a variety of standard formats will enhance the ability of model developers and end-users to achieve their goals. Use of existing tools for data conversion may efficiently provide this capability by avoiding duplication of existing capabilities.

3. Flexible export of modeling results. Risk modeling results may need to be visualized and post-processed in a variety of ways to satisfy stakeholder expectations. The ability to export in a variety of standard formats will allow use of the many existing tools for data processing and visualization. Use of such tools will reduce model development effort by avoiding duplication of existing capabilities.

4. Flexible user interface (UI). Risk models may have multiple types of end-users. Models should be accessible to each type of end-user and satisfy their respective needs. One approach to designing a model's UI is to default to a simplistic interface with a high-level view of the model and options based on standard settings or inputs. As users wish to transition to more advanced descriptions or options, they can expand individual components via a graphical UI, which is often preferable to a detailed, text-based configuration file possibly containing settings outside the scope of the user. The benefit of this approach is that the resulting UI is useful to a variety of users: policy makers, those with very specific needs, and designers or analysts who need to adjust low-level parameters or update functional elements such as BBNs. User documentation should reflect the stratification of the UI and be specialized to each type of user. Additional functionality should always be cross-referenced to aid the user in finding more advanced options. Fundamentally, an "expanding model" reveals what is valuable to the user and hides what is not, preventing any possible confusion.

5. Service-oriented architecture. Current RBM is based on multiple independent tools and databases, though a trend toward comprehensive, integrated tools appears in models such as ISAM and CATS. In a service-oriented architecture, multiple databases can be accessed from a single interface and analysis tools applied as services. This approach requires building an information management infrastructure to provide access to separate databases, defining data ontologies to enable efficient data searches, and implementing a suite of tools to extract and process data for modeling and analysis.

6. Outputs adapted to input quality. Due to initial uncertainty, early risk model versions should focus on providing relative, qualitative results. Example output statements include "operational improvement X will not result in greater risk" or "it is unclear whether operational improvement Y will result in greater risk." These qualitative statements should include appropriate cautionary advisories so users are aware of their limitations; the advisories should be included until the mechanisms and assumptions of the risk calculations can be thoroughly validated. One technical approach for generating qualitative statements is to make use of statistical hypothesis testing to accept or reject a null hypothesis—for example, "operational improvement Z results in few incidents." Qualitative statements could also be based on parameter bounds—considering worst-, expected-, and best-case scenarios, but always tending toward the conservative. Moving forward, as improved supporting data and concept designs allow fine tuning of models, outputs can adjust accordingly—ultimately including more precise quantitative statements such as "the rate of collision is Z events per flight hour."

6.2 DATABASES AND INFORMATION SHARING

Standardization and reusability of data and models should be primary objectives; they will assist in quality control, interpretation of results, reduced duplication of effort, and more complete studies.

By extension, it is beneficial to share a common taxonomy across relevant databases. Whether it is information gathered from the Aviation Safety Reporting System, the ICAO Accident Data Reporting System, or any other accident database (such as those listed in Section 2.2), sharing a common language to describe events will facilitate and ease model building. Given that many of these data sources originate from different organizations and may serve differing purposes, it is unlikely that a common taxonomy will ever exist. In this case, it is vital that any model constructed from multiple data sources provide a clear explanation of the method for normalizing and standardizing information. This step will provide a means for verification and validation by external reviewers; if well-defined it may allow for some level of automation as new data is added.

In addition to concerns about data access, model developers have expressed concern about access to subject-matter experts to provide input on model development and validation. Databases providing access to subject-matter experts and their input are a possible remedy to these concerns [6].

The NextGen risk modeling community needs improved methods of archiving and communicating studies that are performed so that stakeholders are aware of previous work, its assumptions, and results. Improved access to standard data sources, accident/incident taxonomies, model components, and subject-matter experts will reduce duplication of effort; it will also facilitate efficient model development and consistent, accurate results.

