Defence Research and Development Canada Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Analysis of a simulation experiment on optimized crewing for damage control

Renee Chow

Defence R&D Canada
Technical Report
DRDC Toronto TR 2010-128
January 2012

Canada

Analysis of a simulation experiment on optimized crewing for damage control

Renee Chow

Defence R&D Canada – Toronto

Technical Report
DRDC Toronto TR 2010-128
January 2012

Principal Author

Original signed by Renee Chow

Renee Chow

Defence Scientist

Approved by

Original signed by Linda Bossi

Linda Bossi

Section Head, Human Systems Integration

Approved for release by

Original signed by Stergios Stergiopoulos

Stergios Stergiopoulos

Acting Chair, Knowledge and Information Management Committee

Acting Chief Scientist

Abstract

In 2008, a simulation model was developed in the Integrated Performance Modelling Environment (IPME) to evaluate different crew-automation options for naval damage control. This previous work demonstrated the feasibility and value of applying modelling and simulation to explore a large number of factors related to optimized crewing for damage control, but stopped short of performing detailed statistical analysis on the simulation outputs. The current report re-examines the data collected from the 2008 simulation experiment and subjects them to formal hypotheses testing. In particular, it investigates the effects of automation level, automation reliability, and scenario complexity on damage control effectiveness, where damage control effectiveness was measured by time to complete fire response, number of compartments affected by fire, time to complete flood response, and maximal height reached by floodwater. The analyses compared three automation levels (full, medium, and the baseline) that were coupled with three crew sizes (small, medium and large, respectively), two levels of automation reliability (100% and 75%), and two levels of scenario complexity (high, medium). Of the studied factors, automation level was found to have the most significant impact on damage control. Full automation was found to perform best in terms of fire response. Both full automation and the baseline were found to outperform medium automation in terms of flood response. Based on these analyses, this report identified a number of strategies for streamlining future development of related simulation models, as well as future data collection and analysis for related simulation experiments. Finally, this report identified a number of directions for future research on the use of modelling and simulation to inform optimized crewing, including the evaluation of different crew-automation options for whole-ship operation.

Résumé

En 2008, on a élaboré l'environnement intégré de modélisation du rendement (EIMP), un modèle de simulation servant à évaluer différentes formes d'automatisation des équipages aux fins du contrôle des avaries à bord des navires. Ces travaux ont démontré la faisabilité et la valeur de l'application de la modélisation et de la simulation à l'examen d'un grand nombre de facteurs liés à l'optimisation des équipages aux fins du contrôle des avaries, mais sans toutefois élaborer des analyses statistiques détaillées sur les produits de la simulation. Le dernier rapport publié examine à nouveau les données recueillies de l'expérience de simulation de 2008 et les soumet à des vérifications d'hypothèses. Plus précisément, les facteurs examinés sont les effets du degré d'automatisation, de la fiabilité de l'automatisation et de la complexité du scénario sur l'efficacité du contrôle des avaries; l'efficacité du contrôle des avaries étant mesurée en fonction du délai d'exécution de l'intervention en cas d'incendie, du nombre de compartiments touchés par l'incendie, du délai d'exécution de l'intervention en cas d'inondation et de la hauteur maximale atteinte par les dégâts d'eau. Les analyses ont permis de comparer trois degrés d'automatisation (complète, moyenne et de base) selon trois tailles d'équipage (respectivement restreint, moyen et nombreux), deux niveaux de fiabilité de l'automatisation (100 p. 100 et 75 p. 100) et deux niveaux de complexité du scénario (élevé ou moyen). Parmi les facteurs étudiés, on a constaté que le degré d'automatisation avait le plus grand impact sur le contrôle des avaries. On a trouvé que l'automatisation complète donnait les meilleurs résultats pour l'intervention en cas d'incendie. On

a jugé que l’automatisation complète et l’automatisation de base donnaient un rendement supérieur à l’automatisation moyenne pour l’intervention en cas d’inondation. À partir de ces analyses, les auteurs du rapport ont énoncé un certain nombre de stratégies permettant de rationaliser l’élaboration de modèles de simulation connexes, ainsi que la collecte et l’analyse ultérieures de données aux fins d’expériences de simulation semblables. Enfin, les auteurs du rapport ont établi des pistes d’orientation des futurs travaux de recherche sur l’emploi de la modélisation et de la simulation pour documenter l’optimisation des équipages, y compris l’évaluation de différents scénarios d’automatisation de l’ensemble des fonctions du navire.

Executive summary

Analysis of a simulation experiment on optimized crewing for damage control:

Renee Chow; DRDC Toronto TR 2010-128; Defence R&D Canada – Toronto; March 2012.

Introduction or background: In 2008, a simulation model was developed in the Integrated Performance Modelling Environment (IPME) to evaluate different crew-automation options for naval damage control. This previous work demonstrated the feasibility and value of applying modelling and simulation to explore a large number of factors related to optimized crewing for damage control, but stopped short of performing detailed statistical analysis on the simulation outputs. The current report re-examines the data collected from the 2008 simulation experiment and tests specifically for the effects of automation level (full, medium, or baseline), automation reliability (100%, 75%), and scenario complexity (medium, high) on the effectiveness of fire response and flood response.

Results: Automation level was found to have a significant effect on damage control effectiveness. Full automation with small crew size was found to perform best in terms of fire response. In terms of flood response, both full automation with small crew size and the baseline with large crew size were found to outperform medium automation with medium crew size. There was also a significant interaction between automation level and automation reliability. However, main effects of automation reliability and scenario complexity were found only for a subset of the measures.

Significance: A number of strategies were identified for streamlining future development of related simulation models, as well as future data collection and analysis for related simulation experiments. These included the possibilities to apply a reduced set of specific dependent variables, and to use IPME in a standalone mode if task completion times were the primary variables of interest.

Future plans: Future work should investigate the application of modelling and simulation to optimized crewing for whole-ship operation. Supporting work could take the form of comparing multiple crew levels for the same automation level, sensitivity analyses on key simulation parameters such as automation reliability, or comparing different classes or purposes of automation in addition to or instead of levels of automation.

Sommaire

Analyse d'une expérience de simulation de l'équipage optimal aux fins du contrôle des avaries

Renee Chow; DRDC Toronto TR 2010-128; R et D pour la défense Canada – Toronto; Marche 2012.

Introduction ou contexte : En 2008, on a élaboré un modèle de simulation à l'aide de l'outil de l'environnement intégré de modélisation de la performance (EIMP) afin d'évaluer différentes formes d'automatisation de l'équipage aux fins du contrôle des avaries à bord des navires. Ces travaux ont démontré la faisabilité et la valeur de l'application de la modélisation et de la simulation à l'examen d'un grand nombre de facteurs liés à l'optimisation des équipages aux fins du contrôle des avaries, mais sans toutefois élaborer des analyses statistiques détaillées sur les produits de la simulation. Le dernier rapport publié examine à nouveau les données recueillies de l'expérience de simulation de 2008 et vérifie en particulier les effets du degré d'automatisation (complète, moyenne et de base), de la fiabilité de l'automatisation (100 p. 100 et 75 p. 100) et de la complexité du scénario (élevée ou moyenne) sur l'efficacité de l'intervention en cas d'incendie et en cas d'inondation.

Résultats : On a constaté que le degré d'automatisation avait un impact significatif sur l'efficacité du contrôle des avaries. L'automatisation complète d'un équipage restreint donne les meilleurs résultats pour une intervention en cas d'incendie. En ce qui regarde l'intervention en cas d'inondation, on a remarqué que l'automatisation complète d'un équipage restreint et l'automatisation de base d'un équipage nombreux produisent un rendement supérieur à l'automatisation moyenne d'un équipage de taille moyenne. Il y avait aussi une interaction significative entre le degré d'automatisation et la fiabilité de l'automatisation. Cependant, les principaux effets de la fiabilité de l'automatisation et de la complexité du scénario n'ont été constatés que pour un sous-ensemble de données mesurées.

Portée : On a relevé un certain nombre de stratégies permettant de rationaliser l'élaboration de modèles connexes de simulation, ainsi que la collecte et l'analyse ultérieures de données aux fins d'expériences connexes de simulation. Mentionnons, entre autres, la possibilité d'appliquer une série réduite de variables dépendantes précises et celle d'utiliser l'EIMP en mode autonome si les délais d'exécution des tâches sont les variables d'intérêt principales.

Recherches futures : Les travaux à venir devraient porter sur l'application de la modélisation et de la simulation à l'optimisation de l'équipage total du navire. Les travaux connexes pourraient prendre la forme d'une comparaison entre différents niveaux d'équipage pour le même pourcentage d'automatisation, d'analyses de sensibilité des principaux paramètres de simulation comme la fiabilité de l'automatisation, ou d'une comparaison entre différents types ou motifs d'automatisation en plus ou en remplacement des pourcentages d'automatisation.

Table of contents

List of figures

List of tables

Acknowledgements

The author wishes to acknowledge Curtis Coates and Chris Cooper from Esterline | CMC Electronics, Inc. for their assistance with the interpretation of the simulation output files, Dr. Justin Hollands, Dr. David Smith, and Dr. Ann-Renee Blais from DRDC Toronto for their helpful guidance on various aspects of statistical analysis, and Dr. Wenbi Wang for his collaboration on the previous research and for his thoughtful feedback on the current work.

1 Introduction

In recent years, navies around the world have been interested in crew optimization, partly to reduce the operating (and therefore whole life) costs of naval platforms, but also because of the challenge associated with recruiting and retaining sufficient personnel to operate platforms that require very large crew sizes. In addition, advances in technology have opened up the possibility of delivering the same or even enhanced capability with the same or fewer crew members. Therefore, it has become important to investigate how crew and automation can work together to meet the requirements of the modern navy.

1.1 Previous research

In 2005, Defence Research and Development Canada (DRDC) began an Applied Research Project (ARP) on Optimized Crewing for Damage Control (DC) [1]. Although DC is only one of many functions that need to be performed by a ship's crew, it is both a safety-critical and labour-intensive function. It is also a function that needs to be performed on all varieties of naval platforms (e.g., surface combatants, submarines, supply ships, etc.). Therefore, it presents an interesting and potentially generalizable test case for investigating how different crew designs may be complemented by advanced automation to deliver the necessary capabilities.

Within this ARP, a line of research was initiated to determine if modelling and simulation may present a feasible and productive approach to explore the effectiveness of different levels of crew and automation to perform DC. A multi-phase approach was implemented, which included:

1. Functional modelling [2] - where a hierarchy of DC functions were identified without specification of which crew member(s) or automation would be responsible for performing each function. This was essentially a requirements analysis for naval DC;

2. Scenario development [3] - where two scenarios of different complexity were developed to test the effectiveness of any given crew size and automation configuration. The functional model developed in Phase 1 was applied to ensure that the scenarios challenged key DC functions and that each scenario challenged different if overlapping functions. Scenario development also supported the identification of specific tasks that crew and/or automation would be required to perform in each scenario;

3. Options analysis [4] – where three options for crew and automation were specified that would be subjected to an evaluation using the scenarios developed in Phase 2. The three options were:

    a. large crew with baseline automation - this represented traditional practices and mature technologies, and was intended to be reflective of in-service platforms commissioned in the 1980s;

    b. medium crew with medium automation – this represented emerging practices and newly available technologies, and was intended to be reflective of platforms being commissioned in the 2000s; and

c. small crew with full automation – this represented novel practices specifically designed for reduced crewing and emerging technologies, and was intended to be reflective of platforms that may be commissioned in the mid-2010s.

These three phases of analysis then paved the way for the development of a simulation model to assess and compare the effectiveness of the crew-automation options identified in Phase 3 using the scenarios developed in Phase 2.

1.2 Simulation Experiment

In 2008, a simulation model was developed in the Integrated Performance Modelling Environment (IPME) to evaluate different crew-automation options for naval DC [5]. This IPME model, which simulated the activities of crew and automation over the course of different scenarios interacted with a physics-based model of fire and smoke propagation called Fire and Smoke SIMulator (FSSIM) [6] provided by the United States Naval Research Laboratory, Washington, DC. Together, the combined and enhanced models predicted how the activities by the crew and/or automation led to different extents of damage in various compartments aboard the modelled ship. In addition to three crew-automation options mentioned above, the model supported manipulation of other input variables including scenario complexity (i.e., high versus medium), automation reliability (100% vs. 75%), as well as other contextual variables such as fire intensity or permeability of construction materials. The model also produced various forms of output data, including: time to complete specific DC tasks (e.g., extinguishing a fire in a given compartment, removing the source of a flood in a given compartment), and the number of compartments affected (e.g., by smoke, heat). A large number of simulation runs were performed including 25 runs in each of 26 different configurations, and some interesting trends were noted based only on the examination of summary statistics (e.g., means and standard deviations) [7], for example:

- Automation reliability (100% vs. 75%) appeared to make a bigger difference in the full automation option than in the medium automation option;

- For fire response, in particular extinguishing a fire and confirming the extinction of a fire, full automation appeared to perform best;

- For fire response, in particular containing a fire by closing doors and hatches, bounding a fire, and isolating power for personnel safety, full automation appeared to perform best *but only when automation reliability was high*;

- For flood response, in particular containing the flood and removing the source of flood, full automation and the baseline appeared to perform better than medium automation.

However, the most important contribution of the original study [5] was as a proof-of-concept for how modelling and simulation can be applied to the evaluation of crew and automation effectiveness, and to demonstrate the large variety of factors that can be considered in such an evaluation. It was beyond the scope of that study to conduct detailed statistical analyses on the simulation outputs. Therefore, it would appear prudent to re-examine the original data and to subject them to formal hypothesis testing, to verify if significant differences indeed existed between the various experimental conditions, and to identify any ambiguous results that may warrant follow-on investigation through the collection of additional data. In particular, this report tests the following hypotheses:

1. Full automation performs better than medium automation and the baseline.

2. Medium automation option performs better than the baseline.

3. Full automation with high reliability performs better than medium automation with high reliability.

4. Full automation with low reliability performs better than medium automation with low reliability.

5. Medium automation with high reliability performs better than full automation with low reliability.

6. When scenario complexity is high (and heavy casualties are involved), full automation and the baseline perform better than the medium automation option.

7. When scenario complexity is medium (and no casualties are involved), full automation and the medium automation perform better than the baseline.

The motivation behind hypotheses 1. and 2. is to explore whether or not each level of investment in advanced automation (and correspondingly each level of reduction in crew size) can be justified by a performance benefit. It is possible, for example, for a performance difference to exist only between the highest level of automation (i.e., smallest crew size) and the lowest level of automation (i.e., largest crew size), which would raise the question of whether or not an intermediate level of investment in automation (and correspondingly, moderate strategies for crew reduction) can be warranted. Alternatively, the relationship between automation level and performance may not be monotonic, so an intermediate level of investment in automation (and moderate crew reduction) may be associated with a performance benefit, but a high level of investment in automation (and drastic crew reduction) may be associated with a performance decrement. Overall, the posing of hypotheses 1. and 2. does not imply that the author necessarily anticipates advanced automation to be associated with better performance, because advanced automation is coupled with small crew size, and a finding that a larger crew (even one given limited automation) performs better is quite plausible.

