August, 2016

Toward Risk Assessment 2.0: Safety Supervisory Control and Model-based Hazard Monitoring for Risk-informed Safety Interventions

Francesca M. Favaro, *Georgia Institute of Technology*
Joseph H. Saleh, *Georgia Institute of Technology*

Toward Risk Assessment 2.0: Safety Supervisory Control and Model-Based Hazard Monitoring for Risk-Informed Safety Interventions

Francesca M. Favarò, Joseph H. Saleh*
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta GA

Abstract: Probabilistic Risk Assessment (PRA) is a staple in the engineering risk community, and it has become to some extent synonymous with the entire quantitative risk assessment undertaking. Limitations of PRA continue to occupy researchers, and workarounds are often proposed. After a brief review of this literature, we propose to address some of PRA's limitations by developing a novel framework and analytical tools for model-based system safety, or safety supervisory control, to guide safety interventions and support a dynamic approach to risk assessment and accident prevention. Our work shifts the emphasis from the pervading probabilistic mindset in risk assessment toward the notions of danger indices and hazard temporal contingency. The framework and tools here developed are grounded in Control Theory and make use of the state-space formalism in modeling dynamical systems. We show that the use of state variables enables the definition of metrics for accident escalation, termed hazard levels or danger indices, which measure the "proximity" of the system state to adverse events, and we illustrate the development of such indices. Monitoring of the hazard levels provides diagnostic information to support both online and offline safety interventions. For example, we show how the application of the proposed tools to a rejected takeoff scenario provides new insight to support pilots' go/no-go decisions. Furthermore, we augment the traditional state-space equations with a hazard equation and use the latter to estimate the times at which critical thresholds for the hazard level are (b)reached. This estimation process provides important prognostic information and produces a proxy for a time-to-accident metric or advance notice for an impending adverse event. The ability to estimate these two hazard coordinates, danger index and time-to-accident, offers many possibilities for informing system control strategies and improving accident prevention and risk mitigation. Finally we develop a visualization tool, termed hazard temporal contingency map, which dynamically displays the "coordinates" of a portfolio of hazards. This tool is meant to support operators' situational awareness by providing prognostic information regarding the time windows available to intervene before hazardous situations become unrecoverable, and it helps decision-makers prioritize attention and defensive resources for accident prevention. In this view, emerging risks and hazards are dynamically prioritized based on the temporal vicinity of their associated accident(s) to being released, not on probabilities or combination of probabilities and consequences, as is traditionally done (offline) in PRA.

This approach offers novel capabilities, complementary to PRA, for improving risk assessment and accident prevention. It is hoped that this work helps to expand the basis of risk assessment beyond its reliance on probabilistic tools, and that it serves to enrich the intellectual toolkit of risk researchers and safety professionals.

Key Words: Hazard Monitoring; Temporal Contingency; Model-based System Safety; Safety Supervisory Control.

1. Introduction

Probabilistic Risk Assessment (PRA) was first developed in the nuclear industry in the mid 1970s [NUREG, 1975; Apostolakis, 2004; Mosleh, 2014], and in the following decades it was broadly adopted in different industries. Despite its appeal, PRA is not without its flaws, and in recent years researchers have highlighted some of its limitations and proposed several improvements. Two important limitations are related to the difficulties of PRA to account for time-related considerations in relation to accident scenarios (details in Section 2) and to properly handle software issues, just to mention a few. In this work, we propose to address in part these (and other) limitations by developing a novel framework and formal tools for **model-based system safety**, which we also term **safety supervisory control**. Our approach has two fundamental ingredients: (1) the use of state-space models and state variables (from

* Corresponding author. Tel: + 1 404 385 6711; Email address: jsaleh@gatech.edu (J. H. Saleh)

Control Theory) to capture the dynamics of hazard escalation, and to both model and monitor "danger indices" in a system; and (2) the adoption of Temporal Logic (from Computer Science and Software Engineering) to model and verify system safety properties (or their violations, hence identify vulnerabilities in a system). In this work we tackle the first ingredient; the Temporal Logic aspect is developed in a companion article [Favarò and Saleh, 2016b]. We briefly justify the adoption of the state-space models next and leave its detailed examination to Section 3.

The framework and analytical tools here developed are grounded in Control Theory and make use of the state-space representation in modeling dynamical systems. The use of state variables allows the definition of metrics for accident escalation, termed hazard levels or danger indices (used interchangeably hereafter), and they measure the "proximity" of the system to adverse events. Furthermore, the adoption of state-space formalism, as we will show hereafter, allows the estimation of the times at which critical thresholds for the hazard level are (b)reached. This estimation process provides important prognostic information and produces a proxy for a time-to-accident metric or advance notice for an impending adverse event. The hazard levels and the time-to-accident metrics create a portfolio of hazard coordinates that can then be displayed dynamically in a "hazard temporal contingency map" to support operators' situational awareness and help them prioritize attention and defensive resources for accident prevention. **The idea and capability of measuring the proximity to a performance goal is essential for proper control of a system—this is a fundamental concept in Control Theory. By extension, the ability to measure the proximity of a system to adverse events,** proximity in the form of hazard levels or danger indices, **is crucial for accident prevention and sustainment of system safety**. It also makes for improved dynamic risk assessment and management. This measurement of proximity to adverse events is enabled by the use of state variables and a model-based approach to system safety. The monitoring of hazard levels and the estimation of the time window available for safety interventions provide important feedback for various stakeholders and decision-makers to guide safety interventions both on-line (towards accident prevention and/or mitigation) and off-line (towards re-design and re-engineering of safer systems). This approach augments the current perspective in traditional risk assessment and its reliance on probabilities as the fundamental modeling ingredient with the notion of temporal contingency, a novel dimension by which hazards are dynamically prioritized and ranked based on the temporal vicinity of their associated accident(s) to being released.

This work is part of a larger effort whose objective is to introduce novel formalisms and techniques to improve accident prevention and risk assessment. As noted previously, the offline capabilities of our framework, which include temporal logic to model and verify system safety properties (or their violations) early during the lifecycle of a system (design and development stages) to check the presence/adequacy of safety features implemented in the system design, are examined in a companion work [Favarò and Saleh, 2016b]. In this article, we focus on the ingredients of model-based system safety, the associated modeling of danger indices and hazard equations, and the creation of a hazard temporal contingency map. The integration of all these elements provides a safety supervisory control framework (Figure 1), which continuously scans the system and monitors for emerging hazards, providing diagnostic and prognostic information for safety interventions and in support of accident prevention (details in Section 3). We illustrate some of these capabilities by applying the hazard monitoring process to a rejected takeoff scenario, and show how it provides new insight to support the pilots' go/no-go decisions and to inform regulations and policies regarding safety guidelines. The objective of the present work is threefold: (i) to provide a synthesis of key limitations of PRA and the improvements currently proposed in the literature, since these issues constitute the motivation for our efforts; (ii) to make the case for model-based approaches and the use state variables, in particular in relation to the development of danger indices and the modeling/monitoring of hazard dynamics for improved dynamic risk assessment; (iii) to introduce our safety supervisory control framework and develop its analytical tools for guiding safety interventions, and improving accident prevention.

The remainder of this work is organized as follows. Section 2 provides a brief synthesis of the limitations of traditional PRA and a review of some of the improvements recently proposed in the literature. These limitations and the improvements serve both as a motivation and a basis for our work.

Section 3 introduces our model-based safety supervisory control framework and develops its analytical tools. In particular, subsections 3.1 and 3.2 tackle the state variables mapping into danger indices and the development of the hazard equation, and establish the importance of model-based techniques for risk assessment through detailed examples. Section 3.3 expands on the notion of temporal contingency, and the mapping of the hazard coordinates into the hazard temporal contingency map. Section 4 concludes this work.

2. Limitations of traditional PRA and proposed improvements

In this section, we review some of the limitations of traditional PRA and discuss improvements recently proposed in the literature, in particular the emerging Dynamic PRA tools and methods. For a review of PRA itself, the reader is referred to for example Groen et al., [2002], or Stamatelatos and Defzuli [2011]. The material that follows serves as both a motivation and a starting point for our proposed safety supervisory control framework, with its two ingredients (model-based and temporal logic) examined in the present and the companion work.

Limitations of and concerns with PRA have been recently highlighted by several authors, including Aldemir [2013], Mosleh, [2014], and Zio [2014]. The limitations revolve around:

- **Timing and ordering considerations**: the static logic models used in traditional PRA are insensitive to dynamic process failures. For instance, when multiple top events are considered in a fault tree, "the actual final state of a [truly] dynamic scenario depends on the order, timing, and magnitude of the component failure events" [Zio, 2014], which traditional fault trees cannot capture. Similar arguments can be made with Event Trees. Given the scenario postulated and tested by the analyst, the order of occurrence of the failure events is pre-set resulting in potential vulnerable sequences that remain untested [Aldemir, 2013; Zio, 2014]. Additionally, recovery and other time-dependent performances cannot be combined into the static traditional PRA tools.

- **The inclusion of software failures**: the issue is the understanding of how software failures will affect the overall system, and how to include these considerations in risk assessment. Arguably, PRA has been "very much hardware oriented" [Mosleh, 2014]. A gap exists in current methods to account for software failures or contributions to accidents [Leveson, 1995; Favarò et al., 2013] and model them with tools that are compatible with traditional PRA [Apostolakis, 2004; Kirschenbaum et al., 2009]. The need to leverage new tools and perspectives for software safety analysis has been argued by several authors [DOD, 2012; Zio, 2014; Mosleh, 2014].

- **The issue of human response**: Human Reliability Analysis (HRA) is used to estimate the quantitative and qualitative contributions of human performance to the overall system reliability. Current approaches to HRA interface well with traditional PRA tools and are included in many risk assessment efforts [Swain, 1990]. However, HRA does not account for modeling human response *during* the unfolding of an accident scenario [Mosleh, 2014]. Real-time analysis and/or simulation of the human performance are of paramount importance to estimate how operators' actions affect the estimated frequencies of failure events (and hence the risk estimation). Additionally, the study of human response has close ties with the analysis of instrumentation design and layout, which provide the operator a feedback on the system status. New tools and approaches can aim at relating potential operator performance degradation to ineffective instrumentation layouts, with missing information or misleading interpretations of the signals coming from the plant during an accident unfolding.

- **The inclusion of physical models**: current PRA tools have limited capacity for the integration of physical phenomena models (e.g., physics of failure, environment physics description) [Mosleh, 2014]. Whenever the physics behind failure mechanisms can be included, many restrictions are required in terms of simplification. This issue is related to the lack of a mathematical framework that can integrate and handle different domains of analysis at the same time.

- **The issue of completeness**: PRA is executed by means of abstracting events into classes and categories. The level of abstraction is dependent on the type of decisions that PRA is meant to support, the resources allocated to the PRA effort, and the state of the knowledge regarding the system [Mosleh, 2014]. Related to the level of clustering of the events and to the abstraction effort is the issue of how complete the PRA process is in terms of breath of coverage of the risk and accident scenarios, their depth of causality, and the fidelity in the definition of the basic events of the fault and event trees and their associated probabilities [Mosleh, 2014]. Additionally, the scenarios that are tested are postulated, hence developed a priori, by the analyst, so that traditional PRA cannot by itself discover new accident scenarios [Mosleh, 2014].

New techniques and approaches are currently under investigation to address some of these limitations. A brief overview of some of the most notable approaches follows. We divide them in two categories: those that fall under the heading of Dynamic PRA, and those that rely on the addition of time-properties and time-related considerations. Our approach falls at the intersection of the two as it combines aspects that are common to both categories.