Table 6

Summary of Recommendations

Number	Section	Description
1	3.1 Conservatism and Hazard Identification	Safety modeling should avoid under-estimating risk. Modeling should apply concepts of conservatism.
2	3.1 Conservatism and Hazard Identification	As hazards are identified, their likelihood, severity, and any associated mitigation techniques should be documented.
3	3.1 Conservatism and Hazard Identification	A structured and repeatable hazard identification procedure should be utilized. In addition to other traditional methods, application of STPA is recommended when creating FT and ESD.
4	3.2 Systemic Failures	Safety modeling should consider failure of critical systems and test for graceful degradation and recovery.
5	3.3 Uncertainty	Consider uncertainty by selecting the upper and lower of relevant parameters. Perform perturbation analysis. When possible explore the results as a distribution. For sensitive parameters, conduct additional research to refine values.

Number	Section	Description
6	3.4 Emergent Behaviors	To expose emergent behaviors, models should seek to perform exploratory modeling by studying and adjusting lower level behaviors. Agent based models, such as TOPAZ, are well suited for such a task.
7	4.1 Integrating Safety and Risk Modeling into Stakeholder Business Processes	The I2I process should be adjusted to place greater emphasis on safety. Incremental risk modeling should co-evolve with concept development, becoming increasingly detailed as the concept becomes better defined. The I2I process should involve staff with the explicit responsibility of safety.
8	4.2 Recommended Concept Validation Methods	As part of the FAA Concept Development and Validation process, a function analysis of a concept should be performed at CML1. STPA should be applied. As a concept is developed, any functional diagrams and safety analysis can be expanded.
9	4.3 Verification and Validation of Decision Support Algorithms	In CML1 explore potential mathematical and algorithmic failure points of software which are required to support a concept.
10	5.1 Modeling Assumptions for Initial Concept Development	When concepts are not completely defined, make simplifying assumptions. If applying conservatism, ensure all assumptions are consistently conservative. When required, reduce a model's fidelity by increasing the level of abstraction, or by focusing on relative risk (as oppose to absolute risk). Results and non-conservative simplifying assumptions should be included alongside each other to prevent false belief in safety.
11	5.2 Model Validation	Maintain an active publication strategy. Make best use of the peer review process to improve models. Provide easy access to up-to-date documents.
12	5.2 Model Validation	Perform hypothetical "what-if" scenarios. Test model assumptions under credible worst-case conditions.
13	5.2 Model Validation	When involving SMEs, use best practices (e.g., 25 independent SMEs in corresponding field; SMEs not directly involved with project or concept). Prevent or mitigate potential biases when collecting input through surveys or interviews.
14	5.2 Model Validation	Like data, SME input should be well documented and freely available for review.

Number	Section	Description
15	5.2 Model Validation	When combining components of a model, verify compatibility at interfaces (e.g., consistency of assumptions and data sources). Validate as a whole any combined components.
16	6.1 Software Design and Interface of Risk Models	Provide open access to model designs, source code, and documentation. Leverage standard project management practices: common programming languages, application programming interfaces, and software versioning control.
17	6.1 Software Design and Interface of Risk Models	Provide for flexible import of data and expert opinion.
18	6.1 Software Design and Interface of Risk Models	Provide for flexible export of modeling results.
19	6.1 Software Design and Interface of Risk Models	Design flexible user interfaces to adjust to the needs and expertise of the end-user. Provide an expandable GUI interface.
20	6.1 Software Design and Interface of Risk Models	A service-oriented architecture should be taken as part of a risk-based model. Provide information management infrastructure to provide access to databases, define a common ontology, and complete a suite of tools to extract and process data.
21	6.1 Software Design and Interface of Risk Models	Output of the risk model should reflect the uncertainty of the input data and model parameters.
22	6.2 Databases and Information Sharing	In the case of differing taxonomy of data sources, a clear explanation is needed for normalizing and standardizing data.
23	6.2 Databases and Information Sharing	Clearly document and archive work efforts. Communicate studies so stakeholders are aware of previous work, its assumptions, and results. Allow for easy access.

This page intentionally left blank.