The motivation behind hypotheses 3. to 5. is to assess the impact of automation reliability, to assess potential interaction between automation level and automation reliability, and together with the previous hypotheses, to assess the relative importance of automation level versus automation reliability. Automation reliability is an important consideration in the design of any complex system involving both human operators and automation because of the potential for over- or under-utilization of the automation. On one hand, human operators may over-rely on automation and fail to monitor it effectively, possibly because they perceive it to be more reliable than it actually is. On the other hand, human operators may under-utilize automation by ignoring it or turning it off, possibly because they perceive it to be less reliable than it actually is [8, 9]. Various studies have also pointed to the effects of automation reliability and/or the interaction between automation level and automation reliability on performance in military applications such as automated decision aids for command and control [10] and the control of unmanned aerial vehicles [11]. In the current study, we are particularly interested in determining if a high level of automation regardless of its reliability is always associated with performance benefit, or if the performance benefit is only observed when automation reliability is high. It would also be interesting to compare a high level of automation with relatively low reliability against a medium level of automation with relatively high reliability, because it may suggest whether or not it would be more worthwhile to invest in more pervasive and/or powerful automation that may be more prone to failures or in less and/or simpler automation that may not be as prone to failures. Admittedly, these comparisons can only be considered a first step in investigating the effect of automation reliability, since the model simulated system reliability rather than perceived reliability, and did not yet address the issue of misuse or disuse of automation [8] based on mis-calibration by the human operator.

Finally, hypotheses 6. and 7. are intended to explore how scenarios may affect the effectiveness of any crew-automation option in their DC response. When a scenario is relatively straightforward (i.e., a ship is designed to withstand this type of damage with minimal impact on mission effectiveness, and the crew has ample practice and/or experience in handling similar situations), one may expect a high level of performance to be achieved by any crew size, and an even higher level of performance when the crew is supported by advanced automation. However, when a scenario is very challenging (e.g., including the suffering of heavy casualties), it seems more difficult to predict which crew-automation will perform best. For example, the large crew option may perform best because there are enough extra people to take over any duties that would have been assigned to the now-indisposed personnel. Alternatively, the high automation option may perform best because a minimal number of crew is required for the DC response, so even with the casualties, the crew requirement could be met by the still-available personnel.

2 Method

2.1 Overview of Simulation Model

As mentioned in Sub-Section **1.2**, the simulation model analyzed in this study was developed in IPME. IPME is a discrete event simulation environment that can be used to model the activities of human operators as a hierarchical network of tasks that they need to perform. For each task within an IPME network, the model developer can define attributes including but not limited to initiating conditions, a probability distribution for the task completion time, and ending effects (which may include a probability of task failure and specific effects of such failure). Instead of a human operator, it is also possible to assign a task to another resource (e.g., automation), or to have task assignment dependent on criteria that are evaluated at run-time.

IPME supports the detailed modelling of human perceptual and cognitive processes (e.g., by specifying if a task demands visual, auditory, cognitive, and/or psychomotor resources and the expected degree of interference between tasks). It also supports the prediction of cognitive workload (e.g., based on a comparison between the time required and the time available for a given task). However, these capabilities were not utilized in the current simulation model. Instead of in-depth modelling of the tasks for one (or a small number of) operators, the current model focused on representing the broad set of tasks that are required to perform DC on a naval platform (i.e., a function that may involve 70 or more crew members depending on the automation level available) [5]. For example, some of the tasks in the current model include: *Detect fire*, *Contain fire*, and *Confirm fire extinction*. Some lower level tasks associated with *Contain Fire* include: *Shut down ventilation to affected section*, *Close bulkhead isolation valves*, and *Close all relevant doors and hatches* [5]. In this model, once a given human operator is engaged in one task, he/she is considered unavailable for a different task. The model was not as concerned with the workload experienced by any individual operator in any one task, as it was with how the success or failure of a task impacts subsequent tasks and ultimately the performance of the overall system.

IPME was used to track the initiation time and completion time of each task, thereby providing process measures for understanding how DC was performed. In fact, one of the key outputs of IPME was a detailed timeline of all events that occurred during each simulation run. Table 1 shows an excerpt of the timeline produced for one simulation run. As mentioned in Sub-Section **1.2**, IPME was also integrated with FSSIM to produce estimates of how the (timely or delayed) actions of the crew and automation affected how fire and smoke propagated on the simulated ship, thereby providing outcome measures for understanding whether DC was effective. Specifically, IPME together with FSSIM produced data tables showing the maximal temperature, level of carbon monoxide, or soot in each compartment of the simulated ship for each simulation run. Table 2 shows an excerpt of one such data table on maximal temperature in each compartment. These data tables were then processed further to compute measures such as the number of compartments that exceeded a threshold temperature for each simulation run.

STARTING RUN 1					
Task Name	CES Model Task ID	IPME Task ID	Clock	Task Duration	Task Status
Detect Hull Breaches (2.1.3.1) - compartment 164	2.1.3.1	71_3_3_1	10.2	61.43530992	STARTED
Detect Flood Location (2.1.2.1) - compartment 164	2.1.2.1	71_3_2_1	10.4	149.7932748	STARTED
Detect Flood Source (2.1.2.2) - compartment 164	2.1.2.2	71_3_2_2	10.4	173.8716464	STARTED
Detect Flood Volume (2.1.2.3) - compartment 164	2.1.2.3	71_3_2_3	10.4	129.0064062	STARTED
shut down ventilation system to affected section (3.1.1) - compartment 139	3.1.1	72_1_1	20.2	4.500840841	STARTED
detect fire intensity (2.1.1.3) - compartment 139	2.1.1.3	71_3_1_3	20.3	97.975535	STARTED
detect fire type (2.1.1.2) - compartment 139	2.1.1.2	71_3_1_2	20.3	97.975535	STARTED
detect fire location (2.1.1.1) - compartment 139	2.1.1.1	71_3_1_1	20.3	9.181022523	STARTED
shut down ventilation system to affected section (3.1.1) - compartment 139	3.1.1	72_1_1	24.70084	4.500840841	COMPLETE
detect fire location (2.1.1.1) - compartment 139	2.1.1.1	71_3_1_1	29.48102	9.181022523	COMPLETE
Determine Damage Control Strategy (2.4.1) - compartment 139	2.4.1	71_6_1	29.58102	258.9755738	STARTED

Table 1: Sample output data – Task start and completion times

	Compartment Number								
Run Number	99	100	101	102	103	104	105	106	107
1	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
2	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
3	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
4	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
5	298.15	298.15	298.151	298.1511	319.7954	315.997	305.6565	310.4945	311.255
6	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
7	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507
8	298.15	298.15	298.151	298.1511	298.1507	298.1506	298.1507	298.1506	298.1507

Table 2: Sample output data – Maximum compartment temperatures (K)

It is important to note that only one task network was developed to support the entire simulation study. Factors such as crew size, automation configuration, or scenario details (e.g., fire size or location, fire intensity) were specified as part of the configuration file used for each (set of) simulation run(s). Figure 1, re-printed from [5] is a screenshot of the experimenter's interface that had been developed to enable the setup of each simulation run. In particular, this screen enabled the specification of how many crew numbers were available for various DC functions. Other screens were available to specify other simulation parameters. With this experimenter's interface, it would be possible to investigate the impact of other levels of the aforementioned factors without changing the underlying task network model.

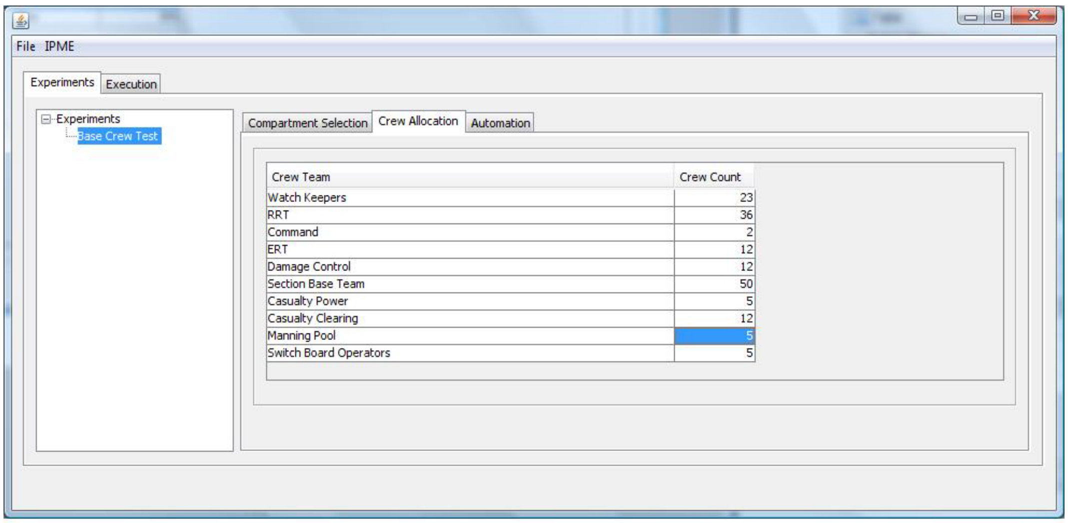

Figure 1: Experimenter's Interface for Specifying Crew Numbers

2.2 Independent Variables

To test the seven hypotheses identified in Section **1**, a 3 x 2 x 2 factorial design was required to examine the main and interaction effects of automation level (full, medium, base), automation reliability (high, low), and scenario complexity (high, medium). An incomplete factorial design was used because it was not meaningful to consider the baseline option (which had only minimal, simple automation) with low automation reliability. Figure 2 illustrates the overall experiment design, and highlights the specific treatments that were excluded.

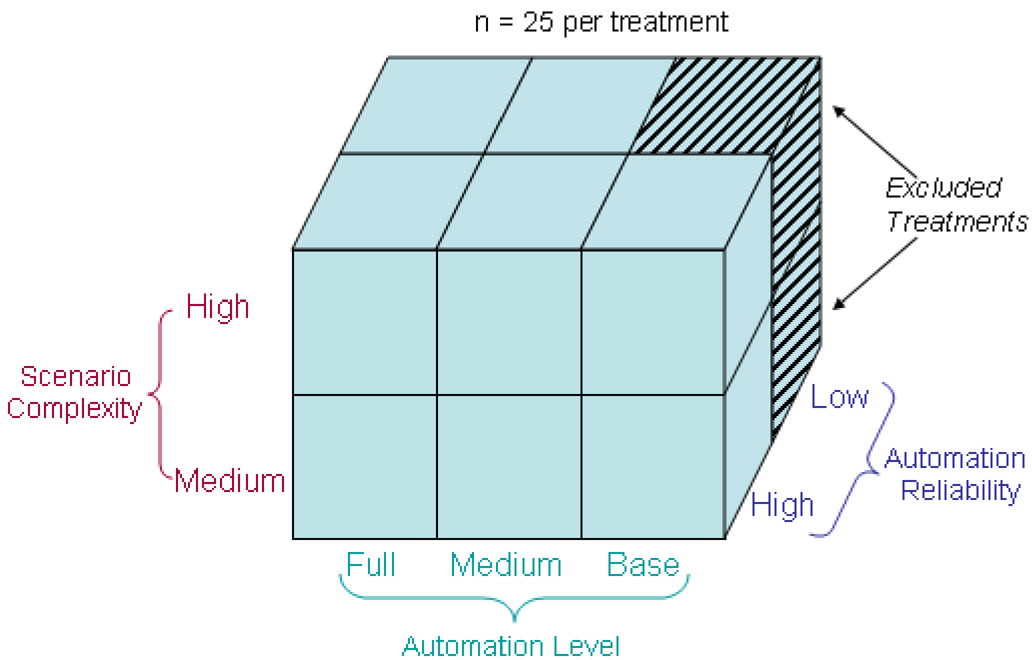

Figure 2. Design of Simulation Experiment

Details of the three automation levels that were simulated were based on the options analysis reported in [4]. To highlight some of the key differences between the levels, full automation included flood detectors in all compartments as well as remote monitoring of liquid levels in tanks; medium automation included flood detectors in all compartments below the water line; while base automation relied on the physical presence of human operators in an affected compartment to detect flood location. To assist in flood response, full automation also included hull integrity sensors and a stress and load detection system that were not available for medium or base automation. In terms of fire response, full automation and medium automation both included automatic shutdown of the ventilation system to the affected section and automatic closure of bulkhead isolation valves, subject to the approval of the DC operator; for base automation, these two actions were performed by the DC operator or the Rapid Response Team. In addition, water mist systems were available to set and maintain boundaries around a fire only for full automation.

Perhaps more importantly, the three automation levels were coupled with three different crew sizes. Table 3, adapted from [5, p.20], shows for each automation level, the number of crew members who were assigned to each DC function. In reality, the ship would sail with many additional crew members who are responsible for non-DC functions. However, these other crew members were not included in this simulation because the objective of this study was only to examine the effectiveness of DC.

	Base Automation (Large Crew)	Medium Automation (Medium Crew)	Full Automation (Small Crew)
Total Crew – available for Damage Control	160	120	70
Command Team	3	3	2
Damage Control – HQ1	5	4	2
Watch Keepers	2	2	1
Rapid Response	4	4	4
Forward Section Base	18	12	10
After Section Base	18	12	10
Section Base 3	11	12	0
Casualty Power	6	4	2
Switch Board Operators	2	2	0
Casualty Clearing	19	10	6
ERT	20	16	8
Manning pool	52	39	25

Table 3: Crewing Levels Corresponding to Each Automation Level

Automation reliability was simulated at 100% (high) and 75% (low). One literature review has shown that with decreasing automation reliability to below a level of around 70%, diagnostic monitoring was worse than had the human not used the automation at all [12]. Another literature review found that there was a level of automation reliability (ranging from 90% and 70% to 60% depending on the system and context) at which trust in automation dropped off sharply [13]. Therefore, while automation had the potential to improve operator safety (e.g., by enabling fire suppression with no or few human operators on scene) and to reduce task times, it was important to acknowledge in the simulation model that automation was fallible, and that a minimal level of automation reliability was required to warrant the appropriate use of automation.

At both levels of scenario complexity, two fires were simulated in the same two compartments of the ship. However, the high complexity of the scenario was characterized by flooding induced by a hull breach that was both deeper (2.0 m vs. 1.0 m below the water line) and larger (15 cm vs. 10 cm in diameter) than the medium complexity scenario. In addition, the high complexity scenario included 20 casualties, while the medium complexity scenario included zero casualties.