2.1 DPRA: an answer to the time-dependency limitations of traditional PRA

Dynamic PRA comprises a set of simulation-based methods that combine deterministic and probabilistic approaches to account for the time-dependency of the events they try to model. For this reason, DPRA tools also go under the name of Integrated Deterministic and Probabilistic Safety Analysis (IDPSA) [Zio, 2014]. DPRA can handle both continuous and discrete time, as well as hybrid systems, depending on the system model of choice [Aldemir, 2013]. Regardless of the method of choice, three basic inputs are needed for DPRA:

1. A time-dependent physical model of the system dynamics;
2. A list of identified normal and abnormal system configurations;
3. The transition probabilities among the normal and abnormal configurations (or a more complex model of the stochastic rules that govern each transition).

Kirschenbaum et al. [2009], Aldemir [2013], and Zio [2014] provide a survey of different DPRA methodologies including dynamic flowgraph methods, Markov/cell-to-cell mapping techniques, and Petri nets. These techniques have the potential to uncover and identify plant vulnerabilities that were a-priori unknown, and that could not be considered with traditional PRA tools. DPRA enlarges the exploration of the possible accident scenarios space, by including ordering and timing of events [Zio, 2014]. Moreover, simulation-based approaches can provide insight into an accident phenomenology and its causal basis for different accident scenarios [Mosleh, 2014]. This is due to the fact that the sequencing of events is no longer pre-determined by the analyst, but derives from the stochastic simulation itself.

DPRA is not an alternative to the traditional PRA, but rather complementary. Traditional PRA is still used in conjunction with the more sophisticated, but more complex, simulations carried out in DPRA. On one hand, DPRA provides additional insight for complex systems; on the other hand, PRA provides a technique popular for its simplicity and clarity in communicating the results of risk assessments [Aldemir, 2013]. Limitations and drawbacks of DPRA include [Aldemir, 2013; Mosleh, 2014; Zio, 2014]:

- Substantial efforts are needed to generate the data for the transition probabilities among the different configurations. Although DPRA seeks to reduce the need for expert judgment, expert opinions are still required. Additionally, the model input data is not always readily available, so that experimental testing and/or components simulation may be required for the computation of the stochastic rules that regulate the system transitions.

- The development of the system models can be computationally intensive. Many of the available modeling tools suffer from the number of states explosion problem, and the size of the system

under consideration is limited by the current computational capabilities (see [Zio, 2014] for a discussion on possible solutions).

- DPRA is a simulation-based methodology. This implies that the verification effort is never completely exhaustive (i.e., not all possible existing scenarios are tested). Generally, dominant-risk scenarios are given higher priority, but completeness of the testing effort is not guaranteed.

- There are difficulties with the output post-processing and with the classification of the various accident scenarios generated by the tools (e.g., problems with clustering of scenarios by similarity of the event sequences and/or the end state of the system). The classification of the scenario is related to the capability of recognizing unanticipated scenarios [Zio, 2014]. Additional concerns regard what kind of output to generate for risk communication and how to organize and communicate the data produced by the simulation in a clear manner.

DPRA is still far from being broadly adopted as an industrial practice, and considerable research remains underway in this field. Benchmark examples continue to be developed for the consistent comparison of the different risk assessment methodologies [Kirschenbaum et al., 2009; Aldemir et al., 2010].

2.2 Current approaches involving the use of temporal properties

Dynamic PRA has not been the only answer provided by the risk and safety community to the previously highlighted limitations of its static counterpart. With the increasing importance of digital systems, frameworks that leverage approaches derived from and inspired by those used in computer science have also surfaced in the academic literature. These approaches introduce formal languages for the definition of temporal properties to be used in conjunction with the tools of static PRA. We briefly review here some notable works in this area[1].

- **Fault Trees (FT) temporal extensions**: we denote under this heading works aimed at extending the classical fault tree analysis within traditional PRA. Notable contributions in this regard have been made by [Hansen et al., 1998], [Palshikar, 2001], and [Magott and Skrobanek, 2012]. The FT temporal extensions were achieved in several ways:

  - By adding temporal gates to the standard FT notation: in this approach new gates were added to the standard pool of static logical gates (e.g., AND, OR, XOR gates). The new gates activation is dependent on the particular sequence or duration of the events that are fed into them. For instance, the "Priority AND" gate requires the ordered occurrence of the events fed into it from left to right; the "For all t instants" gate requires that the events fed into it hold for t instants of time, basically translating into an AND gate of an event holding at t_1, AND again holding at t_2, AND so on up to time t. Additional examples of temporal gates can be found in [Hansen et al., 1998; Palshikar, 2001; Walker and Papadoupulos, 2009].
  - By adding time dependency in the events definition: rather than adding a temporal dependency inside the gate logic (as it was done for the temporal gates of the previous examples), this approach includes time-related considerations inside the definition of the events that are then connected by static logical gates. A common way of doing this is by adding to each event description a duration interval. The duration interval in turn affects the applicability of the logical gate. For instance, it can prevent an AND gate from being activated unless a minimal duration time for the event is achieved.

    Although not always explicitly stated, these extensions make use of temporal logic (or some rudimentary form of it). In general, they do not require the user to be familiar with the formalities of TL and the techniques for the verification of TL properties. This makes their use simple and approachable, but it hinders the user from tapping into and benefitting from the full potential of these techniques.

[1] A mention should be made to current works aimed at integrating DPRA techniques with temporal extensions of fault trees, see for example [Bouisson and Bon, 2003] or [Rao et al., 2009].

- **Formal logics for the analysis of time-critical systems**: these approaches introduce timed logics for the explicit expression of time-dependent considerations. Timed logics add temporal operators to the pool of classical operators from propositional and predicate logics. Different logics can be used for this purpose such as probabilistic computational tree logic [Johnson, 1995] or real-time logic [Jahanian and Mok, 1986, 1994]. The use of timed logics allows to reason about the ordering and timing of events and to specify the desired dynamical behavior of the system. Specifically, timed logics are used to express time-dependent system requirements and performance constraints. These approaches no longer make use of traditional PRA. They require a model for the system under consideration, such as in DPRA. Contributions that resort to timed logics for system properties specification are somewhat more infrequent in the literature when compared to the above-mentioned approaches that extend traditional PRA tools. A separate mention should be given to applications of timed logics to problems that are not strictly related to risk assessment, but still span safety applications. This is the case for instance of Johnson's work on the use of formal methods for accident investigations [Johnson, 2000], or the application of temporal logic to support human factors engineering [Johnson and Harrison, 1992].

The problem of including time considerations in the risk assessment process is tackled in our approach by two complementary features: on the one side, the hazard levels or danger indices model dynamical quantities that enable, through the estimation of the time-to-accident metric, to account for time-dependent considerations for safety interventions; on the other side, the use of TL allows the explicit inclusion of temporal ordering (through the use of TL operators) within the definition of safety properties that act as constraints for the dynamic behavior of the system (see [Favarò and Saleh, 2016a,b]). The ensuing safety supervisory control framework derives from the same spirit that motivated the contributions presented in this section, and it is presented in detail next.

3. Model-based safety supervisory control and hazard monitoring for guiding safety interventions

The framework proposed in this work adopts a model-based approach and state variables to capture the dynamics of hazard escalation and to monitor "danger indices" in the system. The identification and quantification of indices of proximity to adverse events supports the development of a safety supervisory control approach (shown in Figure 1), and it is particularly helpful for triggering pre-emptive safety interventions and improving accident prevention, as we will argue shortly. The monitoring of the hazard level and the estimation of the time-to-accident metric provide important feedback for various stakeholders, from management and designers, to front-line operators and technicians, to guide safety interventions over different time scales. The monitoring of the distinctive macro-state variables "hazard levels" during system operation (i.e., on-line) provides important feedback for operators to recognize a developing adverse situation, prioritize attention, and allocate defensive resources for safety interventions and hazard de-escalation. Additionally, the off-line application of the safety supervisory process can assist in checking the presence/adequacy of safety features implemented in the system, providing an important feedback during the design stages.

A schematic view of the integrated process we propose is shown in Figure 1.

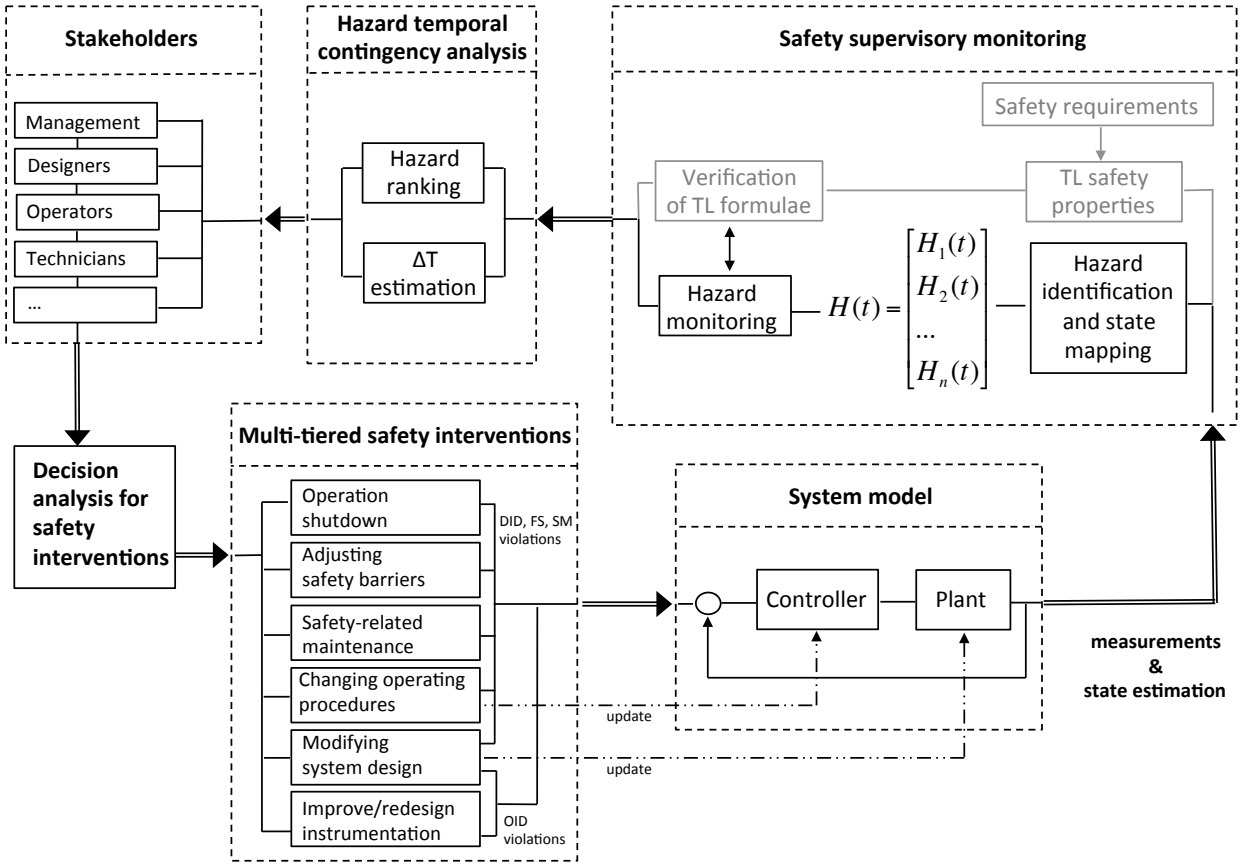

Figure 1. Overview of model-based safety supervisory control, and dynamic hazard monitoring for safety interventions (not meant to be exhaustive; several loops and blocks are not shown to avoid clutter[2]).

We briefly discuss next the various elements in Figure 1 and how they fit together.