GLOSSARY

ADREP	ICAO Accident Data Reporting System
ADS-B	Automatic Dependent Surveillance-Broadcast
AIM	Accident-Incident Model
AMAN	Arrival manager
ANSP	Air Navigation Service Provider
API	Application Programming Interface
ASDE-X	Airport Surface Detection Equipment, Model X
ASRM	Aviation Safety Risk Model
ATC	Air Traffic Control
ATM	Air Traffic Management
AvSSP	Aviation Safety and Security Program
BBN	Bayesian Belief Network
CapSA	Capability Safety Assessment
CATMT	Collaborative Air Traffic Management Technologies
CATS	Causal Model for Air Transport Safety
CD&V	Concept Development and Validation
CFIT	Controlled Flight Into Terrain
CML	Concept Maturity Level
ConOps	Concept of Operations
CSG	Concept Steering Group
CSRL	Complex Systems Research Laboratory
CSS-Wx	Common Support Services–Weather
CWA	Cognitive Walkthrough Analysis
DoD	Department of Defense
ERAM	En Route Automation Modernization
ESD	Event Sequence Diagram

FAA	Federal Aviation Administration
FC	Flight Crew
FIM	Flight deck-based Interval Management
FL	Flight Level
FT	Fault Tree
GPS	Global Positioning System
GUI	Graphical User Interface
HAZOP	Hazard and Operability Study
HITL	Human-In-the-Loop
HFACS	Human Factors Analysis and Classification System
I2I	Ideas to In-Service
ICAO	International Civil Aviation Organization
IRP	Integrated Risk Picture
ISAM	Integrated Safety Assessment Model
ITP	In-Trail Procedure
JPDO	Joint Planning and Development Office
KE	Knowledge Elicitation
LAAS	Local Area Augmentation System
LOSA	Line Operations Safety Audit
MIT	Massachusetts Institute of Technology
MTCD	Medium Term Conflict Detection
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NextGen	Next Generation Air Transportation System
NLR	Netherlands National Aerospace Laboratory
NNEW	NextGen Network-Enabled Weather renamed to CSS-Wx
NOP	Network Operations Portal
NSF	National Science Foundation

NVS	NAS Voice System
OI	Operational Improvement
PLA	Project-Level Agreement
RARM	Risk Assessment & Risk Management
RBM	Risk-Based Modeling
RMP	Research Management Plan
RNAV	Area Navigation
RNP	Required Navigation Performance
RPD	Resource Planning Document
SADT	Structured Analysis and Design Technique
SBS	Surveillance and Broadcast Services
SDE	Stochastic Differential Equations
SESAR	Single European Sky ATM Research
SIA	Safety Impact Assessment
SME	Subject-Matter Expert
SMS	Safety Management System
SSR	Secondary Surveillance Radar
STAMP	System-Theoretic Accident Model and Processes
STCA	Short Term Conflict Alert
STPA	System-Theoretic Process Analysis
SWIM	System-Wide Information Management
TBFM	Time-Based Flow Management
TFMS	Traffic Flow Management System
TBO	Trajectory Based Operations
TCAS	Traffic Alert and Collision Avoidance System
TOPAZ	Traffic Organization and Perturbation Analyzer
VBA	Visual Basic for Applications

This page intentionally left blank.

REFERENCES

[1] N.G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*, MIT Press, 2012.

[2] R. Mendez, "NextGen Trajectory Based Operations Solution Set," in *FAA TBO Conference*, 2009.

[3] "Review of Techniques to Support the EATMP Safety Assessment Methodology – Volume 1," EEC Note No. 01/04, 2004.

[4] M. Everdij, "Unified Framework for Risk Assessment & Risk Management," 2012.

[5] "A Concept Paper for Separation Safety Modeling," 1998.

[6] "Causal Model for Air Transport Safety, Final Report," 2012.

[7] "Concept Development and Validation Guidelines (Final Version 1.0)," 2011.

[8] *NASA Systems Engineering Handbook*, National Aeronautics and Space Administration, 2007.

[9] N.G. Leveson, STPA and CAST Tutorial, 2012.

[10] "Methodology Report for the 2012 Integrated Risk Picture for Air Traffic Management in Europe," 2006.

[11] "Main Report for the 2005/2011 Integrated Risk Picture for Air Traffic Management in Europe," 2006.