Although the entire data set from the original study included additional experimental conditions that varied the contextual variables of fire intensity and construction material permeability, these variables were not of primary interest. Therefore, the current analysis considered only low-intensity (i.e., 100 kW) fire in the medium complexity scenario and only high-intensity (i.e., 1000 kW) fire in the high complexity scenario. The medium and high complexity scenarios already differed in terms of the hull breaches and the number of casualties . It was reasonable to extend these differences to include fires of low versus high intensity in the medium versus high complexity scenarios, respectively. This helped to ensure that the two scenarios differed in terms of the demand for fire response as well as the demand for flood response. In addition, the current analysis examined the output data for only one level of construction material permeability (i.e., 2%).

2.3 Dependent Variables

In terms of simulation outputs, fire response was measured in terms of the following 12 variables, as reported in [5]:

- Times to extinguish fire in compartments 139, 159[1] (V1, V2)[2]
- Times to confirm extinction of fire in compartments 139, 159 (V3, V4)
- Times to contain fire in compartments 139, 159 (V5, V6)
- Times to bound fire in compartments 139, 159 (V7, V8)
- Time to isolate power for personnel safety (V9)
- Number of compartments affected by smoke (V10)
- Number of compartments affected by heat (V11)
- Number of compartments affected by toxicity (V12)

V1 to V9 were extracted from the type of simulation timelines shown in Table 1, while V10 to V12 were based on the type of data tables shown in Table 2. When there is a large number of output variables, and statistical tests are applied to each of the variables individually, then given the probability of Type I error associated with each test, it becomes very likely that at least one (if not more) of the tests will produce a statistically significant result even when one does not really exist (cf., the Bonferroni inequality in [14]). Therefore, it is important to derive meaningful aggregate measures based on the available data to reduce the number of statistical tests required. To this end, new dependent variables (DVs) were defined for this study by aggregating the original output variables as follows:

- DV1: Time to complete fire response (i.e., the simulation time at which the last of the tasks corresponding to V1-V9 above was completed);
- DV2: Number of compartments affected by fire (i.e., the number of compartments in the superset of compartments corresponding to V10-V12 above).

Annex A presents the raw data for these two DVs, but the summary statistics are presented in Table 4 below. For the number of compartments affected by fire (i.e., by smoke, heat, or toxicity), the following operationally relevant thresholds [5] were used. For smoke, a compartment was considered to be affected if there is at least 5×10^{-5} kg soot/ kg gas. For heat, a compartment was

[1] Compartments 139 and 159 were the locations of the simulated fires.
[2] V1 and V2 are variable numbers. A number is assigned to each of the simulation output variables, to make it easier to refer to these variables throughout the report.

considered to be affected if the temperature was at least 85 °C or 358 K which represented the maximal temperature for military standard computer hardware. For toxicity, a compartment was considered to be affected if the level of carbon monoxide was at least 80 parts per million (ppm), which was the level at which it becomes hard to breathe and eyes start to sting. There were 77 compartments in the partial ship model used in this simulation experiment.

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	100%	100%	75%	75%	100%	100%	75%	75%	100%	100%
Scenario	Med	High	Med	High	Med	High	Med	High	Med	High
DV1: Time to complete fire response (seconds)										
Mean	774	782	1282	1421	1817	1817	2033	2052	1996	2037
Std Dev	130	145	315	425	209	226	379	381	243	251
DV2: Number of compartments affected by fire										
Mean	4.1	13.6	20.7	40.9	48.8	54.4	49.3	54.8	48.2	54.8
Std Dev	9.6	16.2	13.1	9.5	1.2	0.8	2.2	0.8	1.2	0.4

Table 4: Summary statistics for the fire-related dependent variables

Similarly, flood response was originally measured in terms of the following variables as reported in [5]:

- Time to contain flood in compartment 164[3] (V13);
- Time to remove / manage source of flood in compartment 164 (V14);
- Number of compartments affected by water (V15).

V13 and V14 were extracted from the type of simulation timelines shown in Table 1, while V15 was based on the type of data tables shown in Table 2. For all of the simulation runs in each experimental condition, exactly one compartment was affected by water (cf., V15). Therefore, V15 was not particularly diagnostic. The following DVs were defined to assess the effectiveness of the flood response:

- DV3: Time to complete flood response (i.e., the simulation time at which the last of the tasks corresponding to V13-V14 above was completed); and
- DV4: Maximal height of flood water (i.e., instead of V15 which always had a value of one, this measure assessed the severity of the flood in that affected compartment)

[3] Compartment 164 was the location of the simulated hull breach (i.e., source of flood).

Annex A presents the raw data for these two DVs, but the summary statistics are presented in Table 5 below.

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	100%	100%	75%	75%	100%	100%	75%	75%	100%	100%
Scenario	Med	High	Med	High	Med	High	Med	High	Med	High
DV3: Time to complete flood response (seconds)										
Mean	1843	1828	1522	1683	1977	2033	2090	2058	1704	1774
Std Dev	217	256	382	408	177	172	253	239	200	239
DV4: Maximum height of flood water (metres)										
Mean	2.36	2.34	1.94	2.15	2.53	2.60	2.68	2.64	2.18	2.27
Std Dev	0.28	0.33	0.49	0.53	0.23	0.22	0.32	0.31	0.26	0.31

Table 5: Summary statistics for the flood-related dependent variables

Conceptually, each of the four DVs of: 1) time to complete fire response, 2) number of compartments affected by fire, 3) time to complete flood response, and 4) maximal height of flood water provide different but complementary ways to assess the effectiveness of DC on a naval platform. The four DVs were expected to be moderately correlated: on one hand, shorter response times are likely to be associated with smaller extents of (fire or water) damage; on the other hand, depending on the strategies employed by the crew and automation (e.g., performing different tasks in series or in parallel), similar response times could still have different damage outcomes.

3 Results

3.1 Multivariate analysis

A multivariate analysis of variance (MANOVA) was conducted using a statistical package called SPSS 17.0 to investigate the main and interaction effects of automation level, automation reliability, and scenario complexity on DC effectiveness, where DC effectiveness was assessed by the four DVs of time to complete fire response, number of compartments affected by fire, time to complete flood response, and maximal height reached by flood water. The MANOVA revealed significant main effects of automation level (Pillai's trace[4] = 1.094, F (8,476) = 71.915, p = 0.000, $\eta_p^2 = 0.547$), automation reliability (Pillai's trace = 0.452, F (4,237) = 48.850, p = 0.000, $\eta_p^2 = 0.452$), and scenario complexity (Pillai's trace = 0.284, F (4,237) = 23.512, p = 0.000, $\eta_p^2 = 0.284$). The MANOVA also revealed significant two-way interaction effects of automation level * automation reliability[5] (Pillai's trace = 0.340, F (4,237) = 30.500, p = 0.000, $\eta_p^2 = 0.340$), and automation level * scenario complexity (Pillai's trace = 0.080, F (8,476) = 2.475, p = 0.012, $\eta_p^2 = 0.040$), as well as a significant three-way interaction effect of automation level * automation reliability * scenario complexity (Pillai's trace = 0.041, F (4,237) = 2.548, p = 0.040, $\eta_p^2 = 0.041$).

Although the above main and interaction effects were statistically significant ($p < 0.05$), the MANOVA also produced partial eta-squares (η_p^2) as indices to describe the "proportion of total variation attributable to (each) factor, partialling out (excluding) other factors from the total nonerror variation" [16, p. 918]. This examination revealed a medium effect size ($\eta_p^2 > 0.50$) [17, 18] for automation level, and small effect sizes ($\eta_p^2 > 0.20$) for automation reliability, scenario complexity and for automation level * automation reliability. The effect sizes for the remaining two-way and three-way interactions were too small to have any practical significance ($\eta_p^2 < 0.05$).

3.2 Univariate analyses

Since the MANOVA found significant main effects of all three independent variables, and a significant interaction effect of automation level * automation reliability, tests of between-subject effects were conducted for each of the four DVs. To prevent inflation of the Type I error rate, a Bonferroni adjustment [14] was made by dividing the original alpha level (0.05) by four to arrive at an adjusted alpha level (0.0125) for the univariate tests corresponding to the four DVs.

As shown previously in Figure 2, this study used an incomplete factorial design where two of the twelve possible treatments had zero observations, making it quite difficult to implement and to interpret a 3 x 2 x 2 Analysis of Variance (ANOVA). Therefore, for each for the four DVs, two complementary ANOVAs were conducted where each ANOVA covered a subset of the treatments as shown in Figure 3 and Figure 4.

---

[4] Although Wilk's lambda is the more commonly used test statistic for a MANOVA, the Pillai's trace is considered to be more robust when the homogeneity of covariances assumption is violated (Box's M = 1386.906, F(90, 62778) = 14.376, p=0.000). [15]

[5] * implies interaction between components.

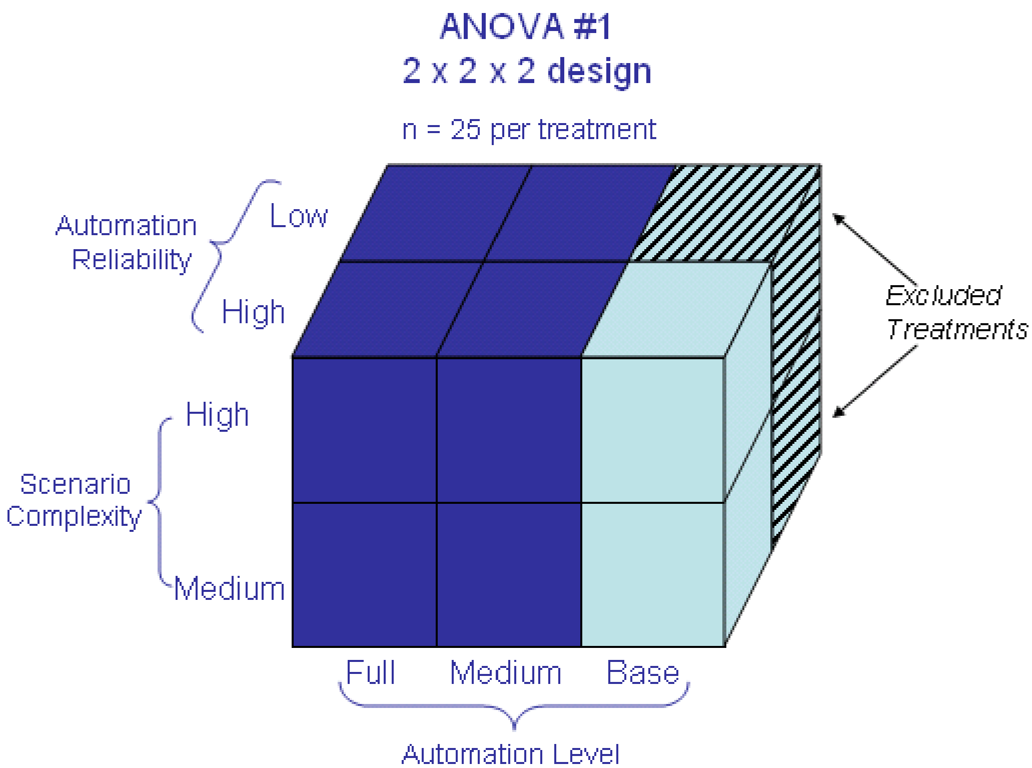

Figure 3. 2x2x2 ANOVA on effects of automation level, reliability, and scenario

Essentially, ANOVA #1 (as shown in Figure 3) enabled an investigation of the main effects of all three independent variables (automation level, automation reliability, and scenario complexity), and their two-way and three-way interactions. However, it does not afford a comparison between the base automation level and the other two automation levels. In a way, the base automation level (tested only at the high automation reliability of 100%) may be viewed as a control condition to which the other conditions can be contrasted. On the other hand, ANOVA #2 (as shown in Figure 4) does afford a comparison between all three automation options (full, medium and base). It also affords opportunities for further investigation of the effect of scenario complexity, and the two-way interaction between automation level and scenario complexity. Since two ANOVAs were conducted for each DV, a further Bonferroni adjustment was made to prevent inflation of the Type I error. Therefore, for each ANOVA reported below, the alpha level was ultimately set at 0.0125 / 2 = 0.006.

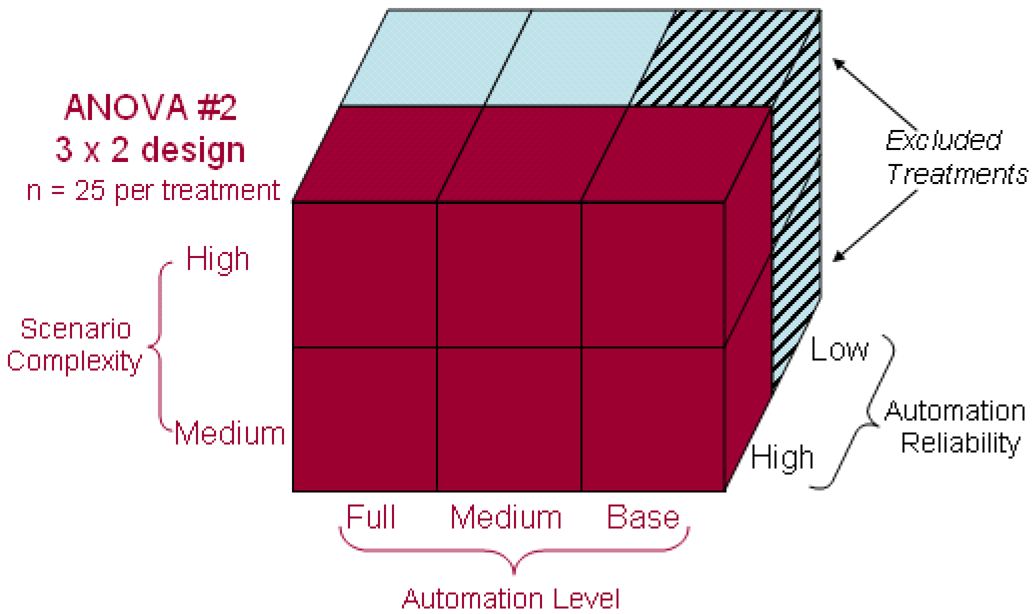

Figure 4. 3x2 ANOVA on effects of automation level and scenario

3.2.1 Time to complete fire response

For the time to complete fire response (DV1), ANOVA #1 which considered all three factors of automation level, automation reliability, and scenario complexity revealed significant main effects of automation level (F (1,192) = 426.798, p = 0.000), and of automation reliability (F (1,192) = 91.005, p = 0.000), and a significant two-way interaction effect of automation level * automation reliability (F (1,192) = 17.302, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Figure 5 presents the means and 95% confidence intervals (CIs) for DV1 as functions of automation level and automation reliability. It shows that the full automation level outperformed the medium automation level, and high automation reliability outperformed medium automation reliability.