The **system model** block in the bottom right corner includes the state-space model of the system under consideration (the "plant" block) and a "system controller" within an inner loop (e.g., a digital controller, a human controller, or a combination of both for varying degrees of automation). The controller provides inputs to the plant, seeking to ensure that the system fulfills its performance requirements (to behave "as expected") and to steer it away from off-nominal hazardous conditions. The reference inputs to the system model block (the input to the comparator upstream the controller box in Figure 1) depend on performance and production requirements, as well as the safety requirements and constraints[3].

Output measurements and system state estimations are undertaken downstream of the "system model" (not shown in a separate "**observer**" block so as not to further clutter the figure). These measurements and state estimation are fed into the macro-block entitled **safety supervisory monitoring** where several functions are performed:

i. **Hazard identification and state mapping**: The hazard level or danger index is an analytic metric for capturing accident escalation, and it reflects the proximity of the system to a particular adverse event (details in 3.2). This block identifies the hazards of interest and maps them into (a subset of) the state variables of the system. The hazard level depends on the system state variables, and this block provides the model and connection between these two analytical concepts. Examples of the mathematical equations that represent the mapping

[2] Gray boxes show the blocks developed in the companion papers [Favarò and Saleh, 2016a,b].

between system states and hazard levels are presented in Section 3.2. Multiple hazard levels (for various risks) are considered at a time, and they are reflected in the vector output H(t) of the Hazard Identification block in Figure 1.

ii. **Hazard level monitoring**: Once the hazard level metrics are defined, they are to be continuously monitored (either by the operator or through an automated process). Monitoring the values and trends of the hazard levels is an important step for the prioritization and triggering of safety interventions.

iii. **TL expression of safety properties** (not presented here): This step accounts for the translation of system safety requirements into TL formulae. These formulae in turn act as constraints on the system behavior. Each safety property is predicated on a particular hazard level function, making the monitoring of the hazard level a fundamental step for the definition and verification of each TL formula.

iv. **Verification of the TL safety properties** (not presented here): The TL safety properties are checked for compliance/violation. The violation of a safety constraint provides diagnostic information for re-engineering the system design, the safety barriers layout, and the system instrumentation, to mention a few possible types of safety interventions that can be triggered by this verification.

The end-objective of the safety supervisory block is to support decision-making, especially in relation to safety interventions on the system, and to improve accident prevention. The functions and blocks discussed in (i–iv) are some of the means for contributing to this end-objective. One particularly important tool in support of this end is shown in Figure 1 downstream the safety supervisory block and is entitled **Hazard temporal contingency map.** This block is both an analysis and visualization tool: it dynamically assesses and displays the "coordinates" of hazards in a system to support operators' sensemaking and help them prioritize attention and defensive resources for accident prevention. The coordinates of hazards include the hazard level or danger index (how hazardous a particular situation is) and the estimated time-to-accident (how much time is left before the accident associated with a particular hazard is released if no changes are made to the system operation). The hazard temporal contingency map provides prognostic information regarding the time-window available for operators to intervene before a hazardous situation becomes unrecoverable, and in so doing it helps prioritize risks and hazards based on their temporal contingency, not based on probability, or some combination of probability and consequence, as is traditionally done in PRA.

This safety supervisory and hazard temporal contingency blocks are not only helpful for system operators, they also affect various **stakeholders** involved in the safety value chain[4] of the system. The outer loop in Figure 1 closes back on the system by providing the hazard information (dynamics/trends) to different stakeholders, and prompts them to assess the need for and trigger **multi-tiered safety interventions**. These interventions can range from immediate actions (e.g., emergency shutdown, adjustment of safety barriers, or safety-related maintenance) to off-line re-engineering of safety features in the system design, the system instrumentation, or the operating procedures for example. These changes affect the **system model** block both in terms of the plant description (state space model) and of the controller definition and operations, thus closing the outer feedback loop in Figure 1. The model-based safety supervisory control and the hazard temporal contingency map support safety interventions over different time-scales and by different agents in the safety value chain. Finally as noted in the caption, Figure 1 is not meant to be exhaustive; several blocks and additional feedback loops are not displayed to avoid visual clutter (for example the "observer and state estimation block, and the feedback loops for monitoring the effectiveness of the safety interventions).

We will revisit some of these considerations in detail in the next subsections. Three key steps can be highlighted in the process here described and are analyzed next:

[4] The safety value chain consists of individuals or groups who contribute to accident prevention and sustainment of system safety. It includes operators, technicians, engineers, system designers, managers and executives, regulators, safety inspectors, and accident investigators, individuals who affect and contribute to system safety over different time-scales [Saleh et al., 2010].

1. **System model development**: development of a mathematical model for the dynamical system under consideration (or for a subset of the system with the safety implications of interest), with identification of state variables and state-space representation (Section 3.1);
2. **Safety supervisory monitoring**: identification of the hazard levels or danger indices of interest and state mapping, along with the monitoring of these indices (Section 3.2);
3. **Hazard temporal contingency analysis (and map)** to guide safety interventions: estimation of the time-to-accident metric and development of the hazard temporal contingency map, for ranking and prioritizing safety interventions (Section 3.3).

3.1 Model Development

The creation of a model for the dynamical system under consideration constitutes the first step of the approach. Similar to Hansen's [1998] work on the extension of temporal fault trees, the framework here proposed makes use of state variables (either continuous or discrete) denoting functions of time, as in Modern Control Theory. As highlighted in [Cowlagi and Saleh, 2013], although well-established safety strategies such as defense-in-depth "reflect an implicit recognition of accident prevention as a control problem" and several authors have articulated and developed this recognition more explicitly [Rasmussen, 1997; Leveson, 2004], actual control theory has been to a large extent absent from the discussion of safety as a control problem. In the following, we make use of some tools from the actual Control discipline, in particular with references to state-space representations of dynamical systems and state estimation. Given the need of a system model, our framework falls under the category of model-based safety analysis, which we briefly review next. Afterwards, we present the state-space formalisms and an example model.

3.1.1 Model-based safety analysis

A dynamical system is one whose properties (or a subset of them) change with time. A model for such system is defined as a set of equations that represent its behavior in time. Once an analytical model is developed, it can be translated/imported into a simulation environment for various types of analyses.

In simple terms, whenever a mathematical model is developed and employed for the analysis of the system under consideration (instead of carrying out experiments on the actual system), the approach is referred to as a model-based analysis. Specifically, model-based *safety* analysis has gained popularity over the past decade. It was first introduced to provide a more formal approach to analysis techniques that had traditionally been performed manually, with a low likelihood of being complete, consistent and error-free [Joshi and Heimdahl, 2005].

The main benefit of model-based analysis is the possibility of interfacing the system model(s) with automated analysis tools that can analyze the system behavior, allowing the verification of different aspects of fault tolerance and potentially the auto-generation of different outputs (e.g., fault trees) [Joshi and Heimdahl, 2005], and the repeatability of the analyses.

Many of the early efforts in model-based safety analysis were aimed at the auto-generation of PRA-types of analysis (see for instance [Papadopoulos et al., 2001]). The interest then expanded towards automated fault-detection and diagnosis [Isermann, 2005], and to the introduction of formal verification of the models to improve system reliability [Bozzano and Villafiorita, 2003; Bozzano et al., 2003]. The need for novel model-based techniques was justified by the increasing complexity of the systems under consideration, and by the need of safety engineers to assess the system behavior in degraded situations without the need to manually develop for example an extensive set of fault trees [Bozzano and Villafiorita, 2003].

Finally, a separate mention should be given to the work by Leveson et al. [2003, 2004] on the development of models for accident investigation and analysis. This approach departs from the previous methodologies that tried to leverage the automation of traditional PRA tools. The Systems Theoretic

Accident Modeling and Processes (STAMP) tool is aimed at better encompassing the role of organizational factors for complex socio-technical systems [Leveson et al., 2003]. STAMP, and the associated analysis tool STPA, views an accident as the breach or absence of suitable safety constraints. This view is referred to as a "safety control perspective", which is also common to our work. This shared trait relates to the notion of "hazardous state" that appears in both STPA (e.g., see [Ishimatsu et al., 2010]) and in our approach. At the same time, there are a number of important differences in relation to how this idea of "safety control" is conceived and implemented, to how the "hazardous states" that the system should stay clear of are quantified, and finally to the analytical tools and formalisms employed in the two approaches[5]. Additionally, our work tackles aspects that do not share similarities with STPA, such as the notion on temporal contingency, the adoption of Temporal Logic, and the verification of the safety constraints in both on-line and off-line contexts.

In the following, we build on ideas from model-based safety analysis, with the distinction that we employ them not in conjunction with PRA tools and techniques, but for our safety supervisory control approach, providing operators and other stakeholders with the means to monitor the system for hazardous conditions and scan for potential unfolding adverse events. The following approach leverages the state-space representation formalism, which we briefly review next.

3.1.2 State-space representation

The state-space representation is a mathematical formalism widely used in Modern Control Theory. It is concerned with three types of variables (all functions of time): *input* variables (denoted by the vector $\mathbf{u}(t)$), *output* variables (denoted by the vector $\mathbf{y}(t)$), and *state* variables (denoted by the vector $\mathbf{x}(t)$). Inputs and outputs are the means by which an external agent can interact with the system: the appropriate control actions are applied through the inputs to ensure the desired system behavior, which in turn is monitored through the output recording and state estimation [Bakolas and Saleh, 2011]. State variables (or simply the "state" of a system) are formally defined as the minimum set of variables that contain all the necessary information of the internal conditions of a system at some time t_0, such that the knowledge of the system state at time t_0 along with the knowledge of the input vector $\mathbf{u}(t)$ for $t \geq t_0$ is sufficient to determine all the system future outputs (for $t \geq t_0$) [Chen, 1995].

A dynamical system can then be represented in terms of its state-space representation through a system of first order differential equations, such as those of Eq. (1).

$$\begin{cases} \dot{\mathbf{x}}(t) = F(\mathbf{x}(t), \mathbf{u}(t)) & \text{state equation} \\ \mathbf{y}(t) = G(\mathbf{x}(t), \mathbf{u}(t)) & \text{output equation} \end{cases} \tag{1a}$$

F and G are generic functions (linear or non-linear) that relate on the one hand the rate of change of the state to the state itself and the input vector (state equation); and on the other hand the output vector to state vector and input vector (output equation). Equation (1a) holds for continuous systems, and can be easily generalized for discrete cases. For the case of linear systems, Eq. (1a) assumes the well-known form

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{1b}$$

where the matrices A, B, C, and D may also be dependent on time.

The role of state variables is central to our discussion, and it will enable the definition of a quantifiable metric for accident escalation, the hazard level function or danger index. The hazards of interest for the

[5] While STPA makes use of descriptive and qualitative control diagrams, our approach leverages a formal analytical model represented with the state-space formalism, which starts from the physics behind a process, and enables the set up of danger indices for the quantification of the "hazardous states".

system are mapped into (a subset of) the system state variables. A simple example is provided next to clarify some of these concepts and the application of the process described in Figure 1.

Figure 2 shows a schematic of a cylindrical oil tank, with an incoming mass flow $\dot{m}_{in}(t)$, and mass outflow $\dot{m}_{out}(t)$. Valves in the feeding and in the outflow line regulate the two mass flows. From the perspective of safety supervisory control, the role of the operator is to monitor the condition of the oil tank, and to apply control actions to steer the system away from dangerous situations should they develop. For instance, for a system such as that of Figure 2 we may want to ensure that: (i) a certain threshold height of oil inside the tower is never (b)reached or simply that the tower does not overflow; and (ii) that correct instrumentation and alarms are set up to inform the operator of potential problems or escalating hazard level in a timely manner. Violation of both these considerations let to the explosion at the Texas City refinery in 2005 in which 15 people were killed and 180 injured [Saleh et al., 2014].