[12] "SESAR Top-Down Systemic Risk Assessment," 2009.

[13] "Note On Risk Model Validation," 2009.

[14] E. Perrin, B. Kirwan, and R. Stroup, "A System model of ATM Safety: The Integrated Risk Picture," in *USA/Europe ATM R&D Seminar*, 2007.

[15] A.L.C. Roelen, H. Zmarrou, H. Eghbali, and A. Sternberg, "Quantification of FAA Causal Model Elements," NLR-CR-2010-102, 2010.

[16] Future Evolution of Aviation Safety Metrics (Draft), 2011.

[17] J.T. Luxhoj and D.W. Coit, "Modeling Low Probability/High Consequence Events: An Aviation Safety Risk Model," in *Reliability and Maintainability Symposium*, 2006.

[18] F. Netjasov and M. Janic, "A review of research on risk and safety modeling in civil aviation," *Journal of Air Transport Management*, vol. 14, no. 4, pp. 213–230, 2008.

[19] J.T. Luxhoj and R. Sarlo, "Decision Analysis for Safety Risk Modeling of Unmanned Aircraft Systems," in *Industrial and Systems Engineering Research Conference*, 2012.

[20] J.T. Luxhoj and K. Topuz, "Sensitivity Analyses of Risk Factors and Mitigation Effects on a Wake Vortex Encounter Flight Scenario," in *Industrial and Systems Engineering Research Conference*, 2012.

[21] H.A.P. Blom et al., "Accident Risk Assessment for Advanced Air Traffic Management," in *Air Transportation Systems Engineering*, George Donohue and Andres Zellweger, Eds.: American Institute of Aeronautics and Astronautics, 2001.

[22] S.H. Stroeve, H.A.P. Blom, and G.J. Bakker, "Contrasting safety assessments of a runway incursion scenario: event sequence analysis versus multi-agent dynamic risk modelling," *Accepted to Reliability Engineering & System Safety*, 2012.

[23] H.A.P. Blom, S.H. Stroeve, and H.H. de Jong, "Safety risk assessment by Monte Carlo simulation of complex safety critical operations," National Aerospace Laboratory.

[24] D. Murray, "A Tiered Approach to Flight Safety Analysis," in *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, 2006.

[25] "Capability Safety Assessment of Trajectory Based Operations," 2011.

[26] (2009, April) Atlanta airport reopens after lightning strike. [Online]. http://articles.cnn.com/2009-04-23/us/ga.airport.storms_1_control-tower-outage-kathleen-bergen?_s=PM:US

[27] E. Paz-Frankel (2007, September), Telecom glitch stops departures at Memphis International. [Online]. http://www.bizjournals.com/memphis/stories/2007/09/24/daily12.html

[28] "FAA Oversight Is Key for Contractor-Owned Air Traffic Control Systems That Are Not Certified," 2011.

[29] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2010.

[30] S. Bankes, “Exploratory Modeling for Policy Analysis,” *Operations Research*, vol. 41, no. 3, pp. 435–449, 1993.

[31] N.N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, 2007.

[32] “Proposed I2I Process Map,” 2012.

[33] R.G. Sargent, “Verification, Validation, and Accreditation of Simulation Models,” 2011.

[34] “Trajectory Based Operations Study Team Report,” 2011.

[35] NextGen Joint Planning Environment [Online]. http://jpe.jpdo.gov

[36] J.P.C. Kleignen, “Theory and Methodology: Verification and validation of simulation models,” *European Journal of Operations Research*, vol. 82, pp. 145–162, 1995.

[37] P.M. Podsakoff, S.B. MacKenzie, J.-Y. Lee, and N.P. Podsakoff, “Common method biases in behavioral research: A critical review of the literature and recommended remedies,” *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879–903, 2003.

[38] R.M. Cooke and L.H.J. Gossens, “Expert judgement elicitation for risk assessments of critical infrastructures,” *Journal of Risk Research*, vol. 7, no. 6, pp. 643–656, 2004.

[39] “DoD Open Systems Architecture Contract Guidebook for Program Managers,” 2011.

This page intentionally left blank.