Two independent sample t-tests were performed to examine further the interaction between automation level and automation reliability: At both the medium automation level and the full level, performance was significantly better for high reliability than for low reliability (t (78.097) = -3.675, p = 0.000 and t (61.590) = -10.118, p = 0.000 respectively)[6]. In other words, the

[6] One would have expected the degrees of freedom for each of these independent sample t-tests to be 98, since there were 50 observations in each of the two experimental conditions that were being compared. However, the t-test assumes equal variances, and this assumption was violated in both cases as per the Levene's test (p = 0.000 in both cases). As a result, the Behren-Fisher T statistic needed to be used instead of t. The statistic T is distributed approximately as t, but on fewer degrees of freedom as determined by the Welch-Satterthwaite solution (or similar) [19, p.30]. Please note that SPSS 17 automatically computed similar adjustments to the degrees of freedom for all subsequent t-tests, which were applied whenever the assumption of equal variances was violated.

interaction between automation level and automation reliability was ordinal, but automation reliability had a more pronounced effect on the full automation level (where mean difference between reliability levels was 573 seconds) than on the medium automation level (where mean difference between reliability levels was 225 seconds).

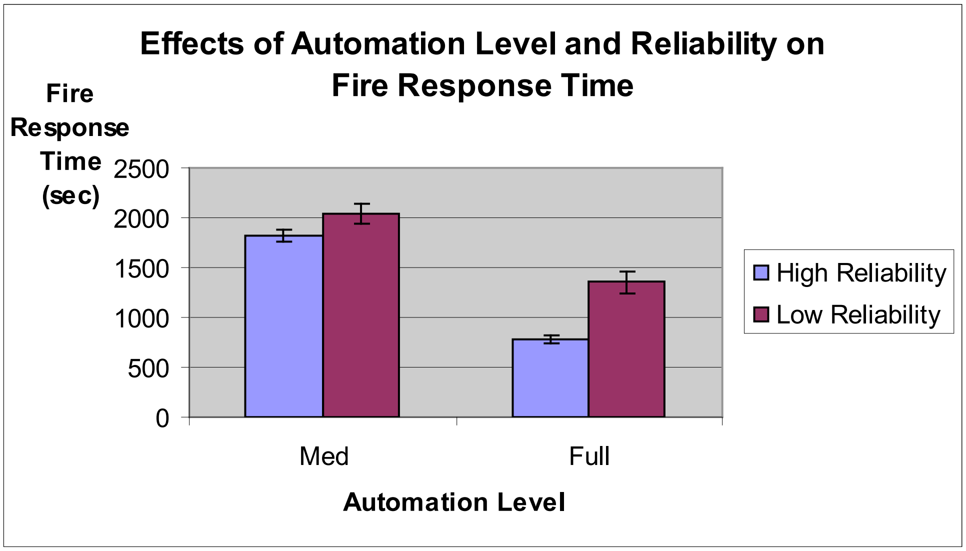

Figure 5: Effects of Automation Level and Automation Reliability on Fire Response Time

ANOVA #2, which considered only the two factors of automation level and scenario complexity, revealed a single significant main effect of automation level (F (2,144) = 520.752, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Post hoc Games-Howell[7] tests [20] found significant differences (p = 0.000) between each pair of automation levels, where full automation outperformed medium automation, and medium automation outperformed base automation. These differences are highlighted in Figure 6.

[7] The Games-Howell test was used instead of the more commonly used Tukey test because the assumption of equal variances was violated as per the Levene's test (F (2,147) = 6.192, p = 0.002).

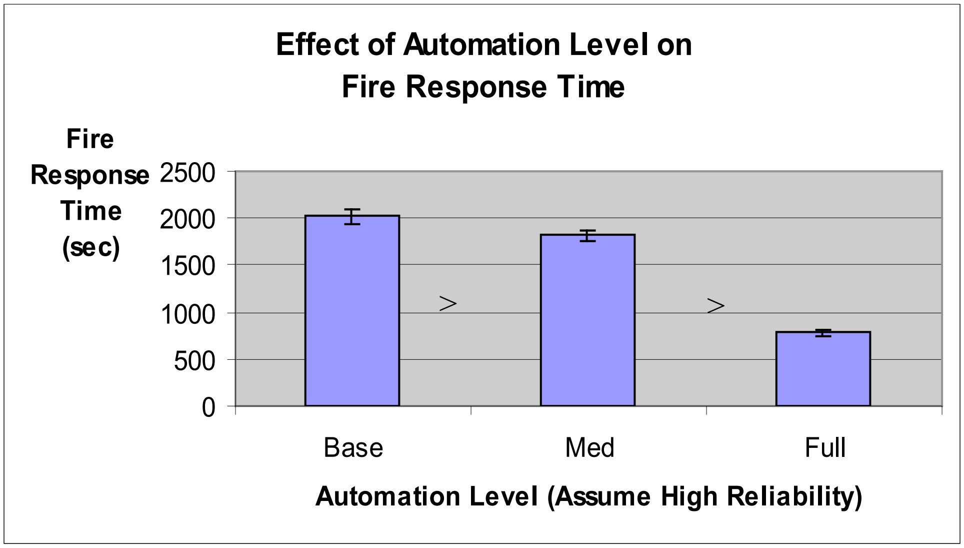

Figure 6: Effect of Automation Level on Fire Response Time, At High Automation Reliability

3.2.2 Time to complete flood response

For the time to complete flood response (DV3), ANOVA #1 revealed a significant main effect of automation level (F (1,192) = 67.668, p = 0.000) and a significant two-way interaction effect of automation level * automation reliability (F (1,192) = 14.969, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Figure 7 shows the effects of automation level and automation reliability, where the full automation level outperformed the medium automation level.

Two independent sample t-tests were performed to examine further the interaction between automation level and automation reliability: At the medium automation level, no significant difference was found between levels of automation reliability. At the full automation level, performance was better at the low reliability level (t (79.345) = 3.546, p = 0.001). This result was counter-intuitive, but no specific explanation could be found except that (as would be expected) there was more variance in the results for full automation with low reliability than for full automation with high reliability (see Figure 8).

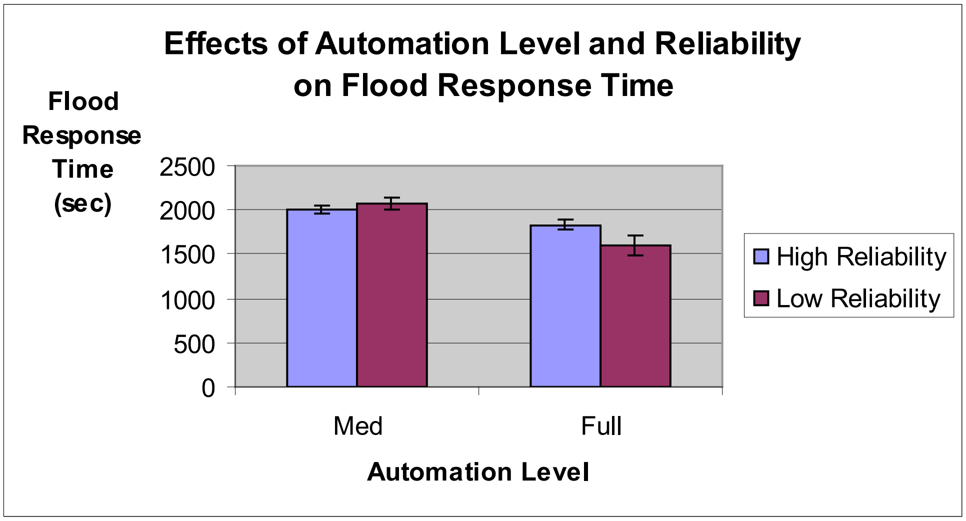

Figure 7: Effects of Automation Level and Automation Reliability on Flood Response Time

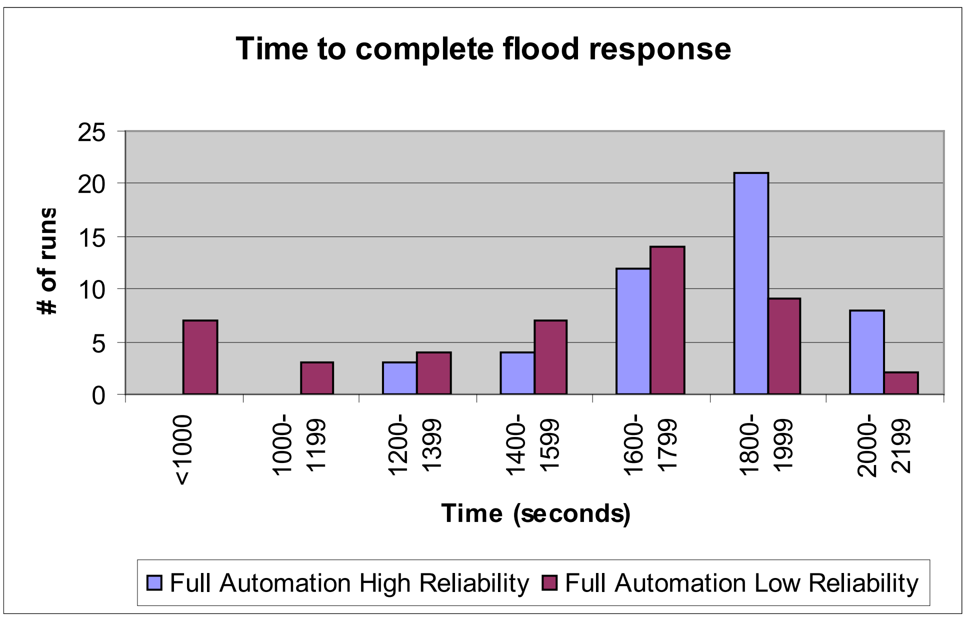

Figure 8: Time to Complete Flood Response for Full Automation with High vs. Low Reliability

ANOVA #2 revealed a single significant main effect of automation level (F (2,144) = 520.752, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Post hoc Tukey[8] tests found significant differences between the medium automation level and base automation level (p = 0.000), and between the medium automation level and full automation level (p = 0.000). These differences are highlighted in Figure 9, which shows that

[8] The Tukey test was used to investigate differences between automation levels for DV3 because the assumption of equal variances was not violated as per the Levene's test (F (2,147) = 1.214, p = 0.300).

both the base automation level and the full automation level outperformed the medium automation level. No significant difference was found between the full automation level and base automation level.

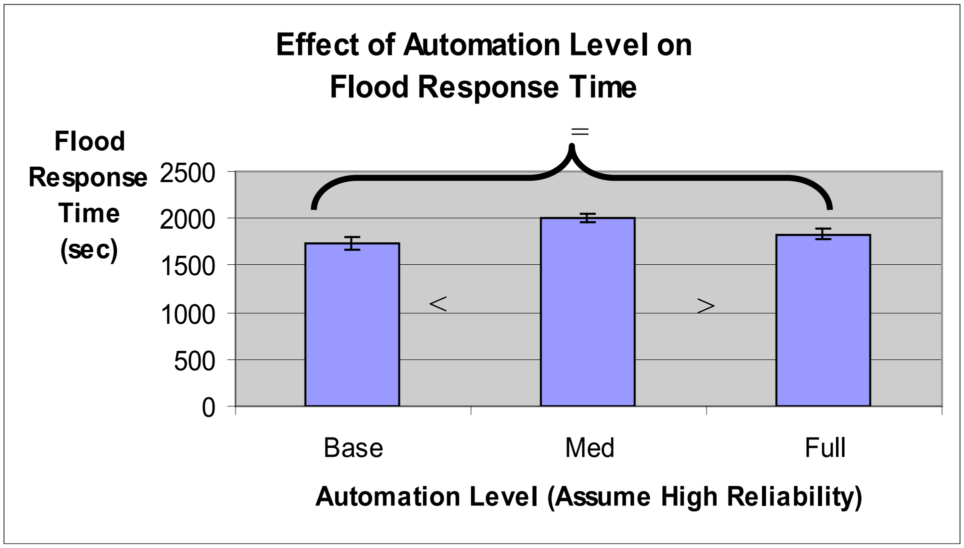

Figure 9: Effect of Automation Level on Flood Response Time, At High Automation Reliability

3.2.3 Number of compartments affected by fire

For the number of compartments affected by fire (DV2), ANOVA #1 revealed significant main effects of automation level (F (1,192) = 655.022, p = 0.000), of automation reliability (F (1,192) = 80.047, p = 0.000), and of scenario complexity (F (1,192) = 66.119, p = 0.000). In addition, significant two-way interaction effects were found for automation level * automation reliability (F (1,192) = 73.876, p = 0.000), and for automation level * scenario complexity (F (1,192) = 13.823, p = 0.000). Figure 10 shows the effects of automation level and automation reliability, where the full automation level outperformed the medium automation level, and high reliability outperformed low reliability. Two independent sample t-tests were performed to examine further the interaction between automation level and automation reliability: At the medium automation level, no significant difference was found between levels of automation reliability. At the full automation level, performance was better at the high reliability level (t (98) = -7.487, p = 0.000).

Figure 11 shows the effects of automation level and scenario complexity on the number of compartments affected by fire, where the full automation level outperformed the medium automation level, and where performance was better in the medium complexity scenario than in the high complexity scenario. Two independent sample t-tests were also performed to examine further the interaction between automation level and scenario complexity: At both the medium automation level and the full level, performance was significantly better in the medium complexity scenario than in the high complexity scenario (t (98) = -20.051, p = 0.000 and t (90.454) = -4.417, and p = 0.000, respectively). In other words, the interaction between automation level and scenario complexity was ordinal, but scenario complexity had a more pronounced effect on the full automation level (where mean difference between scenarios was 14.8 compartments) than on the medium automation level (where mean difference between scenarios was 5.5 compartments).