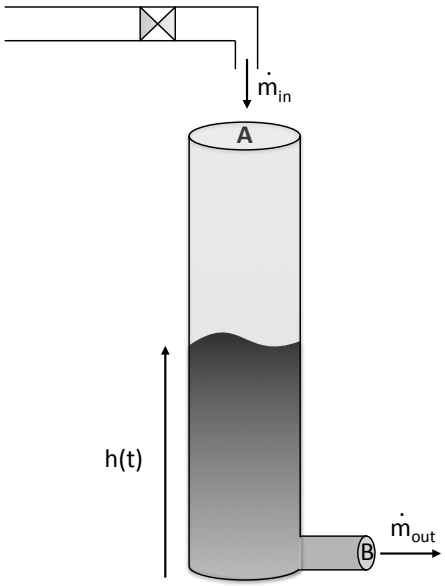

Figure 2. Schematic representation of an oil tank

The distinction between the two considerations, (i) and (ii), is subtle but important: the first requires the operator to monitor the height of the oil in the tower (h(t) in Figure 2). In our framework, we set up the oil height to be one of the system state variables, and then map it into a hazard level function. Monitoring the current value of the oil height against specific thresholds, and allowing the operators the use of appropriate control actions (e.g., regulating the incoming mass flow, or closing/opening the outflow line) provides the operator with the information needed to satisfy the first property. The second property is instead related to the notion of observability-in-depth and the ability to correctly diagnose the hazard level associated with the system. Properties of this type are carefully analyzed in [Favarò and Saleh, 2014, 2016b]).

We can now set up the model in the following way. For simplicity, we consider a one-dimensional problem, and pick the state to be the height of the oil inside the tower. Also, we assume two control inputs given by the incoming mass flow $\dot{m}_{in}(t)$, and the possibility to open up or close out the outflow line (hence zeroing out the outflow cross-section area, B in Figure 2). The output is given by the mass outflow $\dot{m}_{out}(t)$. We have:

$$\left.\begin{cases} x(t) \rightarrow h(t) \\ y(t) \rightarrow \dot{m}_{out}(t) \\ u_1(t) \rightarrow \dot{m}_{in}(t) \\ u_2(t) \rightarrow B(t) \end{cases}\right\} \rightarrow \mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \tag{2}$$

To set up a state-space model, we start by the mathematical model of the physics governing the system under consideration. The mass balance for the tank gives us

$$\frac{dV(t)}{dt} = \frac{1}{\rho}\left[\dot{m}_{in}(t) - \dot{m}_{out}(t)\right] \tag{3}$$

with V being the volume of oil filling the tank, and where we considered for simplicity the density of the oil to be a constant ρ. Given a constant cross-sectional area A for the tank, we have

$$\frac{dh(t)}{dt} = \frac{1}{A\rho}\left[\dot{m}_{in}(t) - \dot{m}_{out}(t)\right] \tag{4}$$

We can express the outflow as

$$\dot{m}_{out}(t) = B(t)\rho\sqrt{2gh(t)} \tag{5}$$

where $\sqrt{2gh(t)}$ represents the velocity of the fluid in the outflow pipe assuming a constant acceleration g for the incoming mass flow. We obtain then the differential equation governing the process:

$$\frac{dh(t)}{dt} = \frac{1}{A\rho}\left[\dot{m}_{in}(t) - B(t)\rho\sqrt{2gh(t)}\right] \tag{6}$$

It is now possible to obtain the non-linear state-space representation of the dynamical system of Figure 2 by considering the choice of states, outputs, and inputs provided in Eq. (2):

$$\begin{cases} \dot{x}(t) = \frac{1}{A\rho}\left[u_1(t) - u_2(t)\rho\sqrt{2g\,x(t)}\right] \\ y(t) = u_2(t)\rho\sqrt{2g\,x(t)} \end{cases} \tag{7}$$

The model of Eq. (7) is the basis for the application of the safety supervisory monitoring analysis that follows according to Figure 1. The state-space representation is a powerful tool for modeling a broad range of dynamical systems. The choice of which variables to select as states of the system is not unique, and in our case is dependent on and informed by the particular hazards to be monitored and safety constraints under consideration. In model-based approaches, the analytical expression of the model (such as that of Eq. (7)) is then imported into a simulation environment, and it enables a broad range of uses such as controller design, (hazard) monitoring, and diagnostic.

3.2 Safety supervisory monitoring

In this section we analyze the safety supervisory monitoring block of Figure 1, focusing on the following two steps: (i) the hazard level(s) identification and its mapping into the system state (section 3.2.1), and (ii) the execution of the monitoring process (section 3.2.2). Section 3.2.3 presents the application of the hazard monitoring process to a rejected takeoff scenario, to illustrate its use and capabilities.

3.2.1 Hazard level identification and state mapping

The hazard level, denoted by H(t), can be intuitively conceived of as the closeness of an accident to being released [Saleh et al., 2014]. Its definition provides an index to quantify "how dangerous" the current system state is, in terms of its proximity to an accident occurrence. In the following we use the terms hazard level and danger index interchangeably.

In order to define the function H(t), we first need to specify what accident we wish to monitor against. For instance, in the oil tank example presented in Figure 2, monitoring against the accident "loss of containment (LoC) through tower overflow" suggests that a suitable danger index maps the state of the system "oil height h(t)" against the maximum height we want to set as threshold. This is captured by

$$H_{LoC}(t) = \frac{h(t)}{h_{max}} \tag{8}$$

where the height of raffinate at time t is divided by the maximum achievable height before overflow occurs, so that the resulting hazard level is dimensionless. The situation H(t) = 1 indicates then overflow of the tower or the onset of the accident "loss of containment".

More generally, a series of adverse events that bring a system from its nominal operational conditions to off-nominal ones and finally to an accident occurrence can be reflected by the dynamics of the hazard level over time (an illustrative example is shown in Figure 3). The dynamics of the hazard level is not necessarily monotonic, and it can consist in a sequence of escalation, de-escalation, and constancy phases. Safety interventions are meant to block or de-escalate a hazardous situation (or its hazard level).

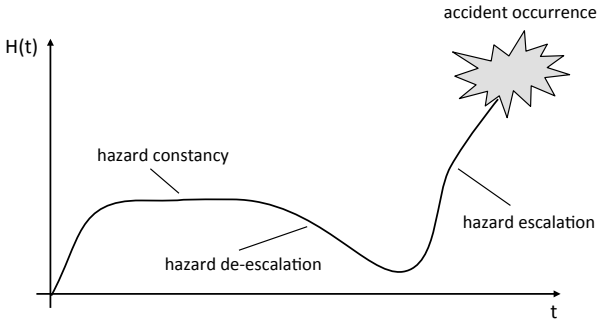

Figure 3. Illustration of hazard level dynamics

More complex danger indices can also be devised for the tower example in Figure 2, for instance by accounting for the velocity at which the tower is filling up, or by considering multiple states such as pressure (p(t)) and temperature (T(t)) of the oil. For example, one can set up a limit for maximum temperature inside the tower (T_{max}), where we take into account also the temperature change due to changing height and pressure of the oil inside the tower[6]:

$$H_T(t) = \frac{T(t)}{T_{max}}\left(1 + \frac{\alpha}{\rho c_p}\dot{p}(t)\,\frac{h(t)}{\dot{h}(t)}\right) \tag{9}$$

The idea of introducing a quantitative index for capturing the hazardousness of a situation is not novel. It is well established and particularly useful in the field of human-robot interaction [Ikuta et al., 2003; Kulić and Croft, 2005] where "danger indices" are devised based on the distance between the agents involved (e.g., patient and robot for human care) and the relative velocity to identify situations in which safety is compromised. For instance, in [Kulić and Croft, 2005] expressions of the following kind appear for the definition of danger levels:

"If (Distance = LOW) and (Velocity = HIGH) -> (DANGER = HIGH)"

Similarly to what we did in Eq. (8), Ikuta et al. [2003] proposed a danger index α based on the force of a potential impact, compared to a critical impact force, where the force is dependent on the velocity, the distance, the shape, and the mass of the agents involved:

$$\alpha = \frac{F(v,d,s,m)}{F_c} \tag{10}$$

These danger evaluation methods are aimed at establishing quantitative metrics to measure and control the hazardousness of a situation during system operation, and to minimize the danger involved in robot tasks. We propose here their extension beyond the specific field of human-robot interactions. These

[6] Equation (9) assumes an isentropic process. α is the volumetric thermal expansion coefficient, c_p is the constant pressure specific heat capacity.

methods are an important tool in support of accident prevention and for sustaining system safety. Regardless of the specifics of their definition, the **notion of quantifiable danger indices is a powerful one for both offline and online safety purposes**: offline these indices can be used within a simulation environment to assess for example how often the system approaches or breaches critical hazard levels and whether additional safety features ought to be embedded in the design of the system; online, **these danger indices add an important real-time dimension to the problem of risk assessment and hazard monitoring** (details in the next subsection). In both cases, quantifiable danger indices are an important piece in the view of safety as a control problem since accident prevention requires maintaining danger indices within safe bounds.

In our approach the definition of the hazard level is dependent on (a subset of) the state of the system. Equations such as (8), (9), and (10) can be generalized in the case of a N-dimensional state vector by the functional definition:

$$H(t) = f(x_1, x_2, \dots, x_N, t) \tag{11}$$

The estimation of the system state enables to measure the proximity to particular adverse events, an important step for accident prevention. Other authors have in the past advocated the need to include state variables dependencies in the notion of risk. For example, according to Haimes [2009] the reason why a universally agreed-upon definition of risk, a complex multidimensional concept, is still lacking is to be found in the missing understanding of some requisite ingredients, such as the state variables of the system. As the "performance capabilities of a system are a function of its state vector" [Haimes, 2009], then by the same token so is the safety or lack thereof and the hazardous condition of the system at any point in time. The notion of a danger index enables one to make explicit this (dynamic) risk dependence on the state vector of the system, and it becomes important to ensure the proper control of the system.

Based on the previous considerations, we propose that our model-based approach augments the system model shown in Eq. (1) with an additional *hazard equation*, which captures the dependency of the hazard level on (a subset of) the state vector **x**(t), and of the dynamics of the hazard level on the control variables **u**(t). For the case of linear systems we obtain:

$$\begin{cases} \dot{\mathrm{x}}(\mathrm{t}) = \mathrm{Ax}(\mathrm{t}) + \mathrm{Bu}(\mathrm{t}) & \text{state equation} \\ \mathrm{y}(\mathrm{t}) = \mathrm{Cx}(\mathrm{t}) + \mathrm{Du}(\mathrm{t}) & \text{output equation} \\ \text{----------} & \\ \dot{\mathrm{H}}(\mathrm{t}) = \Phi\mathrm{x}(\mathrm{t}) + \Psi\mathrm{u}(\mathrm{t}) & \text{hazard equation} \end{cases} \tag{12}$$

where the matrix Φ derives from the mapping of the hazard level into the state vector and the matrix Ψ embodies the dependence of the hazard level dynamics on the inputs vector. Adjusting the values of the matrix Ψ (whether done "manually" by the operator, or through an automated controller) results in different control actions on the hazard level. This process, called input shaping, is generally carried out in modern control theory to achieve specific performance goals. In our case, input shaping for the hazard equation allows to control the system and steer it away from hazardous states or dangerous conditions (e.g., de-escalate hazardous levels).

For many years, the guiding principle behind the control synthesis problem was that of output feedback (i.e., the observation of the system output). After the seminal works of Kalman [1960] and Bellman [1957], it became evident that the selection of control inputs is more efficient when based on the knowledge of the actual internal state of the system, rather than on its output [Bakolas and Saleh, 2011]. This consideration is reflected in our approach in the mapping of the state vector into the hazard level, and thus in the fundamental role of state estimation to capture the dynamics of the danger indices. The process of hazard monitoring is thus a form of state estimation, and it provides the proper feedback upon which to base control actions for safety interventions.