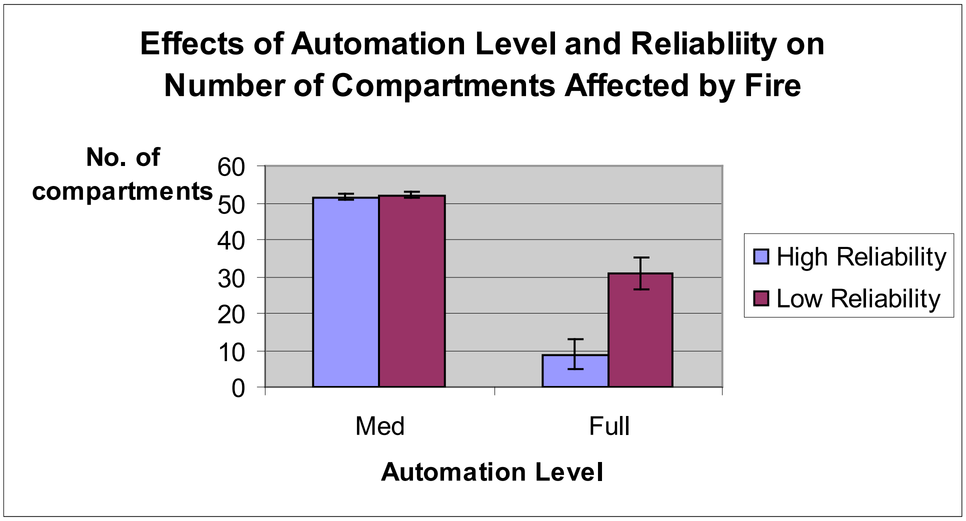

Figure 10: Effects of Automation Level and Reliability on Compartments Affected by Fire

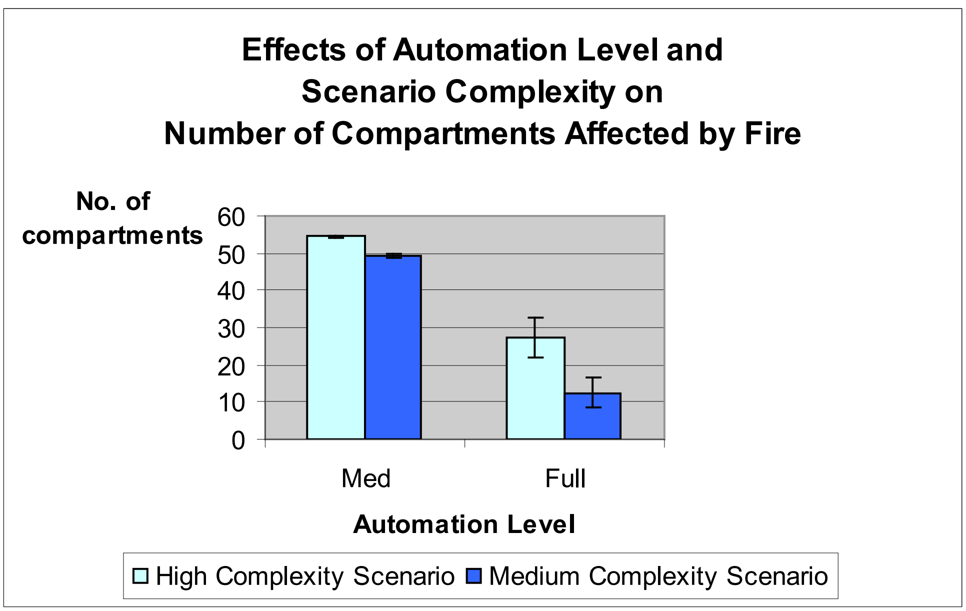

Figure 11: Effects of Automation Level and Scenario on Compartments affected by Fire

ANOVA #2 revealed significant main effects of automation level (F (2,144) = 508.392, p = 0.000) and of scenario complexity (F (1,144) = 32.779, p = 0.000). No significant interaction effect of automation level * scenario complexity was found. Figure 12 shows the main effect of scenario complexity, where performance was better in the medium complexity scenario than in the high complexity scenario. To further investigate the main effect of automation level, post hoc

Tukey[9] tests found significant differences between the full automation level and medium automation level (p = 0.000), and between the full automation level and base automation level (p = 0.000). No significant difference was found between the medium automation level and base automation level. Figure 13 shows the main effect of automation level, and highlights significant differences between specific automation levels.

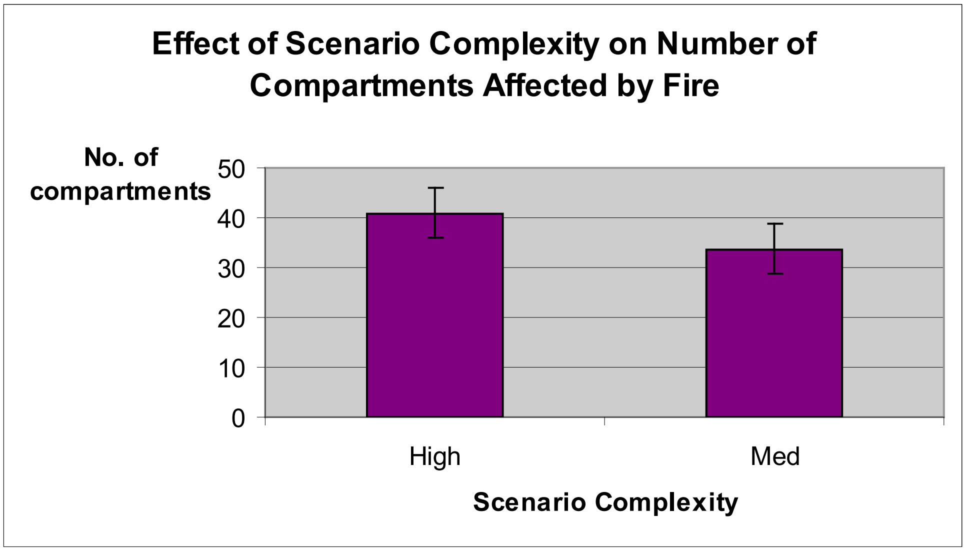

Figure 12: Effect of Scenario Complexity on Compartments Affected by Fire

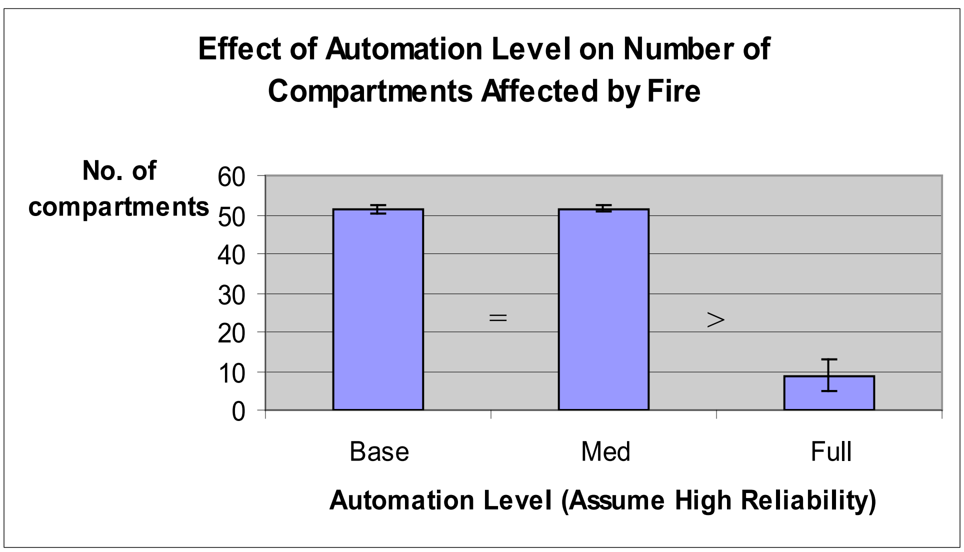

Figure 13: Effect of Automation Level on Compartments Affected by Fire, At High Reliability

[9] Similar to the case of DV1, the Games-Howell test was used because the assumption of equal variances was violated as per the Levene's test (F (2,147) = 101.816, p = 0.000).

3.2.4 Maximal height of flood water

For maximal height of flood water (DV4), ANOVA #1 revealed a significant main effect of automation level (F (1,192) = 68.449, p = 0.000) and a significant two-way interaction effect of automation level * automation reliability (F (1,192) = 15.422, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Figure 14 shows the effects of automation level and automation reliability, where the full automation level outperformed the medium automation level. Two independent sample t-tests were performed to examine further the interaction between automation level and automation reliability: At the medium automation level, no significant difference was found between levels of automation reliability. At the full automation level, performance was better at the low reliability level (t (79.376) = 3.556, p = 0.001). As with the time to complete flood response, this result was counter-intuitive, but no specific explanation could be found except that as would be expected, there was more variance in the results for full automation with low reliability than for full automation with high reliability (see Figure 15).[10]

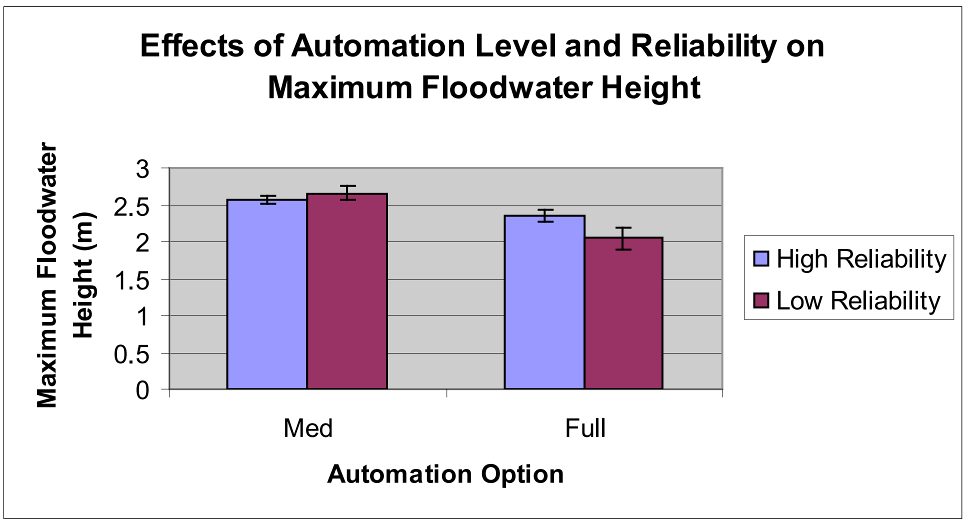

Figure 14: Effects of Automation Level and Reliability on Maximum Floodwater Height

ANOVA #2 revealed a significant main effect of automation level (F (2,144) = 20.096, p = 0.000). No significant main or interaction effect associated with scenario complexity was found. Post hoc Tukey[11] tests found significant differences between the medium automation level and base automation level (p = 0.000), and between the medium automation level and full automation level (p = 0.000). No significant difference was found between the full automation level and base automation level. These findings are presented in Figure 16, which shows that both the full and base automation levels outperformed the medium automation level.

[10] It would be prudent, before further application and extension of the simulation model, to investigate the possibility of a software bug causing this pattern of results.

[11] Similar to the case of DV2, the Tukey test was used because the assumption of equal variances was not violated as per the Levene's test (F (2,147) = 1.274, p = 0.283).

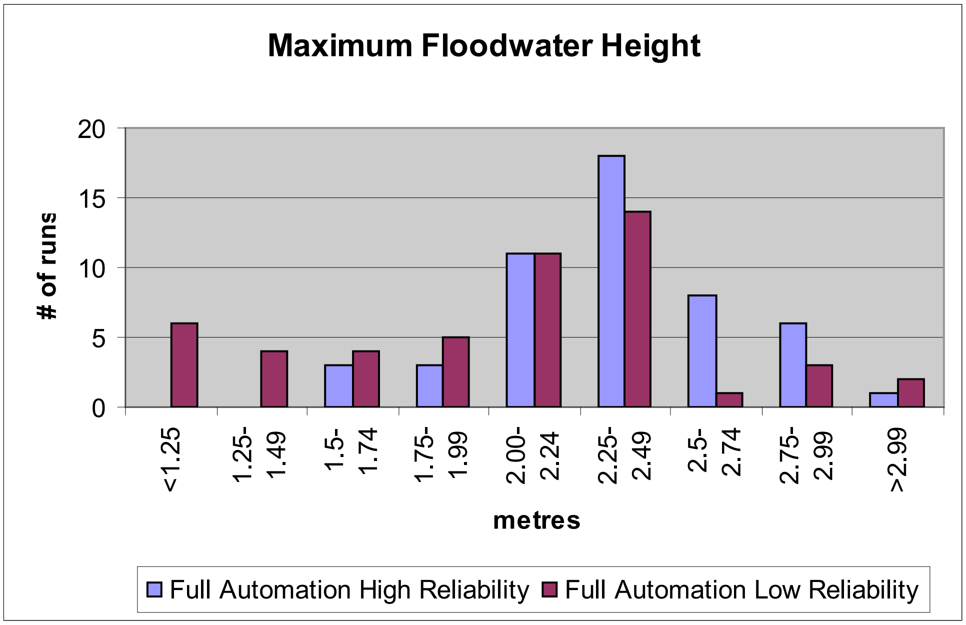

Figure 15: Maximum Floodwater Height for Full Automation with High vs. Low Reliability

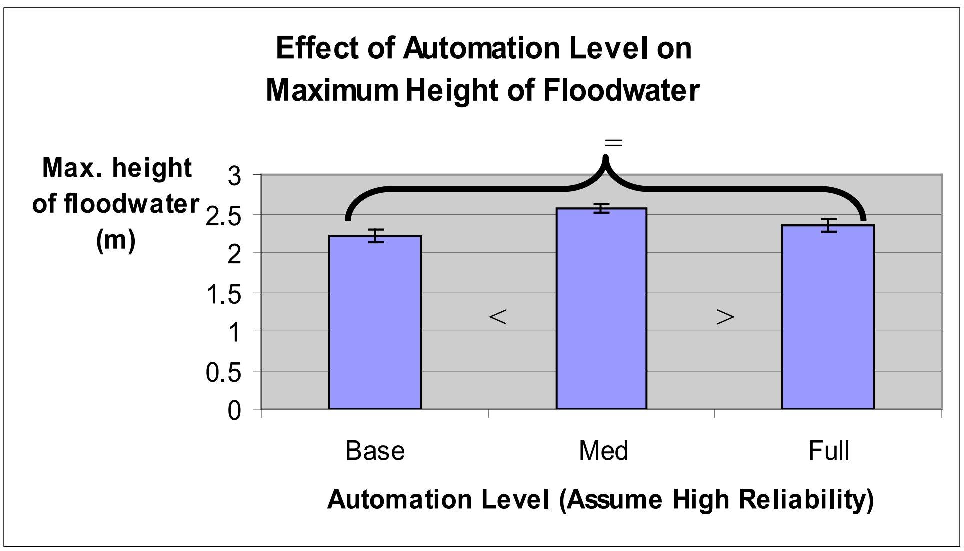

Figure 16: Effect of Automation Level on Floodwater Height, At High Automation Reliability

3.2.5 Comparison of five automation options

One other reasonable perspective on the two factors of automation level and automation reliability would be to view each combination of the two factors as a distinct and meaningful automation option to be compared directly with the other combinations. This comparison could be of practical value because each of these options could potentially represent the product offering from a particular vendor at a specific cost. For example, vendor A may propose a very comprehensive and sophisticated set of DC automation that had relatively low reliability (i.e., full-automation-low-reliability) at price point X; while vendor B may propose a less ambitious set of DC automation that had relatively high reliability (i.e., medium-automation-high-reliability) at a similar price point Y; and vendor C may propose similarly comprehensive and powerful automation as vendor B but with relatively low reliability (i.e., medium-automation-low-reliability) and at a lower price point Z. It would be important to assess the effectiveness of the options proposed by different vendors (e.g., A, B, C) to enable further cost-benefit analysis. In fact, direct comparison of the five tested combinations of automation level and automation reliability may yield results that are more readily interpreted and acted upon by decision makers than comparisons that speak to the main and interaction effects of the two factors.