A final remark is worth noting. The hazard level provides an index of accident escalation, regardless of the sequence of events that leads to that particular accident. In other words, the hazard level spans every sequence and scenario of escalation that will lead to such an accident occurring. The choice of setting up a metric based on the system state (i.e., a proxy of its internal condition) allows to eliminate the path-dependency implicit in traditional PRA, where the computation of the conditional probabilities that lead to an accident occurrence has to account for the specific path followed by the system. In short, **danger indices are agnostic to the series of events that led to their particular value at any given instant of time, and as such they are independent on the specific accident trajectory followed by the system**. The set-up of danger indices for the system hence shifts the reliance of the risk assessment process from the identification of all possible accident trajectories and their associated probabilities to the identification of suitable hazard levels, whose choice is informed by the particular safety requirements imposed for the system.

3.2.2 Hazard level monitoring

The last step in the Safety Supervisory block in Figure 1 is the hazard level monitoring. To illustrate its role, consider the first requirement that was set up for the oil tank example, i.e., ensuring that the tower does not overflow. Intuitively, the implementation of this requirement in a quantifiable form implies the verification of the following constraint for the hazard level:

$$H(t) < H_A \tag{13}$$

where H_A represents the hazard level associated with the onset of the accident "loss of containment", thus $H_A = 1$ in our example of H(t) provided in Eq. (8). Properties such as that expressed in Eq. (13) allow the set up of safety bounds (or safety envelopes for higher dimensions than 1D) and criticality thresholds for the hazard level. Safety margins can also be accounted for in the definition of the threshold values, so that in general we require $H(t) < H_{crit}$, for a pre-defined H_{crit} criticality threshold.

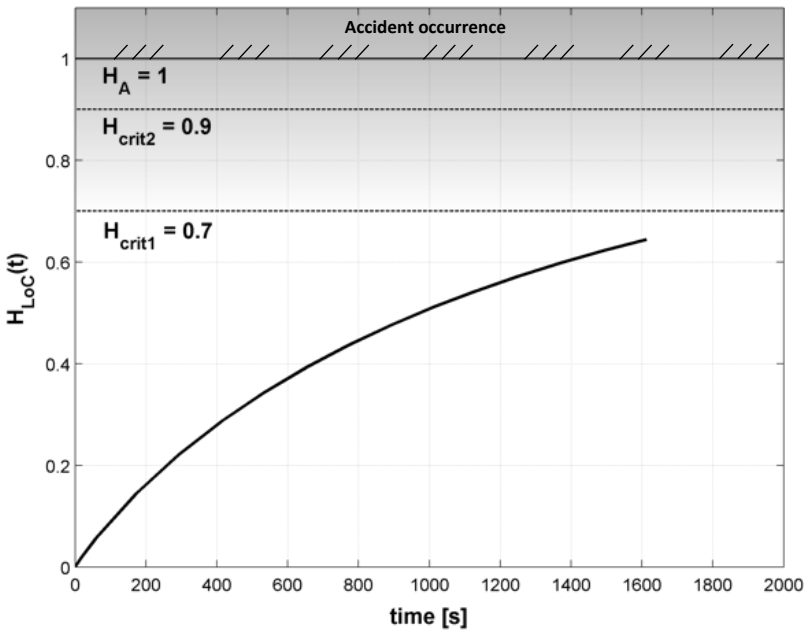

Figure 4. Hazard level dynamics for the oil tank example and comparison with criticality thresholds

Monitoring of the hazard level informs the operators of developing dangerous situations, and thus supports their situational awareness by capturing the specific hazard dynamics and escalation (and the particular accident the system is approaching). As an illustration, Figure 4 provides an example of hazard level dynamics for the oil tank example, compared in this case to three criticality thresholds: one

corresponding to 70% of the tower filled up (H_{crit1}), one corresponding to 90% of the tower filled up (H_{crit2}), and one corresponding to actual overflow conditions (H_A). In the case of Figure 4, the hazard level value is obtained by computing the height of oil inside the tower through direct integration of Eq. (7), using as input values a constant incoming mass flow of 35 kg/s and considering a partially closed outflow line.

Plots such as the one of Figure 4 can serve as a diagnostic tool to inform *on-line* safety interventions. For instance, in this case a value of H(t) too close to a critical threshold H_{crit}, and a sustained positive slope for H(t), suggests to the operator that a safety intervention is warranted—at a minimum to block the dynamics of hazard escalation through emergency shutdown for example, or fully open the outflow line to de-escalate the hazardousness of the situation and decrease the height of the oil in the tower away from the critical thresholds. This corresponds (from a control/mathematical perspective) to adjusting the values of the control matrix Ψ in the hazard equation. By comparing the current value of H(t) to the criticality thresholds, the operator is also able to get a real-time estimate of the time when the thresholds will be (b)reached, as we examine in the next section.

When the hazard level monitoring is executed *off-line*, a detailed analysis of the history of hazard dynamics can help answer important questions regarding on the one hand, the occurrence and ranking of near misses (frequency and severity or hazardousness—how close the situation got to critical thresholds), and on the other hand, the identification of missing or ineffective safety features, that allowed the increase in the hazard level, including inadequate operator training. Although the following topic is tangential to our purposes, we believe the connection between the proposed safety supervisory control and model-based hazard monitoring on the one hand, and near miss management systems on the other hand [Gnoni and Lettera, 2012; Gnoni et al., 2013] offer many possibilities for meaningful contributions and is a rich area for further research and investigation.

To further illustrate the capabilities and insight that can be derived from the hazard monitoring process, we provide next an application of the presented tools in support of the "go/no-go" decision-making in rejected takeoff situations (RTO).

3.2.3 Example application of hazard level monitoring

Traditionally, the thinking about the problem of setting regulations and policies for rejecting a takeoff has revolved around the notion of the decision speed V_1. Pilots are advised against rejecting a takeoff after the decision speed V_1 is achieved unless they have reason to believe "the aircraft cannot be safely airborne" [ECAST, 2016].

Statistics show that there is more to the "go/no-go" decision than the simple "stop before V1" and "go after V1" strategy [TSTA, 2016]. The fact that the V_1 limit is not sufficient in of itself is recognized by both air manufacturers and regulators, who advocate new metrics to expand on the current thinking about these issues. For instance [Airbus, 2005] shows that about 54% of runway excursions occur when RTOs are initiated at speed above V1, but also highlight that about 26% of them occur for RTOs initiated below V1.

The set up of hazard levels and criticality thresholds can support pilots in their decision to reject the takeoff versus "take the problem into the air" strategies. Consider for instance the hazard level defined in Eq. (14).

$$\mathrm{H(t)} = \frac{d_{STOP}(t)}{l_{run}+d_{RESA}-d(t)} \tag{14}$$

This hazard level quantifies and relates the distance required for the aircraft to come to a stop (once a RTO is initiated) to the total length available to the aircraft before encountering an obstacle on its path. This length is computed as the runway length still available (given by the runway length l_{run} minus the

distance already traveled d(t)) plus the runway end safety area (d_{RESA})[7]. Rather than defining the accident as a simple runway overrun, this danger index identifies the accident as that condition for which the stopping distance required would bring the aircraft beyond the limit of the RESA. In other words, the situation H(t) = 1 would thus identify either a collision with an obstacle and/or the encounter of highly uneven terrain.

The calculation of the stopping distance $d_{STOP}(t)$ depends on several factors, such as the velocity at which the RTO is initiated, the position of the aircraft along the runway, the conditions of the runway (e.g., wet, dry,...), and the availability of the brakes and thrust reversers among other things. In order to compute such distance, it is necessary to set up a model for the aircraft dynamics during the RTO. For simplicity, we will only consider the longitudinal motion of the aircraft along the runway, and make some simplifications for the aerodynamic coefficients of interest. The governing equation is provided by

$$m\frac{d^2x}{dt^2} = T - D - \mu_r(W - L) \tag{15}$$

m is the vehicle mass; T the thrust provided by the engine(s); D the drag, and it is dependent on the aircraft configuration (e.g., with flaps and slats deployed) and the velocity $\frac{dx}{dt}$; μ_r is the rolling friction coefficient (and for the RTO case its increase models the brakes application); W is the aircraft weight; L the lift. Equation (15) can be translated in the state-space representation formalisms as follows:

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = \frac{T(t) - D(x_2(t)) - \mu_r(t)(W - L(x_2(t)))}{m} \end{cases} \tag{16}$$

where the first state x_1 represents the distance traveled along the runway (x-axis), and the second state x_2 the instantaneous velocity of the aircraft. The following plots are obtained applying the model to the data of a Learjet 60. We consider here that full braking power and thrust reversers are available, and the runway is dry.

The model of Eq. (16) is integrated to compute distance, velocity, and acceleration of the aircraft at any point in time. Specifically for the RTO, when brakes and thrust reversers are applied, the stopping distance is computed as the distance corresponding to a zero velocity. This procedure can be repeated for a range of different initial conditions, i.e., for a range of different velocities v_0 and positions along the runway d_0 at which the RTO is initiated. Plotting the hazard level as a function of these initial conditions (which we normalize for convenience with respect to V_1 and to the runway length) yields plots such as the one of Figure 5, where two criticality thresholds are highlighted. The first threshold represents situations in which the aircraft comes to a stop within a 15% safety margin from the end of the RESA, while the second threshold corresponds to the accident unfolding.

[7] The runway end safety area (RESA) accounts for an additional region beyond the end of the runway before sudden changes in the terrain gradient and/or obstacles are encountered.

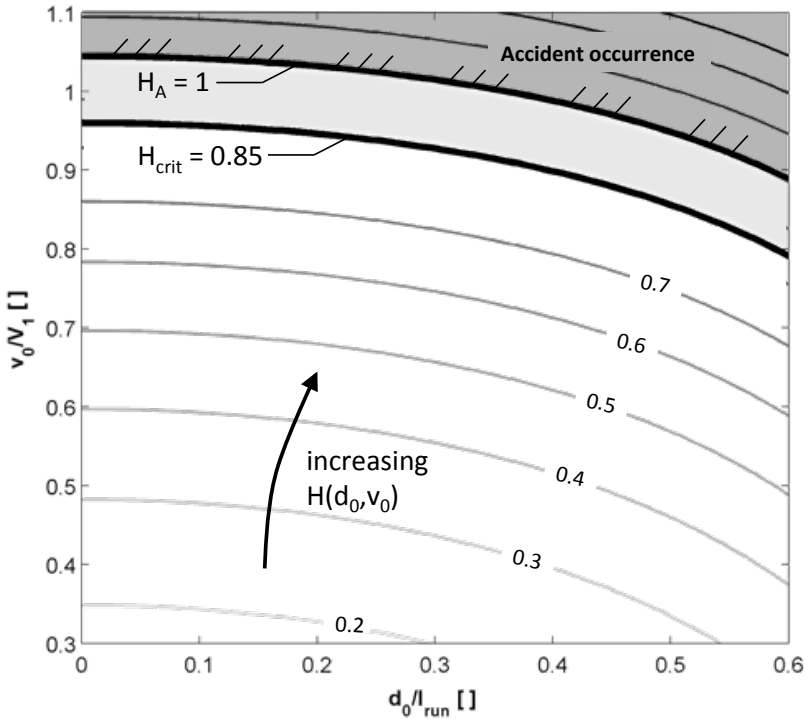

Figure 5. Contours of the hazard level of Eq. (14) plotted as a function of the initial conditions for the RTO

The accident threshold $H_A = 1$ can be compared to the traditional limit imposed on the decision speed V_1 (Figure 6).

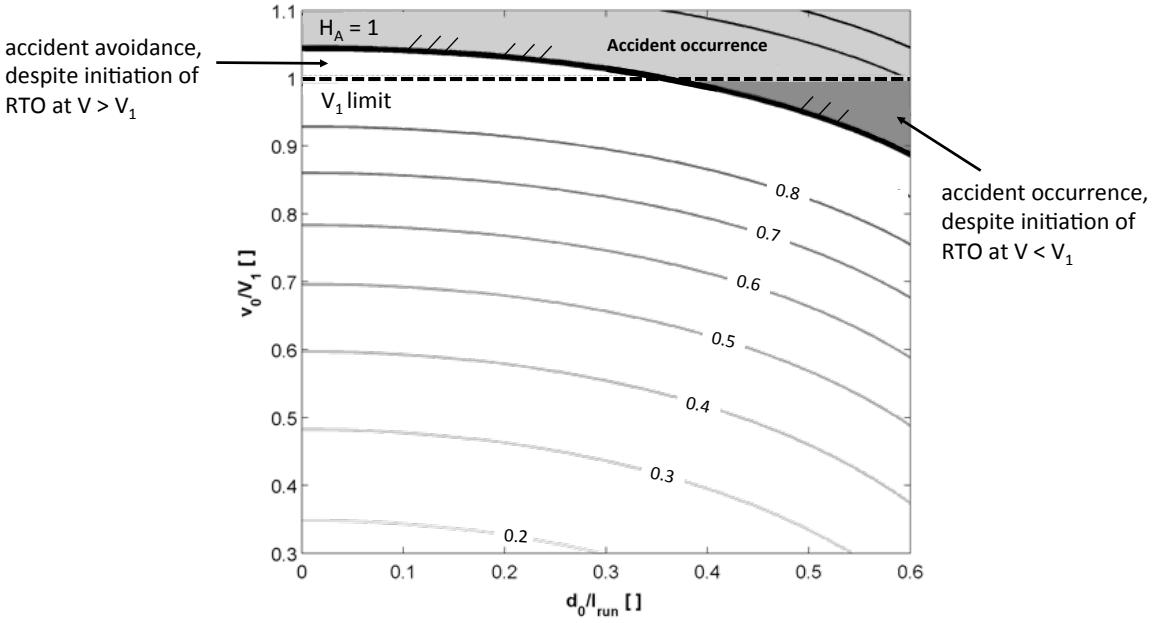

Figure 6. Contours of the hazard level of Eq. (14) and comparison with V_1 limit

Figure 6 provides a clear visualization of how the metric established by Eq. (14) informs traditional approaches for the RTO decision-making problem. Specifically, by accounting for the stopping distance dependence on the state of the system when the RTO is initiated (in terms of velocity and position), **the selected hazard level can account for both situations in which RTOs are initiated below V_1 and still result in an accident, and situations for which RTOs are initiated above V_1 that do not**.

Finally, Figure 7 superimposes a typical aircraft trajectory during takeoff to the mapping of Figure 6. It can be seen that for this particular scenario (dry runway and thrust reversers deployed), the trajectory briefly enters the new "danger area" highlighted in Figure 6. More so will be in the case when full braking power is not available, or the runway conditions are less than ideal. As the possibility of an RTO should always be considered by the pilots before the initiation of takeoff procedures, a situation such as the one of Figure 7 can advise the pilots to reconsider the suitability of that particular runway and/or make sure that the entire available length of the runway is exploited (e.g., not starting the takeoff from an intersection with a taxiway).

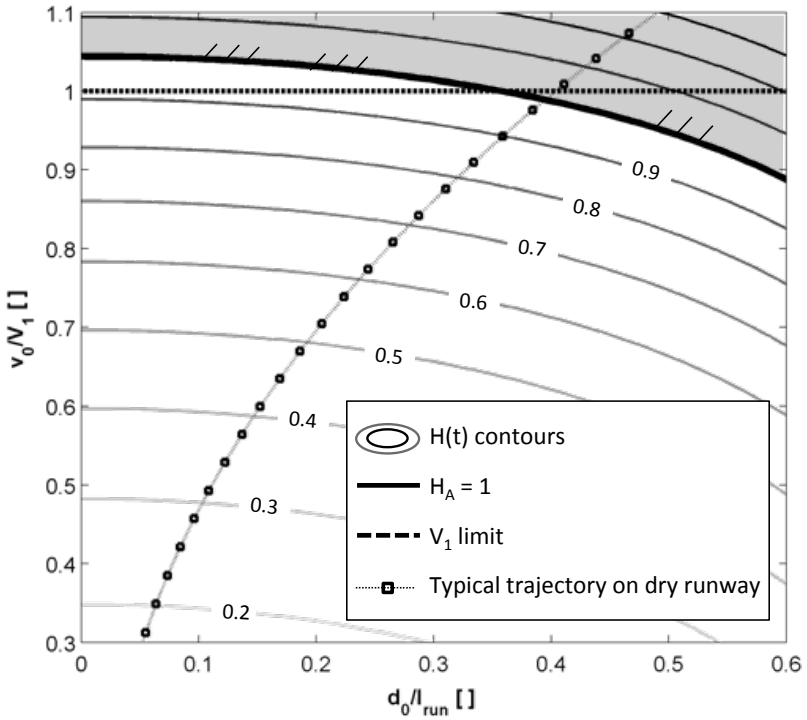

Figure 7. Comparison of "danger areas" and typical aircraft takeoff trajectory for best-case scenario

Metrics and diagnostic tools such as the one here considered can also be used by regulators and policy makers to inform safety guidelines, and at the same time they can be applied on-line to support real-time decision-making in critical situations (in this case, some interesting avionics development and user interface/displays can also be devised to support this approach). In short, the safety supervisory monitoring process presented in this section offers many advantages that complement the traditional approaches to risk assessment. Other than the diagnostic information presented in this section, the monitoring of the hazard level within a model-based approach supports a *prognostic dimension* as well, which we introduce next.

3.3 Hazard temporal contingency analysis (and map)

This step is shown downstream of the safety supervisory monitoring block in Figure 1. It is shown as a separate entity to highlight its importance. The development of a hazard equation (Eq. 12), which is enabled by the adoption of a model-based approach, allows one to estimate the time at which critical thresholds for the hazard level are (b)reached. This estimation process provides prognostic information and produces a proxy for a time-to-accident metric or advance notice for an impending adverse event. This temporal metric[8] can also be construed as providing an estimate for the time-window available for

[8] The time-to-accident metric can be described as a random variable. One objective of a dynamic risk assessment and accident prevention is to monitor and control the set of such metrics in a system, and keep them at a safe temporal distance away from 0.

safety interventions, assuming no changes are made to the system operation/inputs. This helps with the identification of the temporal criticality of different hazards on the one hand, and the prioritization of attention and defensive resources for hazards that warrant more timely intervention on the other hand.

To illustrate this estimation process, consider one more time the oil tank example. Given the current value of the hazard level at time t_e (the time at which the estimation will take place), the remaining time before the LoC accident occurs, assuming no change of inputs, can be derived using various estimators, the simplest one is expressed as follows:

$$\widehat{\Delta T_{LoC}}(t_e) = \frac{h_{max} - h(t_e)}{\dot{h}(t_e)} = \frac{1 - H_{LoC}(t_e)}{\dot{H}_{LoC}(t_e)} \quad (17)$$

The knowledge of these two "coordinates" of a hazard, $H_{LoC}(t_e)$ and $\Delta T_{LoC}(t_e)$, provides an important feedback for operators and decision-makers to dynamically monitor and actively manage the hazard of loss of containment in real time. Furthermore, when other potential accidents are identified and their associated hazard coordinates are estimated, the result is a portfolio of hazard coordinates, which roughly translates into "how hazardous is a particular situation" and "how much time is left before their corresponding accident occur". This collective information can then be displayed dynamically in a **hazard temporal contingency map** (Fig. 8) to support operators' sensemaking and help them prioritize attention and defensive resources for safety interventions and accident prevention[9].

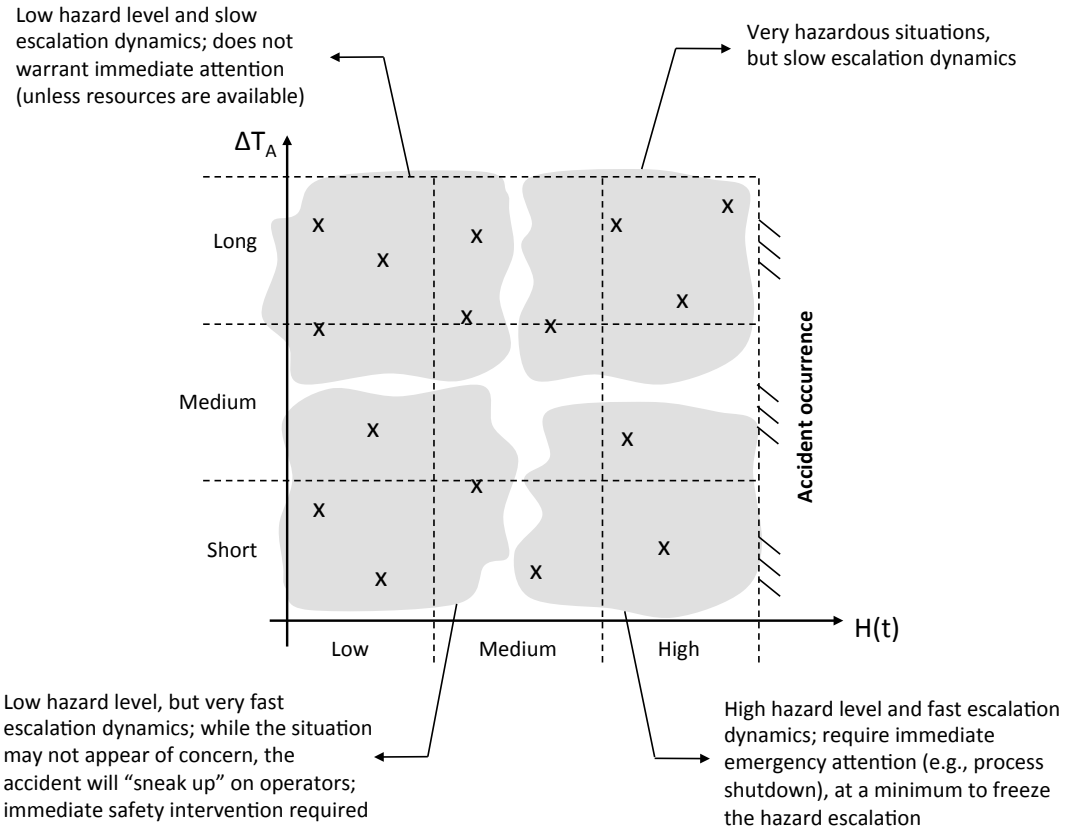

Figure 8. Illustrative hazard temporal contingency map

Figure 9 provides a graphical illustration of how the estimate for the time-to-accident ΔT_{Ai} can be achieved in practice (shown here for two accidents A_1 and A_2). The plots represent the evolution in time of the quantity $H_{Ai} - H(t)$, which reflects how far the current hazard level is from the level associated

[9] Trends over time and uncertainty bars in the estimates of both hazard coordinates can also be assessed and displayed.

with each accident (normalized at 1 for simplicity and consistency with the similar feature discussed in the previous examples). The two panels in Figure 9 show the situation at two instants of time. The top and the bottom plots relate to two different hazard indices $H_1(t)$ and $H_2(t)$. At the beginning of the monitoring period (left panel), both indices indicate no hazardous condition developing ($\Delta T_{Ai} \rightarrow \infty$). At time t_2, both hazard levels $H_1(t)$ and $H_2(t)$ escalate, the former faster than the latter (right panel). In this situation a simple estimation of the time to accidents for both indices informs the operators which sequence deserves more timely attention or immediate intervention ($H_1(t)$ in this case). The time-window available for safety interventions can be simply estimated according to Eq. (18):

$$\widehat{\Delta T_A}(t_2) = t_A - t_2 = \frac{H_A - H(t_2)}{\dot{H}(t_2)} \tag{18}$$

More elaborate estimators can be devised to account for the persistency of increase in H(t) as well as its slope and other dynamic features. Furthermore when the estimate is conducted repeatedly over time, a probability density function of ΔT_{Ai} can be obtained, thus reflecting the true nature of this time-to-accident metric as a random variable. Several uses can be made of this random variable and its features to inform safety-related decision-making, for example the shrinking of its standard deviation would reflect an increasing certainty of an impending accident (should business-as-usual in the operation of the system be maintained, or no safety intervention triggered). These issues are left as fruitful venues for future work.