As a result, a third type of ANOVA (as illustrated in Figure 17) was conducted to investigate potential differences between the five tested automation options, where each "option" is defined by a specific automation level (full, medium, or base) as well as a specific automation reliability (i.e., high or low). Since ANOVA #3 can be seen as an alternative analysis to the ANOVA results reported in Sub-Sections **3.2.1-3.2.4**, a Type I error rate of 0.05 / 4 = 0.125 was employed for the test corresponding to each of the four DVs.

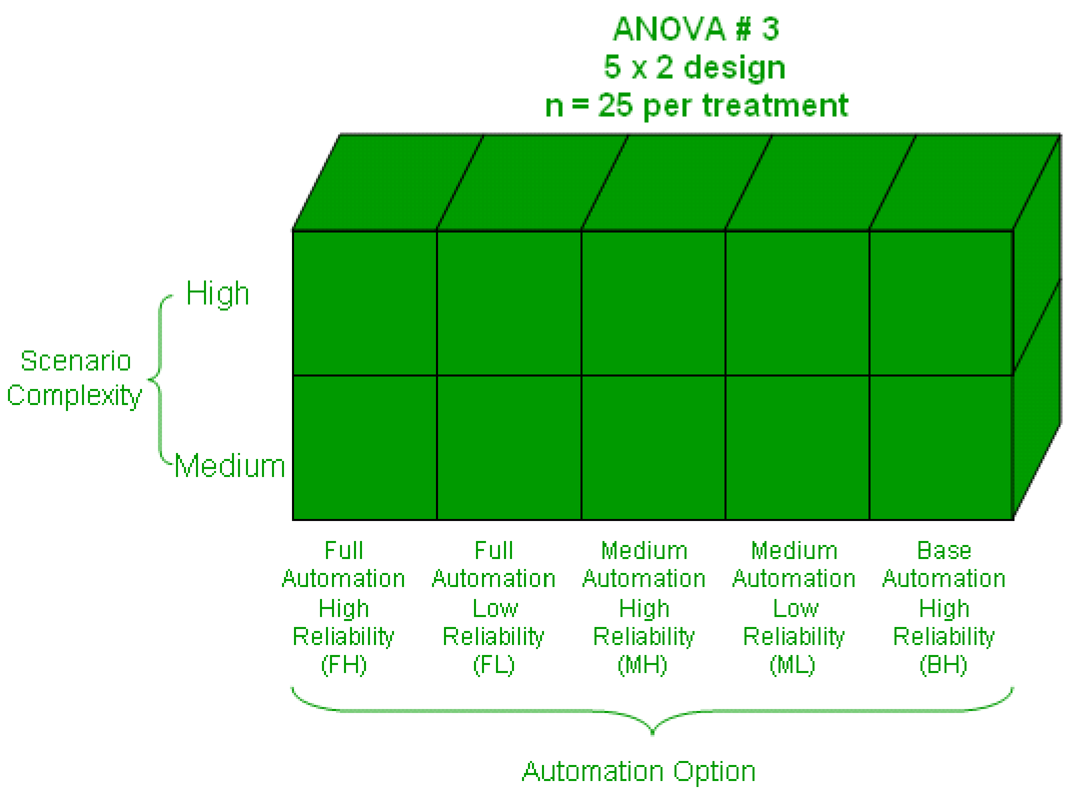

Figure 17. 5x2 ANOVA on effects of automation option and scenario

For three of the four DVs (except DV2 – number of compartments affected by fire), ANOVA #3 revealed only a significant main effect of automation option. Figure 18 shows the means and 95% CIs for the time to complete fire response (DV1), with the five automation options ordered from the best-performing to the worst-performing. Post hoc Games-Howell[12] tests indicated significant differences between all but one pair of automation options. Specifically, the two worst-performing options (i.e., base automation with high reliability, and medium automation with low reliability) were not significantly different.

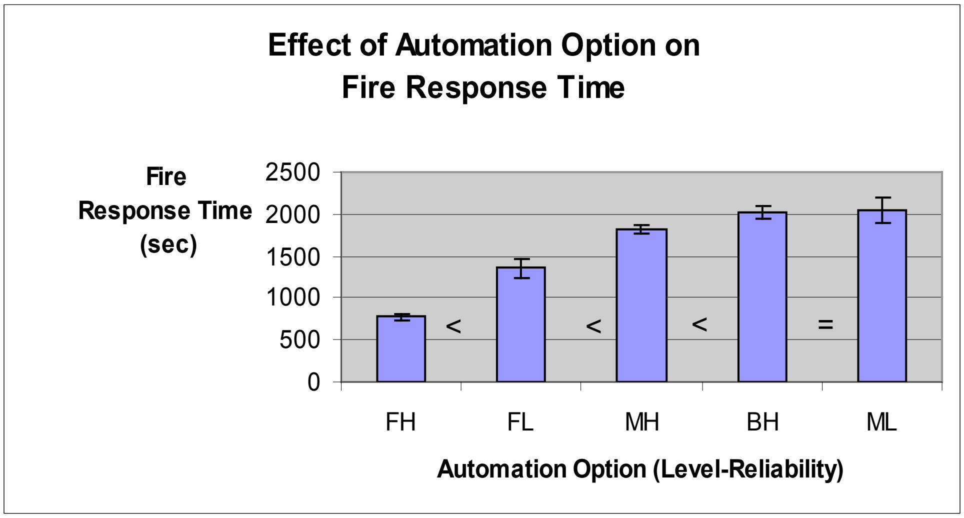

Figure 18: Comparison of Automation Options by Fire Response Time

Figure 19 shows the means and 95% CIs for the time to complete flood response (DV3), with the five automation options ordered from the best-performing to the worst-performing. Post hoc Games-Howell tests indicated no significant difference between three pairs of adjacent options (i.e., best and second-best option, second-best and third-best option, and the two worst options), but significant differences between all other pairs of options.

Figure 20 shows the means and 95% CIs for the maximal height reached by flood water (DV4), with the five automation options ordered from the best-performing to the worst-performing. Similar to the results for the time to complete flood response, post hoc Games-Howell tests indicated no significant difference between three pairs of adjacent options (i.e., best and second-best option, second-best and third-best option, and the two worse options), but significant differences between all other pairs of options.

For the number of compartments affected by fire (DV2), ANOVA #3 found significant main effects of the automation option (F (4,240) = 291.111, p = 0.000) and of scenario complexity (F (1,240) = 89.219, p = 0.000), as well as a significant interaction effect of automation option * scenario complexity (F (4,240) = 7.645, p = 0.000). Figure 21 shows the means and 95% CIs for DV2, with the five automation options ordered from the best-performing to the worst-performing. Post hoc Games-Howell tests indicated no significant differences between the three worst-performing options, but significant differences between all other pairs of options.

[12] The Games-Howell test was used for all post hoc pairwise comparisons associated with ANOVA #3 because for each of the four DVs, the assumption of equal variances was violated as per the Levene's test (F (4,245) > 9.004, p = 0.000).

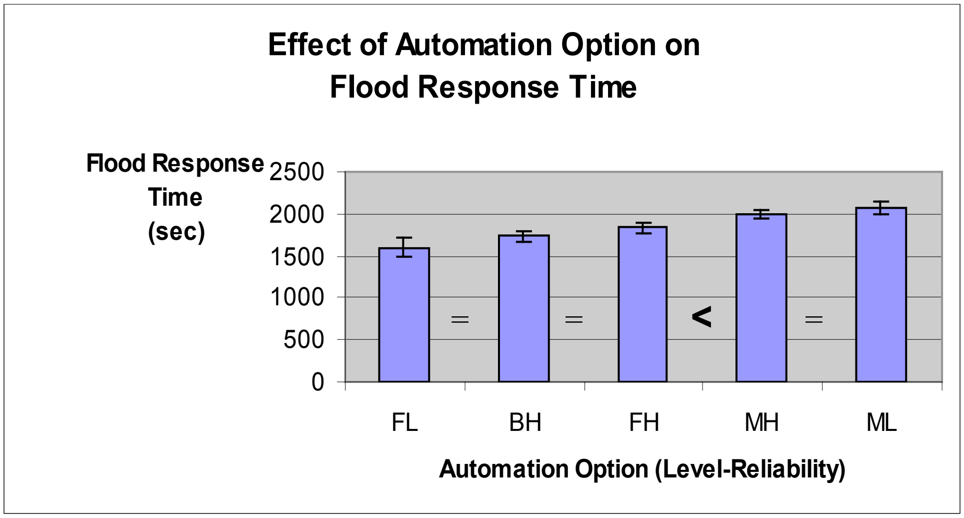

Figure 19: Comparison of Automation Options by Flood Response Time

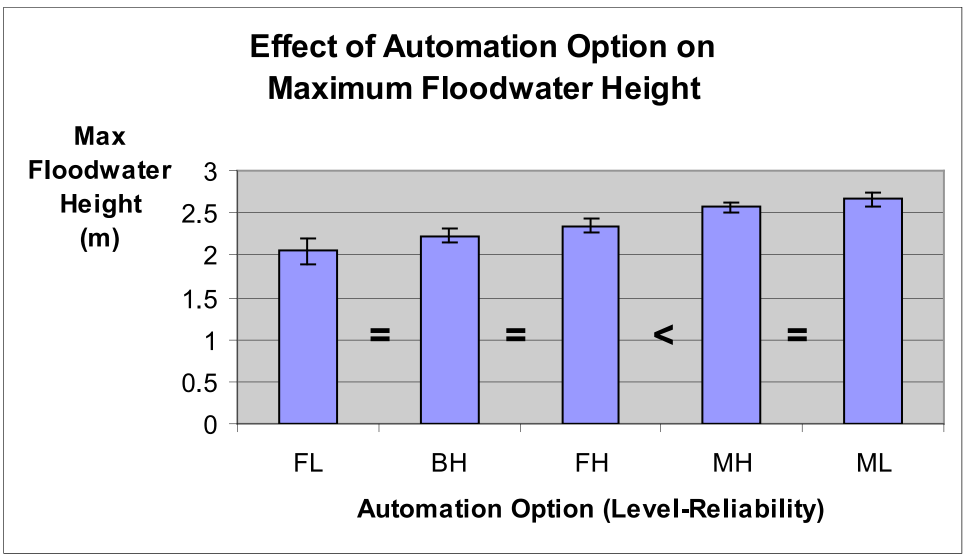

Figure 20: Comparison of Automation Options by Maximum Floodwater Height

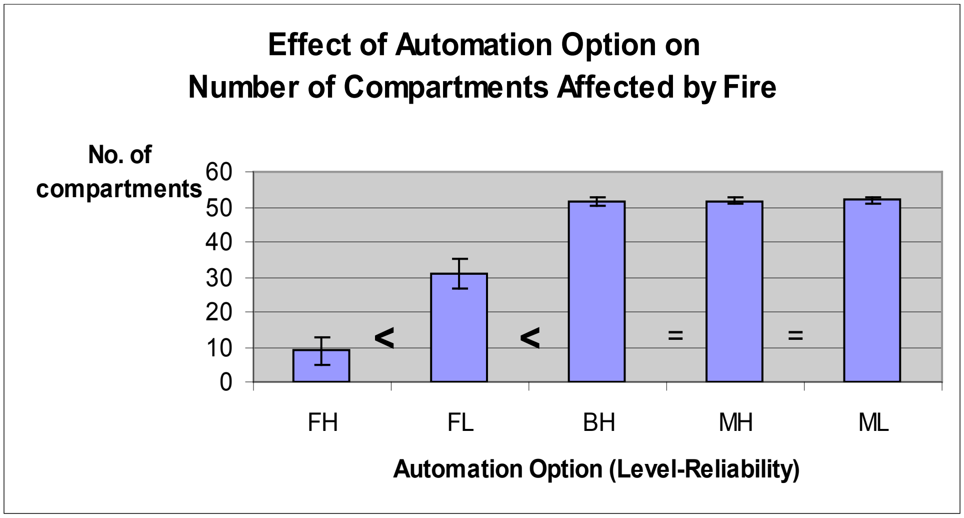

Figure 21: Comparison of Automation Options by Compartments Affected by Fire

As for the main effect of scenario complexity, performance in the medium complexity scenario (mean = 34.2 compartments) was found to be better than performance in the high complexity scenario (mean = 43.7 compartments). Five independent sample t-tests were performed to further investigate the interaction between the automation option and scenario complexity: For each automation level, performance was better in the medium complexity scenario than in the high complexity scenario ($p < 0.05$), so the interaction between automation option and scenario complexity was ordinal. However, the mean differences between levels of scenario complexity varied from 20.2 compartments (in the case of full automation with low reliability) to 5.5 compartments (in the case of medium automation with low reliability).

3.3 Summary

In summary, every relevant multivariate or univariate test that was conducted indicated a significant main effect of automation level. The main effect of automation reliability was consistently found for fire-related measures (DV1, DV3) but not for flood-related measures (DV2, DV4). The main effect of scenario complexity was found for only one fire-related measure (i.e., DV3 - number of compartments affected by fire).

With regards to automation level as a standalone factor, the full automation level outperformed the medium and base automation levels in fire response; while the medium automation level underperformed relative to the full and base automation levels in flood response. With regards to automation reliability as a standalone factor, automation with high reliability outperformed automation with low reliability on fire-related measures (DV1, DV2), but not on flood-related measures (DV3, DV4). With regards to scenario complexity as a standalone factor, results from all three types of ANOVAs found significantly better performance in the medium complexity scenario than in the high complexity scenario but only in terms of number of compartments affected by fire (DV2).

Every relevant multivariate or univariate test that was conducted also indicated a significant interaction between automation level * automation reliability. On the fire-related measures (DV1, DV2), automation reliability had greater effects on performance at the full automation level than at the medium automation level. Higher performance was observed at high reliability than at low reliability, but the performance differences between reliability levels were not always significant (e.g., no significant difference on DV2 at the medium automation level). On the flood-related measures (DV3, DV4), performance differences between reliability levels were only significant at the full automation level, where performance was better at low reliability.

A significant interaction between automation level * scenario complexity was found for only one measure (i.e., DV2 – number of compartments affected by fire). Performance was better in the medium complexity scenario than in the high complexity scenario, and scenario complexity had a greater effect at the full automation level than at the medium automation level. On a similar note, when the factors of automation level and automation reliability were used in combination to produce five complete, distinct definitions of automation options (cf., ANOVA #3), a significant interaction between automation option * scenario complexity was found for the same DV. Performance was always significantly better in the medium complexity scenario than in the high complexity scenario, but the magnitudes of the performance differences between scenarios varied across the automation options.

Finally, when five distinct automation options were defined based on a combination of automation level and automation reliability and these options were compared, a main effect of automation option was found for all four DVs. In terms of fire-related measures (DV1, DV2), FH performed best, and FL performed second-best, while the remaining three options performed more poorly. In terms of flood-related measures (DV3, DV4), the five automation options could be divided into two groups, with Full-Automation-High-Reliability (FH), Full-Automation-Low-Reliability (FL), and Base-Automation-High-Reliability (BH) in the higher-performing group, and Medium-Automation-High-Reliability (MH) and Medium-Automation-Low-Reliability (ML) in the lower-performing group.