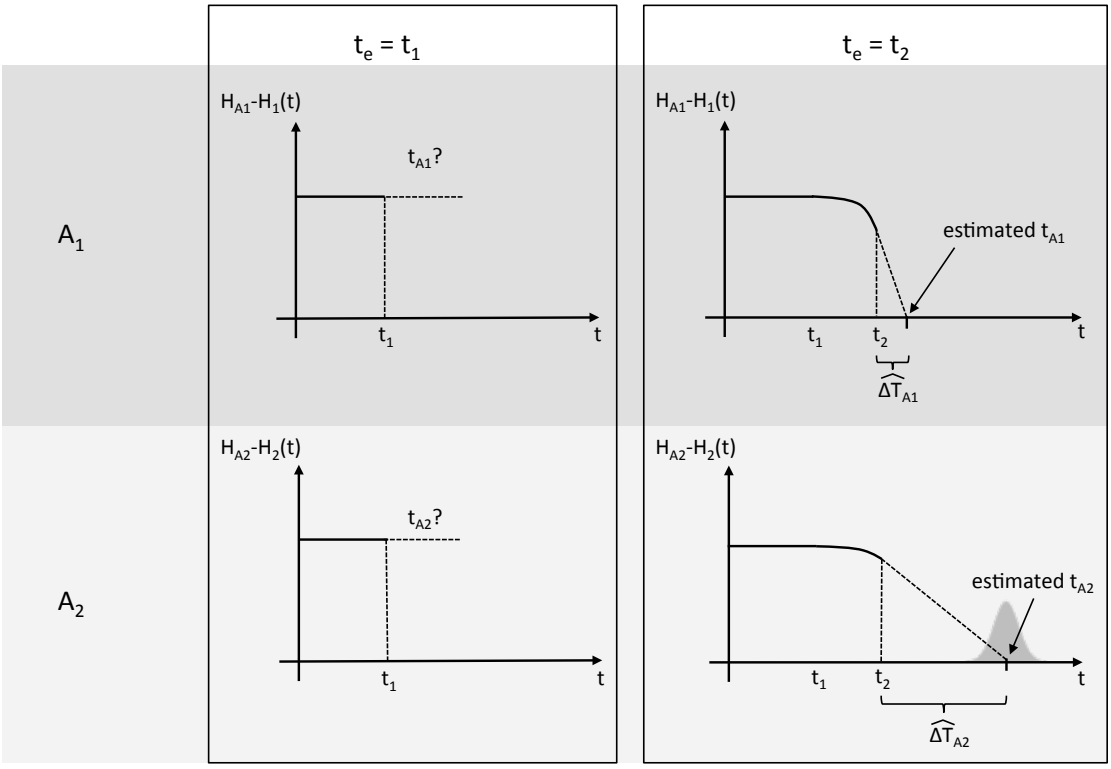

Figure 9. Estimation of the time-to-accident for two hazard indices

The considerations presented in this section also allow for the identification of areas were additional response capacity is required in order to improve the response time to emerging hazards and/or other mitigating actions, as advocated in [Mosleh, 2014]. The effect of safety interventions directly translates into decreases in the hazard level, and hence new estimations of the time-to-accident metric (i.e., extension of ΔT_A). Scenario-based testing can ensure that the safety features included in the system provide the operator with enough time to either trigger a safety intervention and abate the hazard level

(i.e., block an accident from unfolding), or to mitigate its consequences should it occur (e.g., with a timely evacuation).

3.4. Safety Monitoring and Fault Detection: off-line and on-line applications

The previous sections provided an overview of the proposed framework, which we termed safety supervisory control. The problem of safety monitoring and the subsequent control of off-nominal situations is in the safety literature tied to that of fault diagnosis and detection [Venkatasubramanian et al., 2003a]. The approaches to fault diagnosis can be divided into two big families: those that are aimed at detecting problems in off-line contexts, and those that do so on-line (or in real-time) [Kelly and Bartlett, 2006]. Among off-line approaches, it is possible to recognize sequential testing of system (during development stages) and the classical approaches of PRA (e.g., FMEA, FTA, and HAZOP techniques). On-line approaches, conversely, deal with the real-time monitoring of the system during operations. These approaches require fast analysis techniques, given the importance of providing the fault information to the system operators as quickly as possible. For example, qualitative trends analysis of the system output and noise control of sensor collected data figure among the possible approaches[10]. Another important distinction between off-line and on-line approaches exists, and it is related to the use they make of the notion of "system state". The importance of state estimation is recognized in on-line fault detection [Chen and Modarres, 1992], and adequate observability of the system state is a precondition for effective monitoring [Dheedan and Papadopoulos, 2010].

The framework and tools introduced in this work share some traits with online monitoring, but they also extend beyond it. Section 3.1 highlighted the role of a second ingredient in our work (Temporal Logic) towards off-line applications of the proposed framework. As a matter of fact, the proposed framework is capable of providing a set of diagnostic information that can also be employed off-line. The capabilities of off-line applications are further explored in the companion papers [Favarò and Saleh, 2016a,b], where a set of constraints for the system, predicated on the hazard levels, are expressed in temporal logic and can be verified for satisfaction/compliance in both on-line and off-line contexts (e.g., during the development stages of the system). Broadly speaking, the mapping in Figure 8 summarizes the two dimensions of prognostics and diagnostic feedback that the framework is capable of providing. While the prognostic information is fundamental for the framework on-line use, the diagnostic information that stems from the evaluation of the hazard level in testing scenarios as wells as from the verification of the safety constraints in temporal logic can serve to guide off-line safety interventions for system re-engineering and design modifications.

A final important distinction between the framework here proposed and other approaches to fault diagnosis relates to the novel dimension of temporal contingency introduced in Section 3.3. The traditional problem of fault detection and administration requires the user/analyst to pre-define a set of ratings for potential faults that are related to the probabilities and likelihoods associated to the identified faults [Chen and Modarres, 1992; Hu et al., 2003]. These rating are then employed for prioritization of safety interventions. In our framework on-line safety interventions are prioritized based on a real-time and not pre-defined schedule, evaluated according to the proximity of each danger index to the criticality thresholds of interest. Maruya et al. [2010] recognized that "modern plants are data rich and information poor". We believe that the addition of the notion of temporal contingency and of a hazard equation to the set of equations that represent and model the system behavior provide a useful contribution towards better-informed safety interventions.

4. Conclusions

The end-objective of our work was to contribute to improving (dynamic) risk assessment and accident prevention. To this effect, we first provided a synthesis of key limitations of PRA and the improvements

[10] For a detailed survey of fault detection methodologies see [Venkatasubramanian et al., 2003 a,b,c; Ng et al., 2010; Hashemian, 2011].

currently proposed in the literature, since these issues constitute the motivation for our efforts. We then made the case for model-based approaches and the use state variables, in particular in relation to the development of danger indices and the monitoring of hazard dynamics for improved risk assessment. We introduced a novel safety supervisory control framework, developed its analytical tools, and presented the notion of hazard temporal contingency for dynamic risk assessment and for guiding safety interventions to improve accident prevention.

The framework and analytical tools we developed were grounded in Control Theory and made use of the state-space representation in modeling dynamical systems. The use of state variables allowed the definition of hazard levels or danger indices, which measured the "proximity" of the system to adverse events. Furthermore, we showed that the adoption of state-space formalism enables the estimation of the times at which critical thresholds for the hazard level are (b)reached. This estimation process we argued provides important prognostic information and produces a proxy for a time-to-accident metric or advance notice for an impending adverse event. These hazard coordinates were displayed in a hazard temporal contingency map to support operators' situational awareness, and help them prioritize attention and defensive resources for accident prevention. The monitoring of hazard levels and the estimation of the time window available for safety interventions provide important feedback for various stakeholders and decision-makers to guide safety interventions both on-line (towards accident prevention and/or mitigation) and off-line (towards re-design and re-engineering of safer systems).

The work here presented sought to augment the current perspective in traditional risk assessment and its reliance on probabilities as the fundamental modeling ingredient with the notion of temporal contingency, a novel dimension by which hazards are dynamically prioritized and ranked based on the temporal vicinity of their associated accident(s) to being released.

Several new challenges are raised with this new approach. For example, more reliance is placed on the analysts who develop the model of the system and identify the hazard levels of interest (i.e., high level of modeling expertise is required, as well as in-depth knowledge of the system). Note that the choice of the H(t) functions of interest can be informed by the particular safety requirements imposed for the system. Another challenge for the practical implementation of any model-based approach to system design and operation is related to the proliferation of the number of states to consider (known as the state explosion problem). This problem requires careful consideration of model order reduction and computational implementation (especially for real-time hazard monitoring and estimations). A set of challenges are raised in relation to the verification and validation of such analytics, as well as human factors considerations in using/interfacing with our proposed safety supervisory control approach.

One final challenge with our proposed model-based approach to safety is worth pointing, that of its scope and scalability. The examples used previously were confined to a subsystem (oil tank) or system in a particular operational phase (aircraft during takeoff). As one reviewer rightly pointed out on a previous version of this work and questioned: how practical is this approach to be "applied at the level of a complex system… or a system of systems"? In general, model-based approaches do not easily scale up to large complex systems, in particular because of the "state explosion" phenomenon and the resulting modeling and computational complexities [Dheedan and Papadopoulos, 2010]. While this will affect the scalability of our approach, several of its redeeming features can help it cope with larger systems, to some extent similar to what traditional PRA and DPRA typically handle in the following manner. First our approach is hazard centric, which implies that state-space models need to be developed only for the hazards of interests, not the entire system, or for a subset of the system with the safety implications of interest[11]. As such, only a subset of the system states whose monitoring is required for the evaluation of the hazards of interests are needed. If other parts or subsystems are needed, they can be modeled with different techniques (e.g., transfer functions for linear components, state charts for digital/software components), which can then be integrated within the same simulation environment (e.g., Simulink). Second, the use of Temporal Logic (intrinsic to our framework, but

[11] This is similar to the fact that different fault tress have to be developed for different top events, or different event trees for different initiating events.

examined in detail in a companion article) can significantly reduce the number of state variables needed to express a hazard level and capture the dynamics of hazard escalation (without requiring an extensive state space model of the entire system). Despite these two redeeming features, we recognize that scalability of our model-based approach is likely to remain a (manageable) challenge.

However, we believe the prospects and potential advantages offered by the framework and tools here introduced outweigh the challenges they raise, and they constitute a rich area for further development. Several paths forward are possible and some were outlined throughout this text. Some authors have recently argued for the need to leverage automation for risk assessment and management; this model-based approach provides one step in this direction. We hope our work will expand the basis of risk assessment beyond its reliance on probabilistic tools, that it enriches the intellectual toolkit of risk and safety researchers, and that it invites further contributions from the community to improve (dynamic) risk assessment and accident prevention.

References:

Airbus (2005). *Flight Operations Briefing Notes* – Supp. Tech., Rev. 2 – May 2005. Available at http://www.airbus.com/fileadmin/media_gallery/files/safety_library_items/AirbusSafetyLib_-FLT_OPS-SUPP_TECH-SEQ02.pdf , accessed on 01/21/2016.

Aldemir, T., Guarro, S., Mandelli, D., Kirschenbaum, J., Mangan, L. A., Bucci, P., ... & Arndt, S. A. (2010). *Probabilistic risk assessment modeling of digital instrumentation and control systems using two dynamic methodologies.* Reliability Engineering & System Safety, 95(10), 1011-1039.

Aldemir, T. (2013). *A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants*. Annals of Nuclear Energy, 52, 113-124.

Apostolakis, G. E. (2004). *How useful is quantitative risk assessment?* Risk Analysis, 24(3), 515-520.

Bakolas, E., and Saleh, J. H. (2011). *Augmenting defense-in-depth with the concepts of observability and diagnosability from control theory and discrete event systems*. Reliability Engineering & System Safety, 96(1), 184-193.

Bellman, R. (1956). *Dynamic programming and Lagrange multipliers*. Proceedings of the National Academy of Sciences of the United States of America, 42(10), 767.

Bouissou, M. and Bon, J.L., (2003). *A new formalism that combines advantages of fault-trees and Markov models: Boolean logic driven Markov processes*. Reliability Engineering & System Safety, 82(2), pp.149-163.

Bozzano, M., and Villafiorita, A. (2003). *Improving system reliability via model checking: The FSAP/NuSMV-SA safety analysis platform*. In: Computer Safety, Reliability, and Security (pp. 49-62). Springer Berlin Heidelberg.

Bozzano, M., Villafiorita, A., Åkerlund, O., Bieber, P., Bougnol, C., Böde, E., ... and Zacco, G. (2003). *ESACS: an integrated methodology for design and safety analysis of complex systems*. In Proceedings of ESREL 2003 Conference, Maastricht, The Netherlands, 15-18 June 2003.

Chen, C. T. (1995). *Linear system theory and design*. Oxford University Press, Oxford, UK.

Chen, L.W. and Modarres, M., (1992). *Hierarchical decision process for fault administration*. Computers & chemical engineering, 16(5), pp.425-448.

Cowlagi, R. V., and Saleh, J. H. (2013). *Coordinability and consistency in accident causation and prevention: formal system theoretic concepts for safety in multilevel systems*. Risk analysis, 33(3), 420-433.

Dheedan, A. and Papadopoulos, Y., (2010). *Multi-agent safety monitor*. In proceeding of Intelligent Manufacturing Systems (IMS): 10th IFAC Workshop on Intelligent Manufacturing Systems, Lisboa, Portugal.

DOD – Department of Defense Standard Practice System Safety. (2012). *MIL – STD – 882E*. Available at http://www.system-safety.org/Documents/MIL-STD-882E.pdf , accessed on 06/17/2015.

ECAST – European Commercial Aviation Safety Team (2016). *Runway Excursion Preventions*. Available at http://easa.europa.eu/essi/ecast/main-page-2/runway-safety/ , accessed on 01/21/2016.

Favarò, F. M., Jackson, D. W., Saleh, J. H., & Mavris, D. N. (2013). *Software contributions to aircraft adverse events: Case studies and analyses of recurrent accident patterns and failure mechanisms*. Reliability Engineering & System Safety, 113, 131-142.

Favarò, F. M., and Saleh, J. H. (2013). *Observability in Depth: novel safety strategy to complement defense-in-depth for dynamic real-time allocation of defensive resources*. In Proceeding of ESREL 2013 Conference, Amsterdam, The Netherlands, 29 September – 2 October 2013.

Favarò, F. M., and Saleh, J. H. (2014). *Observability-in-depth: an essential complement to the defense-in-depth safety strategy in the nuclear industry*. Nuclear Engineering and Technology, 46(6), 803-816.

Favarò, F. M., and Saleh, J. H. (2016a). *Temporal Logic for System Safety Properties and Hazard Monitoring*. Submitted to the Journal of Loss Prevention in the Process Industries, January 2016.

Favarò, F. M., and Saleh, J. H. (2016b). *An Application for Safety Supervisory Control and Model-based Hazard Monitoring for Risk-informed Safety Interventions using Temporal Logic*. Submitted to Reliability Engineering and System Safety, January 2016.

Gnoni, M.G. and Lettera, G., (2012). *Near-miss management systems: A methodological comparison.* Journal of Loss Prevention in the Process Industries, 25(3), pp.609-616.

Gnoni, M.G., Andriulo, S., Maggio, G. and Nardone, P., (2013). *"Lean occupational" safety: An application for a Near-miss Management System design.* Safety science, 53, pp.96-104.

Groen, F. J., Smidts, C. S., Mosleh, A., & Swaminathan, S. (2002). *Qras-the quantitative risk assessment system.* In Proceedings of RAMS 2002 Conference, Seattle, WA, USA, 28-31 January 2002.

Haimes, Y. Y. (2009). *On the Complex Definition of Risk: A Systems Based Approach*. Risk analysis, 29(12), 1647-1654.

Hansen, K. M., Ravn, A. P., & Stavridou, V. (1998). *From safety analysis to software requirements*. Software Engineering, IEEE Transactions on, 24(7), 573-584.

Hashemian, H.M., (2011). *On-line monitoring applications in nuclear power plants*. Progress in Nuclear Energy, 53(2), pp.167-181.

Hu, W., Starr, A.G. and Leung, A.Y.T., (2003). *Operational fault diagnosis of manufacturing systems*. Journal of Materials Processing Technology, 133(1), pp.108-117.

Ikuta, K., Ishii, H. and Nokata, M. (2003). *Safety evaluation method of design and control for human-care robots*. The International Journal of Robotics Research, 22(5), 281-297.

Isermann, R. (2005). *Model-based fault-detection and diagnosis–status and applications*. Annual Reviews in control, 29(1), 71-85.

Ishimatsu, T., Leveson, N.G., Thomas, J., Katahira, M., Miyamoto, Y. and Nakao, H., (2010). *Modeling and hazard analysis using STPA*. Conference of the International Association for the Advancement of Space Safety 2010. Huntsville, Alabama.

Jahanian, F., and Mok, A. K. L. (1986). *Safety analysis of timing properties in real-time systems*. Software Engineering, IEEE Transactions on, (9), 890-904.

Jahanian, F., and Mok, A. K. (1994). *Modechart: A specification language for real-time systems*. Software Engineering, IEEE Transactions on, 20(12), 933-947.

Johnson, C. W. (1995). *Decision theory and safety-critical interfaces*. In Proceeding of Interact 1995, Lillehammer, Norway, 25-29 June 1995.

Johnson, C. W. (2000). *Proving properties of accidents.* Reliability Engineering & System Safety, 67(2), 175-191.

Johnson, C. W., & Harrison, M. D. (1992). *Using temporal logic to support the specification and prototyping of interactive control systems*. International Journal of Man-Machine Studies, 37(3), 357-385.

Kalman, R.E (1960). *Contributions to the theory of optimal control*. Boletin Sociedad Matematica Mexicana, Vol. 5, 102–119.

Kelly, E.M. and Bartlett, L.M., (2006). *Application of the digraph method in system fault diagnostics*. In Availability, Reliability and Security, 2006. ARES 2006. The First International Conference on (pp. 8-pp). IEEE

Kirschenbaum, J., Bucci, P., Stovsky, M., Mandelli, D., Aldemir, T., Yau, M., Guarro, S., Ekici, E. & Arndt, S. A. (2009). *A benchmark system for comparing reliability modeling approaches for digital instrumentation and control systems*. Nuclear Technology, 165(1), 53-95.

Kulić, D. and Croft, E.A. (2005). *Safe planning for human-robot interaction*. Journal of Robotic Systems, 22(7), 383-396.

Leveson, N. (1995). *Safeware: System Safety and Computers, Sphigs Software*. Addison-Wesley Professional.

Leveson, N. (2004). *A new accident model for engineering safer systems*. Safety science, 42(4), 237-270.

Leveson, N., Daouk, M., Dulac, N. and Marais, K., (2003). *Applying STAMP in accident analysis*. Workshop on the Investigation and Reporting of Accidents, Sept. 2003.

Magott, J., & Skrobanek, P. (2012). *Timing analysis of safety properties using fault trees with time dependencies and timed state-charts*. Reliability Engineering & System Safety, 97(1), 14-26.

Maurya, M.R., Paritosh, P.K., Rengaswamy, R. and Venkatasubramanian, V., (2010). *A framework for on-line trend extraction and fault diagnosis*. Engineering Applications of Artificial Intelligence, 23(6), pp.950-960.

Mosleh, A. (2014). *PRA: A Perspective on Strengths, Current Limitations, And Possible Improvements*. Nuclear Engineering and Technology, (1), 1-10.

Ng, Y.S. and Srinivasan, R., (2010). *Multi-agent based collaborative fault detection and identification in chemical processes.* Engineering Applications of Artificial Intelligence, 23(6), pp.934-949.

Palshikar, G. K. (2002). *Temporal fault trees*. Information and Software Technology, 44(3), 137-150.

Papadopoulos, Y., McDermid, J., Sasse, R., & Heiner, G. (2001). *Analysis and synthesis of the behaviour of complex programmable electronic systems in conditions of failure*. Reliability Engineering & System Safety, 71(3), 229-247.

TSTA – Takeoff Safety Training Aid (2016). *Pilot Guide to Takeoff Safety*. Available at https://www.faa.gov/other_visit/aviation_industry/airline_operators/training/media/takeoff_safety.pdf , accessed on 01/21/2016.

Rao, K.D., Gopika, V., Rao, V.S., Kushwaha, H.S., Verma, A.K. and Srividya, A., (2009). *Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment.* Reliability Engineering & System Safety,94(4), pp.872-883.

Rasmussen, J. (1997). *Risk management in a dynamic society: a modelling problem*. Safety science, 27(2), 183-213.

Saleh, J. H., Marais, K. B., Bakolas, E., & Cowlagi, R. V. (2010). *Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges*. Reliability Engineering & System Safety,95(11), 1105-1116.

Saleh, J. H., Haga, R. A., Favarò, F. M., & Bakolas, E. (2014). *Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design*. Engineering Failure Analysis, 36, 121-133.

Swain, A. D. (1990). *Human reliability analysis: Need, status, trends and limitations.* Reliability Engineering & System Safety, 29(3), 301-313.

US Nuclear Regulatory Commission [NUREG]. (1975). *Reactor safety study: An assessment of accident risks in US commercial nuclear power plants*. WASH-1400, NUREG-75/014. Washington, DC.

Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S.N., (2003a). *A review of process fault detection and diagnosis: Part I: Quantitative model-based methods*. Computers & chemical engineering, 27(3), pp.293-311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N. and Yin, K., (2003b). *A review of process fault detection and diagnosis Part II: Qualitative models and search strategies*. Computers and Chemical Engineering, 27(3), pp.313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N. and Yin, K., (2003c). *A review of process fault detection and diagnosis: Part III: Process history based methods*. Computers & chemical engineering, 27(3), pp.327-346.

Walker, M., and Papadopoulos, Y. (2009). *Qualitative temporal analysis: Towards a full implementation of the Fault Tree Handbook*. Control Engineering Practice, 17(10), 1115-1125

Zio, E. (2014). *Integrated deterministic and probabilistic safety assessment: Concepts, challenges, research directions*. Nuclear Engineering and Design, 280, 413-419.