4 Discussion

This chapter will begin by re-visiting the seven hypotheses presented in Sub-Section **1.2** in light of the evidence gathered in Section **3**, and noting the implications of the acceptance or rejection of these hypotheses.

4.1 Implications re: Automation Levels

Hypothesis (1): Full automation performs better than medium automation and the baseline.

For fire response (where good performance includes both a fast response time and fewer affected compartments), full automation did perform better than both medium automation and the baseline (refer to Figures 5, 6, 10, 13). For flood response (where good performance includes both a fast response time and less severe flooding), full automation did perform better than medium automation, but performed similarly to the baseline (refer to Figures 7, 9, 14, 16).

Hypothesis (2): Medium automation performs better than the baseline.

For fire response, there was some, incomplete evidence that medium automation performed better than the baseline (i.e., in terms of response time but not necessarily in terms of affected compartments) (refer to Figures 6, 13). For flood response, the available evidence pointed to the opposite situation where the baseline performed better than medium automation (refer to Figures 9, 16).

Looking across the evidence related to Hypotheses (1) and (2), investment in full automation appeared worthy of consideration because of its performance benefit over both of the other automation levels in fire response, and at least over the medium automation level in terms of flood response. Investment in full automation would be especially attractive if the life cycle costs associated with the advanced automation (as compared to the baseline) were comparable or lower than the life cycle costs associated with the large crew size required by the baseline (as compared to the much smaller crew size enabled by full automation). However, there was little support for investment in medium automation because overall, it did not seem to perform better than the baseline.

4.2 Implications re: Automation Reliability

Hypothesis (3): Full automation with high reliability performs better than medium automation with high reliability.

In all aspects of DC, full automation with high reliability performed better than medium automation with high reliability.[13]

[13] Sub-Sections **3.1-3.4** reported on the automation level * automation reliability interaction for all four DVs. The two t-tests that were reported for each DV investigated differences between reliability levels for each automation level. However, for each DV, two complementary t-tests were also conducted on the

Hypothesis (4): Full automation with low reliability performs better than medium automation with low reliability.

In all aspects of DC, full automation with low reliability performed better than medium automation with low reliability.[14]

Hypothesis (5): Medium automation with high reliability performs better than full automation with low reliability.

There was no evidence to support this hypothesis. In fact, in all aspects of DC, full automation with low reliability performed better than medium automation with high reliability (see Figures 18-21).

Looking across the evidence related to Hypotheses (3)-(5), it appeared that automation level was a more important determinant of DC performance than automation reliability. It is important to keep in mind, however, that this study only examined two levels of automation reliability (100% vs. 75%), so it is possible that for automation with still lower reliability (i.e., < 75%), the benefit of advanced automation may start to become eroded by frequent automation failures. It would be more prudent to conclude that if multiple automation options all meet a reasonable threshold in terms of reliability, then more advanced automation would be expected to produce a higher level of performance.

4.3 Implications re: Scenario Complexity

Hypothesis (6): When scenario complexity is high (and heavy casualties are involved), full automation and the baseline perform better than medium automation.

Regardless of scenario complexity, full automation performed better than medium automation in all aspects of DC (see Figures 5, 6, 7, 9, 10, 13, 14, 16). The baseline did perform better than medium automation for flood response (see Figures 9, 16). But medium automation performed better than the baseline for fire response. [15]

Hypothesis (7): When scenario complexity is medium (and no casualties are involved), full automation and medium automation perform better than the baseline.

---

differences between automation levels at each reliability level. At high automation reliability, each t-test indicated a significant difference between the medium automation level and the full automation level, where performance was better for the full automation level.

[14] As mentioned in the previous footnote, for each DV, a t-test was also conducted on the difference between automation levels at low reliability. Each of these four t-tests indicated a significant difference between the medium automation level and the full automation level, where performance was better for the full automation level.

[15] In terms of fire response time (DV1), the difference between medium automation and the baseline can be seen in Figure 6 as there was no significant interaction between automation level * scenario complexity. In terms of number of compartments affected by fire (DV2), there was a significant interaction between automation level and scenario complexity. Therefore, an independent sample t-test was conducted at the high scenario complexity, to compare medium automation with the baseline. The t-test found a significant difference between the two automation levels (t (35.294) = -2.191, p = 0.035), where medium automation (mean = 54.4 compartments) performed better than the baseline (mean = 54.8 compartments).

Regardless of scenario complexity, full automation performed better than the baseline for fire response (see Figures 6, 11), but not for flood response where the two automation levels had similar performance (see Figures 9, 16). Contrary to the hypothesis, the baseline performed better than medium automation for flood response (see Figures 9, 16). But the results were mixed for fire response, where medium automation performed better than baseline in terms of fire response time (see Figure 6), but not in terms of affected compartments where performance was similar between the two automation levels (see Figure 13). [16]

Looking across the evidence related to Hypotheses (6) and (7), scenario complexity seemed to have little or no impact on the relative merits of the different automation levels. In fact, the complexity of a scenario (i.e., fire size, number of casualties, severity of hull breach) seemed to be less important than the breadth of the scenario – i.e., the inclusion of both fire and flood. Looking across the evidence related to all hypotheses, the two fire-related measures appeared correlated: In most cases, a higher performing automation level based on one measure was also higher performing based on the other measure; in the few remaining cases, a performance difference was noted in terms of fire response time but not in terms of affected compartments. The two flood-related measures appeared highly correlated, in that a higher performing automation level based on one measure was always higher performing based on the other measure. However, the performance results related to fire often followed a different pattern than the performance results related to flood. Therefore, it would be critical for future simulation experiments to use scenarios that involve both fire and flood, and to apply measures of performance related to both types of damage events. However, it may be sufficient to use only one measure related to fire (probably response time since that appeared more discriminatory) and only one measure related to flood.

4.4 Limitations

There were several noteworthy limitations to the current study: First, automation level and crew size were confounded – i.e., full automation was coupled with a small crew, medium automation with a medium crew, and base automation with a large crew. Although the assumption that as automation level increases, crew size will decrease is valid from a practical perspective (in reality, future naval platforms will be designed to operate with more advanced automation and smaller crews; and DC automation is often advocated as an enabler for crew size reduction), it was not possible to determine from the study whether the performance benefit observed at any one automation level was (primarily) due to the available automation or to the available crew. This limitation should not be of great concern at the full automation level, since performance was consistently high despite the small crew size. However, at the medium automation level, it may be informative to investigate if different crew sizes coupled with the same automation level would produce different performance.

Second, automation level was only one of several possible distinctions that can be drawn between the design options that were tested and compared. The specific implementations of the full automation, medium automation, and the baseline were based on an in-depth study reported in [4], and the ordering of these options based on automation level should not be controversial.

[16] An independent sample t-test was conducted at medium scenario complexity, to compare medium automation with the baseline. The t-test did not find a significant difference between the two automation levels (t (47.847 = 1.981, p = 0.053).

However, one might wonder if the different options were optimized for different purposes. For example, was one option optimized for fire response while another optimized for flood response? Or was one option designed for both fire and flood management while another option was designed for one type of event but not the other? Or did one option include automation that both provided information and acted on the environment, while another option provided information only (but relied on the human operators to take action), or vice versa? Before acting on the finding that full automation performed best, it would be important to probe deeper into where the medium automation fell short especially in terms of flood response. Perhaps a different variation of "medium" automation that included different mechanisms for flood management would produce a very different level of performance. Also, depending on what decision makers deem to be of higher or ultimate importance (i.e., fire response, flood response, or both), the relative merits of the tested options could be different.

Third, as with all simulation experiments, the outputs were only as valid as the inputs that had been entered into the simulation. The current simulation was based on four years of extensive research into optimized crewing and damage control, including consultations with subject matter experts in various relevant disciplines (cf., [2]-[5]), as well as integration with a validated, physics based simulation of fire and smoke propagation (cf., [6]). However, where possible, it would be important to validate the simulation outputs using data from human-in-the-loop experiments, and to adjust and re-run the simulation where necessary. Given the size, complexity, and cost of naval platforms including their equipment and personnel, the conduct of live experiments at the scope of the current simulation study is highly unlikely. However, data gathered from experiments focusing on one or more specific aspects of damage control and optimized crewing can still be of tremendous value. For example, an experiment may be conducted to study the impact of automation failure on crew activities including the actual time required for the crew to perform specific actions that the failed automation would have performed, and the actual variance in the time required.

Besides validation against empirical data, sensitivity analysis on key simulation parameters may be conducted to identify the ranges of input values over which the simulation study findings would remain unchanged. Then, even if a decision maker was not totally in agreement with or totally confident about the choices of input values used in the original simulation model, he/she could consider instead if what he/she believed or knew to be the true input values still fell within larger range of values over which the same conclusions could be drawn. In any case, steps had been taken to ensure that the input values used in the original simulation model were as realistic as possible; for example, all task timings in the current model had been validated by an experienced Marine Engineering Officer and an experienced Marine Engineering Operator (Petty Officer) who were employed in the Directorate of Maritime Ship Support. The development of the simulation model was also led by a retired naval officer (Lieutenant Commander) with 11 years of experience, who was command-qualified and trained in damage control.

Rather than thinking of the simulation model as a completely faithful representation of how an actual crew or an actual suite of DC automation would perform, it would be more appropriate to think of the simulation model as a decision aid intended to 1) make explicit knowledge or assumptions that were held implicitly by decision makers, 2) aid the integration and interpretation of these knowledge or assumptions, and 3) reveal gaps in knowledge or assumptions that would be needed to enhance future iterations of the model. The reality is that the reduced-size crews being considered have not yet been assembled, the DC automation being considered has not yet been acquired, and the ships that these crews and automation are intended to operate do not yet exist, so no empirical data on the performance of these overall systems could be available. Yet decisions on crew sizes and automation still need to be made in the absence of such empirical data. With the help of a simulation model, it would at least be possible to distinguish between more or less promising options given what the decision makers believe to be the capabilities held by the human operators and/or automated systems, to track what and how specific options have been considered, and to identify constraints associated with each option.

4.5 Future Research

Based on the findings and limitations of the current simulation study, there are several interesting directions that can be pursued in future research, including:

- Comparison of different automation types – this may take the form of automation for fire versus flood management, or "information" automation versus "action" automation (see [4] for detailed definitions);

- Comparison of different crew sizes for the same automation configuration;

- Sensitivity analysis on automation reliability as a key simulation parameter – instead of comparing only 100% automation reliability with 75% automation reliability, it may be valuable to explore a larger range of values and finer-grained comparisons between maximal and minimal values (e.g., 100%, 95%, 90% …. 50%) as this may help to determine a minimal acceptable value for automation reliability, or to determine a threshold value where the relative importance of automation level versus automation reliability begins to change; and

- Sensitivity analysis on other simulation parameters – e.g., completion time for tasks performed by crew members, completion time for tasks performed by automation, error rates for tasks performed by crew members, time penalties for completion of previously failed tasks, as well as the variances associated with these parameters.

Perhaps most importantly, it would be prudent to apply the simulation approach developed in the current study to investigate the impact of crew size and automation design on other (non-DC) naval functions. For example, simulation experiments can be performed to investigate optimized crewing for combat operations or combat systems engineering. In fact, it would be most important to simulate and compare the effectiveness of different crew and automation options on the operation of the entire ship, even if some or all of the functions may not be modelled in as much detail as was done for DC in the current study.

5 Conclusion

The current simulation experiment demonstrated that amongst the factors of automation level, automation reliability and scenario complexity, automation level appeared to have the highest impact on DC effectiveness. Full automation (with small crew size) consistently produced a high level of performance. In contrast, medium automation (with medium crew size) performed well in fire response but poorly in flood response.

It was important to consider both fire and flood both in the design of the DC scenarios and the selection of performance measures. Relative merits of the automation (and crew) configurations changed depending on whether a fire or flood-related measure was used. Instead of measuring and analyzing the large number of variables described in the original contract report [5] (i.e., 12 fire-related variables and 3 flood-related variables for a total of 15 variables), it was feasible and informative to analyze only four aggregate variables (i.e., 2 fire-related variables and 2 flood-related variables). In fact, it appeared that using only the two variables of total time to complete fire response and total time to complete flood response would be sufficient to identify all the significant effects found in this simulation experiment. This finding has the potential to simplify greatly the data collection and analysis for similar simulation experiments in the future.

In addition, although one of the original motives for the development of this simulation was to explore the feasibility of integrating IPME (which modelled crew and automation activities) with FFSIM (which modelled the propagation of fire and smoke), both of the dependent variables that were deemed most informative (i.e., total time to complete fire response and total time to complete flood response) were produced by IPME rather than FSSIM. This finding has the potential to simplify the development of similar simulations in the future, by de-emphasizing the criticality of real-time integration between the IPME and FSSIM modelling tools if the primary purpose of a study is to evaluate crew performance (e.g., in terms of task completion times) or workload. Of course, integration with FSSIM can still be tremendously useful to explore other (especially design-related) factors.

To inform decision making on the design or acquisition of future naval platforms, it would be most important to simulate different automation (and crew) options for other (non-DC) naval functions, and to develop integrated simulations that would enable comparison of automation (and crew) options for the whole ship. To support such a research effort, it would also be important to examine in greater detail different ways to define and compare different automation (and crew) options (e.g., by going beyond full, medium, or baseline automation, or by decoupling automation and crew size in future experiments).

References

[1] Hiltz, J. A. (2005). *Damage control and crew optimization* (TM2005-010). Dartmouth, NS: DRDC.

[2] Torenvliet, G., & Jamieson, G. A. (2007). *Functional modelling, scenario development and options analysis to support optimized crewing for damage control - Phase 1: Functional modelling (Revision A)* (CR2006-071). Toronto, ON: DRDC Toronto.

[3] Torenvliet, G., Jamieson, G. A., & Cournoyer, L. (2007). F*unctional modelling, scenario development, and options analysis to support optimized crewing for damage control - Phase 2: Scenario development* (CR2006-090). Toronto, ON: DRDC Toronto.

[4] Torenvliet, G., Jamieson, G. A., & Coates, C. E. (2008). *Functional modelling, scenario development, and options analysis to support optimized crewing for damage control - Phase 3: Options analysis* (CR2008-175). Toronto, ON: DRDC Toronto.

[5] Coates, C., & Cooper, C. (2009). *Crewing effectiveness: Full scale model development* (CR2009-048). Toronto, ON: DRDC Toronto.

[6] Floyd, J., Hunt, S., Williams, F., & Tatem, P. (2004). *Fire + Smoke Simulator (FSSIM), Version 1 - Theory manual* (NRL/MR/6180-04-8765).

[7] Chow, R., Hiltz, J., Coates, C., & Wang, W. (2010). Optimized crewing for damage control: A simulation study, *Maritime Systems and Technology (MAST) Americas 2010 Conference.* Arlington, VA.

[8] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.

[9] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, 30*(3), 286-297.

[10] Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors, 49*(1), 76-87.

[11] Dixon, S., & Wickens, C. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*(3), 474-486.

[12] Wickens, C., & Dixon, S. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201-212.

[13] Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*, 50-80.

[14] Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics. N*ew York: Harper & Row. (Chapter 9; more recent editions are available)

[15] Stevens, J. (1986). *Applied multivariate statistics for the social sciences. H*illsdale, NJ: Erlbaum.

[16] Pierce, C.A., Block, R.A. & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64*(6), 916–924.

[17] Cohen, J. (1992). Statistics a power primer. *Psychology Bulletin, 112*, 155–159.

[18] Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.).

[19] Dunnett, C. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association, 75*(372), 796-800.

[20] Grimm, L., & Yarnold, P. (Eds.). (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.

[21] Kinnear, P. R., & Gray, C., D. (2010). *PASW 17 statistics made simple*. New York: Psychology Press.

Annex A Data Tables for Dependent Variables

This annex contains the raw data for each of the four dependent variables that were analyzed.

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**
Scenario	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**
Run #										
1	732	1057	1236	1816	1688	1816	2261	1789	1868	1852
2	758	847	1000	1136	1462	1815	2315	1984	2226	2267
3	606	779	1180	801	1893	1828	1471	1757	2483	2273
4	793	999	1687	733	1813	1941	2760	2197	2010	2028
5	852	623	1289	2596	1557	1714	2111	2036	1964	2314
6	896	873	797	1190	1941	1482	1937	2117	2067	1804
7	624	856	1360	1867	1903	1588	1624	1430	2109	1490
8	577	962	1568	1736	1882	1823	1641	2575	2266	1834
9	718	490	1175	1503	1570	2133	2356	2151	1966	1864
10	842	719	1347	872	1605	2084	2794	2118	2362	1657
11	690	792	1486	1004	1766	2178	1694	1657	1838	2368
12	823	762	1290	1472	1750	1928	1941	2072	2058	2053
13	839	713	1766	1486	1474	1933	2465	1607	1905	2171
14	988	703	1125	1580	2067	1539	1756	1759	1545	1922
15	858	895	1813	1786	1909	2274	1838	2497	2090	1736
16	646	585	830	1095	1599	1514	2565	2439	1719	2286
17	825	1000	1604	1744	1986	1851	1727	2445	1896	2285
18	575	868	1069	1395	1795	1663	1720	2093	2271	1989
19	600	884	934	1626	2058	2066	2201	1840	1868	1772
20	1008	613	1875	1518	1765	1544	1766	1740	2368	1972
21	939	580	1404	1527	1799	1684	2482	1540	1957	2204
22	847	718	1351	1523	1783	1895	1571	2791	1665	1999
23	704	708	1040	793	2252	1982	1994	2858	1788	2321
24	937	706	1001	1667	2234	1518	1969	1777	2017	2460
25	665	819	812	1069	1872	1637	1863	2029	1594	1991
Mean	774	782	1282	1421	1817	1817	2033	2052	1996	2037
Std Dev	130	145	315	425	209	226	379	381	243	251

Table A-1: Fire response completion time (in seconds)

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**
Scenario	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**
Run #										
1	0	33	13	40	49	53	49	56	49	55
2	0	33	0	14	48	54	54	55	48	55
3	0	1	25	39	49	54	48	54	49	55
4	0	33	11	29	49	53	49	54	49	55
5	26	2	46	31	48	54	50	54	47	54
6	0	0	23	43	49	54	49	54	47	55
7	0	0	6	50	44	56	48	55	47	55
8	0	0	40	48	49	53	50	54	49	54
9	0	33	7	49	49	55	49	56	48	55
10	0	1	30	38	51	54	48	55	49	55
11	0	33	10	50	49	54	54	54	48	55
12	25	33	38	52	49	55	53	54	50	55
13	0	1	25	50	49	54	49	54	46	55
14	0	0	5	44	49	56	46	56	49	55
15	0	1	23	39	49	54	49	56	46	54
16	0	0	23	37	49	55	46	55	48	54
17	0	0	6	23	49	54	49	55	49	55
18	0	0	21	52	50	55	48	55	50	55
19	0	1	20	36	49	55	50	54	48	55
20	0	33	32	40	48	55	49	55	49	55
21	0	33	36	46	49	55	48	55	48	55
22	0	3	42	32	49	54	50	54	49	54
23	26	33	19	46	50	55	49	56	48	55
24	0	33	10	50	49	54	53	54	49	55
25	26	0	7	44	49	55	46	56	45	55
Mean	4.1	13.6	20.7	40.9	48.8	54.4	49.3	54.8	48.2	54.8
Std Dev	9.6	16.2	13.1	9.5	1.2	0.8	2.2	0.8	1.2	0.4

Table A-2: Number of compartments affected by fire

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**
Scenario	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**
Run #										
1	1963	1746	2182	1674	2101	2024	2302	1959	1287	1212
2	1829	2221	984	1875	2019	2029	2145	2133	1828	1963
3	2076	2193	1444	1878	2130	2148	1650	2082	1735	1709
4	2117	1582	899	1798	2153	2131	2182	2162	1557	1952
5	1876	1888	1333	1762	2134	1807	2407	2362	2139	1334
6	1939	1817	904	2395	1850	1950	2041	2098	1415	1756
7	1954	1807	1757	1689	2372	1818	1765	1700	1433	2007
8	1725	2148	1308	943	1957	2043	1997	2130	1789	1790
9	1997	2105	1714	1777	1903	2162	2069	2540	1616	1935
10	2145	1950	1554	1942	1719	2326	1965	1802	1690	2080
11	2347	1654	1554	1696	2266	2099	1718	1778	1538	2109
12	1912	1688	1865	1805	1830	2294	2021	2306	1774	1467
13	1711	1922	1840	930	2096	1892	2348	2033	2011	1844
14	1799	1748	2144	1201	2011	1999	2018	2261	1793	1633
15	1802	1527	1599	1437	2059	2176	2073	2456	1770	1753
16	1473	1624	1773	1648	1812	1585	2096	1991	2011	1781
17	1464	2168	1920	2200	1788	2263	2022	1725	1515	1999
18	1842	1797	1743	2200	1860	2237	2167	1739	1784	1717
19	1824	1820	914	1542	1885	1959	2080	1733	1737	1692
20	1728	2167	1141	2373	2158	2028	2049	1834	1660	1310
21	1355	1286	1898	1665	1864	1947	1543	1999	1628	1922
22	1863	1951	1683	956	1640	1985	2083	2068	1579	1880
23	1834	1924	1251	1143	1790	1920	2456	2077	1973	1875
24	1805	1621	1564	1677	2033	1861	2562	2096	1723	1991
25	1686	1345	1089	1879	1996	2140	2498	2378	1604	1626
Mean	1843	1828	1522	1683	1977	2033	2090	2058	1704	1774
Std Dev	217	256	382	408	177	172	253	239	200	239

Table A-3: Flood response completion time (in seconds)

Automation	Full	Full	Full	Full	Med	Med	Med	Med	Base	Base
Reliability	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**	**75%**	**75%**	**100%**	**100%**
Scenario	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**	**Med**	**High**
Run #										
1	2.51	2.23	2.79	2.14	2.69	2.59	2.95	2.77	1.64	1.54
2	2.34	2.84	1.25	2.4	2.58	2.6	2.75	2.73	2.34	2.51
3	2.66	2.81	1.84	2.4	2.73	2.75	2.11	2.66	2.22	2.18
4	2.71	2.02	1.14	2.3	2.76	2.73	2.79	2.77	1.99	2.5
5	2.4	2.42	1.7	2.25	2.73	2.31	3.08	3.03	2.74	1.7
6	2.48	2.32	1.15	3.07	2.37	2.5	2.61	2.69	1.81	2.25
7	2.5	2.31	2.25	2.16	3.04	2.32	2.26	2.17	1.83	2.57
8	2.21	2.75	1.67	1.2	2.5	2.61	2.56	2.73	2.29	2.29
9	2.56	2.69	2.19	2.27	2.44	2.77	2.65	3.26	2.07	2.48
10	2.75	2.5	1.98	2.48	2.2	2.98	2.52	2.3	2.16	2.66
11	3.01	2.11	1.98	2.17	2.9	2.69	2.2	2.27	1.96	2.7
12	2.45	2.16	2.39	2.31	2.34	2.94	2.59	2.95	2.27	1.87
13	2.19	2.46	2.35	1.18	2.68	2.42	3.01	2.6	2.57	2.36
14	2.3	2.23	2.74	1.53	2.57	2.56	2.58	2.9	2.29	2.09
15	2.3	1.95	2.04	1.83	2.63	2.79	2.65	3.15	2.26	2.24
16	1.88	2.07	2.27	2.11	2.32	2.02	2.68	2.55	2.57	2.28
17	1.87	2.78	2.46	2.82	2.29	2.9	2.59	2.2	1.93	2.56
18	2.36	2.3	2.23	2.82	2.38	2.86	2.78	2.22	2.28	2.2
19	2.33	2.33	1.16	1.97	2.41	2.51	2.66	2.21	2.22	2.16
20	2.21	2.78	1.45	3.04	2.76	2.6	2.62	2.35	2.12	1.67
21	1.73	1.64	2.43	2.13	2.38	2.49	1.97	2.56	2.08	2.46
22	2.38	2.5	2.15	1.21	2.1	2.54	2.67	2.65	2.02	2.41
23	2.35	2.46	1.59	1.45	2.29	2.46	3.15	2.66	2.53	2.4
24	2.31	2.07	2	2.14	2.6	2.38	3.28	2.68	2.2	2.55
25	2.16	1.72	1.39	2.4	2.56	2.74	3.2	3.05	2.05	2.08
Mean	2.36	2.34	1.94	2.15	2.53	2.60	2.68	2.64	2.18	2.27
Std Dev	0.28	0.33	0.49	0.53	0.23	0.22	0.32	0.31	0.26	0.31

Table A-4: Maximum height of flood water (in m)

List of symbols/abbreviations/acronyms/initialisms

ANOVA	Analysis of Variance
ARP	Applied Research Project
BH	Base Automation with High Reliability
CI	Confidence Interval
DC	Damage Control
DRDC	Defence Research & Development Canada
DV	Dependent Variable
DV1	Dependent Variable 1: Fire Response Time
DV2	Dependent Variable 2: Number of Compartments Affected by Fire
DV3	Dependent Variable 3: Flood Response Time
DV4	Dependent Variable 4: Maximal Floodwater Height
FL	Full Automation with Low Reliability
FH	Full Automation with High Reliability
FSSIM	Fire and Smoke Simulator
IPME	Integrated Performance Modelling Environment
MANOVA	Multivariate Analysis of Variance
ML	Medium Automation with Low Reliability
MH	Medium Automation with High Reliability
US	United States

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Toronto 1133 Sheppard Avenue West P.O. Box 2000 Toronto, Ontario M3M 3B9	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC-A REVIEW: CGEC JUNE 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Analysis of a simulation experiment on optimized crewing for damage control:

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Chow, R.

5. DATE OF PUBLICATION (Month and year of publication of document.) June 2012	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 57	6b. NO. OF REFS (Total cited in document.) 21

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Toronto TR 2010-128	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)
Unlimited distribution.

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))
Unlimited announcement.

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

In 2008, a simulation model was developed in the Integrated Performance Modelling Environment (IPME) to evaluate different crew-automation options for naval damage control. This previous work demonstrated the feasibility and value of applying modelling and simulation to explore a large number of factors related to optimized crewing for damage control, but stopped short of performing detailed statistical analysis on the simulation outputs. The current report re-examines the data collected from the 2008 simulation experiment and subjects them to formal hypotheses testing. In particular, it investigates the effects of automation level, automation reliability, and scenario complexity on damage control effectiveness, where damage control effectiveness was measured by time to complete fire response, number of compartments affected by fire, time to complete flood response, and maximal height reached by floodwater. The analyses compared three automation levels (full, medium, and the baseline) that were coupled with three crew sizes (small, medium and large, respectively), two levels of automation reliability (100% and 75%), and two levels of scenario complexity (high, medium). Of the studied factors, automation level was found to have the most significant impact on damage control. Full automation was found to perform best in terms of fire response. Both full automation and the baseline were found to outperform medium automation in terms of flood response. Based on these analyses, this report identified a number of strategies for streamlining future development of related simulation models, as well as future data collection and analysis for related simulation experiments. Finally, this report identified a number of directions for future research on the use of modelling and simulation to inform optimized crewing, including the evaluation of different crew-automation options for whole-ship operation.

En 2008, on a élaboré l’environnement intégré de modélisation du rendement (EIMP), un modèle de simulation servant à évaluer différentes formes d’automatisation des équipages aux fins du contrôle des avaries à bord des navires. Ces travaux ont démontré la faisabilité et la valeur de l’application de la modélisation et de la simulation à l’examen d’un grand nombre de facteurs liés à l’optimisation des équipages aux fins du contrôle des avaries, mais sans toutefois élaborer des analyses statistiques détaillées sur les produits de la simulation. Le dernier rapport publié examine à nouveau les données recueillies de l’expérience de simulation de 2008 et les soumet à des vérifications d’hypothèses. Plus précisément, les facteurs examinés sont les effets du degré d’automatisation, de la fiabilité de l’automatisation et de la complexité du scénario sur l’efficacité du contrôle des avaries; l’efficacité du contrôle des avaries étant mesurée en fonction du délai d’exécution de l’intervention en cas d’incendie, du nombre de compartiments touchés par l’incendie, du délai d’exécution de l’intervention en cas d’inondation et de la hauteur maximale atteinte par les dégâts d’eau. Les analyses ont permis de comparer trois degrés d’automatisation (complète, moyenne et de base) selon trois tailles d’équipage (respectivement restreint, moyen et nombreux), deux niveaux de fiabilité de l’automatisation (100 p. 100 et 75 p. 100) et deux niveaux de complexité du scénario (élevé ou moyen). Parmi les facteurs étudiés, on a constaté que le degré d’automatisation avait le plus grand impact sur le contrôle des avaries. On a trouvé que l’automatisation complète donnait les meilleurs résultats pour l’intervention en cas d’incendie. On a jugé que l’automatisation complète et l’automatisation de base donnaient un rendement supérieur à l’automatisation moyenne pour l’intervention en cas d’inondation. À partir de ces analyses, les auteurs du rapport ont énoncé un certain nombre de stratégies permettant de rationaliser l’élaboration de modèles de simulation connexes, ainsi que la collecte et l’analyse ultérieures de données aux fins d’expériences de simulation semblables. Enfin, les auteurs du rapport ont établi des pistes d’orientation des futurs travaux de recherche sur l’emploi de la modélisation et de la simulation pour documenter l’optimisation des équipages, y compris l’évaluation de différents scénarios d’automatisation de l’ensemble des fonctions du navire

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) optimized crewing; damage control; modelling and simulation; IPME

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca