Institution: University of Bristol

Unit of Assessment: 12) Engineering

Title of case study: Statistical testing validates new protection software for nuclear power stations

Period when the underpinning research was undertaken: 2001-2007

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Dr John May	Reader in Safety Systems	01/1996-present
Dr Silke Kuball	Research Fellow	02/1998-11/2007
Dr Luping Chen	Senior Research Associate	11/1997-present
Dr Gordon Hughes	Director of Safety Systems Research	06/1996-12/2002

Period when the claimed impact occurred: 1 August 2013 – 31 July 2020

Is this case study continued from a case study submitted in 2014? N

1. Summary of the impact

Statistical Testing (ST) of systems, developed by Dr May and his team at the University of Bristol, has provided a scientific basis for increased confidence in new control & instrumentation (C&I) software for the UK nuclear energy sector. Crucial to the future of nuclear power, the software performs critical safety functions, and its reliability must be demonstrated. Traditional techniques are unable to do this, thus May and his team developed ST to fill the gap and provide the key evidence necessary to support the licensing of this increasingly complex software. During the REF2021 impact period, ST has: (i) become a key component in the governance and regulation of critical nuclear systems, including within Office for Nuclear Regulation guidance and nuclear licensees' practice; (ii) been incorporated into EDF infrastructure and expertise; and (iii) stimulated nuclear C&I policy analysis and review in national and international government and industry bodies.

2. Underpinning research

For all nuclear technologies, and particularly after events such as Three Mile Island, Chernobyl and Fukushima, independent and robust demonstration of safety is paramount, and is required to obtain an operating licence. New nuclear plants use significantly more programmable (software based) digital C&I than previous generations. Such C&I is more complex, and it is relied on to a greater extent to keep the public safe. However, traditional random failure reliability models, originally developed for hardware, cannot be used to assess programmable systems, because the design (also called 'systematic') software failures exhibit a different type of randomness.

Dr John May and his research group at the University of Bristol (UoB) have advanced ST to provide a new approach to the reliability assurance of programmable digital systems. It is a form of what the industry calls 'software substantiation' and is unique in two respects: it applies the scientific method (it is empirical testing of software design), and it generates a scientifically defensible reliability estimate. It therefore provides very strong support for regulator confidence in critical software.

The research was conducted within UoB's South West Nuclear Hub (https://southwestnuclearhub.ac.uk/) , for the most part funded by the C&I Nuclear Industry Forum (CINIF), the UK's leading body responsible for advancing nuclear C&I safety in the UK. CINIF funded the research, led by Dr May, within the Hub's "System Reliability, Resilience, Security & Risk" theme, continuously between 2001-2017. Papers [1]-[6] all derive from this programme, the results of which are disseminated to CINIF members (UK nuclear licensees and regulators) [A] [C].

Specifically, from 2001 to 2004, May, Kuball and Hughes showed how ST tests can be defined to simultaneously satisfy reliability targets and traditional test adequacy (coverage) criteria [1]. This was important because code coverage was and remains an important aim of software testing. The ST was applied to a software system for turret control which had already been tested using industry standard techniques. This found all faults uncovered by the standard techniques and additional faults not discovered previously, demonstrating for the first time the power of ST in a real application [2], and persuading nuclear engineers in industry and the regulator of its power and value.

Between 2004 and 2007, the research evolved to accommodate a trend in C&I, namely, increasing use of ‘smart devices’ with embedded software. May and Kuball were the first to solve the underlying research problems and hence demonstrate ST in software substantiation of smart devices, thereby unlocking a scientifically defensible method for reliability assessment [3] [4]. ST samples the set of possible challenges or ‘demands’ on a system, to estimate the proportion which fail, but the research problems did not lie in the inference statistics. They concerned the need for 1) new stochastic system environment simulation models (e.g. of reactor accident transients), and 2) new ways to define system tests to achieve random sampling (the software test methods existing at the time did not satisfy ST requirements). Both were needed to perform ST on C&I systems, and on smart devices in particular, which is why the solutions have been able to underpin the actions of organisations performing ST, government bodies initiating ST policy synthesis, and regulators advocating ST in the guidance for assessing these systems.

During 2014-2016, May and Chen advanced understanding of the reliability efficacy of software failure diversity in systems with redundant channel architectures [5]. Without failure diversity, the software in the channels has the potential to defeat the most fundamental system reliability design technique available to C&I engineers, namely, redundancy. Again, this is because, in contrast to hardware, software fails systematically; broadly speaking, if each channel contains identical software, then when one channel fails due to a software fault, all channels fail coincidentally. Paper [5] gives the fundamental theory underpinning demonstration of failure diversity by injecting faults in diversely coded software channels.

Finally, May and Chen produced (2015-2017) a world-first model capable of estimating system reliability based on software executing on simulated hardware platforms [6]. The importance of this is that simulation raises the possibility of massively concurrent testing (hence massive test acceleration) and is capable of breaking current limits for reliability demonstrations proposed in the literature. Equally important, it also allows design-time estimation of reliability, by moving much of the testing prior to hardware platform availability. This is a crucial advantage because C&I testing often comes at a later stage in plant build schedules and it is necessary to de-risk the possibilities for finding faults at this stage.

3. References to the research

1. **Kuball S, May J** (2004). Test-Adequacy and Statistical Testing: Combining Different Properties of a Test-Set, *15th International Symposium on Software Reliability Engineering (ISSRE)*, pp.161-172 https://doi.org/10.1109/ISSRE.2004.40
2. **Kuball S, Hughes G, May J,** Gallardo J, John A (2004). The effectiveness of statistical testing when applied to logic systems, *Safety Science*, **42**, pp.369-383 https://doi.org/10.1016/j.ssci.2003.09.006
3. **Kuball S, May J** (2006). Building statistical test cases for smart device software, *Institution of Engineering and Technology International Conference on System Safety*, **1**, pp.269-274 https://doi.org/10.1049/cp:20060227
4. **Kuball S, May J** (2007). A discussion of statistical testing on a safety-related application, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **221:2**, pp.121-132 https://doi.org/10.1243/1748006XJRR43
5. **Chen L, May J** (2016). A Diversity Model Based on Failure Distribution and its Application in Safety Cases, *IEEE Transactions on Reliability*, **65:3**, pp.1149-1162 https://doi.org/10.1109/TR.2015.2503335
6. **Chen L, May J** (2017). Theoretical Feasibility of Statistical Assurance of Programmable Systems Based on Simulation Tests, *IEEE International Conference on Software Quality, Reliability and Security Companion, QRS-C*, pp.630-631 https://doi.org/10.1109/QRS-C.2017.123

Funding Information

- **May J** (PI), *Improving C&I Design for Testability* and *Reliability Limits of Programmable Protection Systems* (two projects), BEIS/Fraser Nash Consultancy, 2019-2021, GBP350,000
- **May J** (PI), *DDT* and *DISPO* projects, CINIF, 2000-2020, GBP1.6 million
- **May J** (PI), *SCARLETT*, European Commission FP7 project 211439, 2008-2011, EUR200,000 (EUR43M project overall, led by Airbus & Thales with 39 partners)
- **May J** (PI), *Re-use of reliable software components*, EPSRC EP/D061652/1, 2007-2009, GBP117,887
- **Hughes G** (PI), **May J** (CI), *SSRC Generic Research Programme*, Lloyds Register, British Energy, Railtrack, Health & Safety Executive, 1997-2002, approximately GBP225,000

4. Details of the impact

In 2019, nuclear accounted for 18.4% of the UK energy supply (https://www.ofgem.gov.uk/publications-and-updates/infographic-energy-security) . There are 15 operating plants, at 8 sites, all operated by licensee EDF Energy (http://www.onr.org.uk/civil-nuclear-reactors/index.htm; https://www.edfenergy.com/about/nuclear). The UK intends to maintain a significant level of nuclear power production to provide a baseline carbon neutral energy source for the UK. This will require the construction of new plants, possibly including small/advanced modular reactor designs, as well as modifications to extend the life of existing plants. In all cases, software C&I plays an important safety role, hence the significant impact that UoB's ST research has had on policy described in sections 4.1-4.3. In REF 2014, Dr May and his team articulated the impact of their pioneering use of ST to substantiate key critical control and instrumentation (C&I) systems, such as protection (shutdown) systems. Sections 4.1-4.3 describe new impacts for REF 2021.

4.1 Changes to regulatory policy and safety guidance lead to improved confidence in the safety of UK nuclear systems

The Superintending Inspector for Nuclear Safety at the Office for Nuclear Regulation (ONR) stated: "*The work of Dr May's research group on ST […] made important contributions to our understanding that ST is effective in terms of more familiar test adequacy criteria (statistical testing is not a traditional or common form of software testing) and to demonstrate how it can be applied to smart devices. With its capability to provide a scientifically defensible reliability estimate for software systems, it has enhanced scientific rigour in the assurance of complex programmable electronic and software-based systems important to safety. It has therefore provided the nuclear industry sector with a method of demonstrating high quality additional evidence in support of licensing nuclear plants and/or modification to safety systems containing increasingly complex C&I systems that can be assessed by ONR. This research has been a major factor supporting ONR's decision to recommend the use of ST as part of the measures to justify the design of complex C&I safety systems and equipment in guidance, namely Technical Assessment Guide (TAG) NS-TAST-GD-046 revision (commonly referred to as TAG 046), and therefore within nuclear licensees' practice. In addition, it was incorporated into ONR's Safety Assessment Principles (SAPs) that form our highest level of technical policy that reflects relevant good practice in the nuclear sector…*" [A]

The ONR publications rely mainly on standards, and rarely contain academic references (although TAG 046 does cite paper [5] as a specialist research source). It also references international standard IEC61508-7 "Functional safety of electrical/electronic/programmable electronic safety-related systems", which cites [4] on this topic in Section C5.1. Dr May has also provided direct input to TAG 046, most recently in 2018, giving guidance to ONR on ST and coding practices to support ST, which was included in clauses 9.26 and 10.26-10.29 [A]. As a result of his contributions, in 2015, Dr May was invited to sit on the ONR Chief Inspector's Independent Advisory Panel to advise on C&I policy and remains a member to date [A].
The impact of ST resulting from its inclusion in the SAPs (from 2014), TAG 046 and the IEC61508-7 is exemplified by two important changes in the nuclear landscape:

a. <u>The UK New Nuclear Build (NNB) programme</u> – All new projects in this programme will follow the ONR guidance regarding ST. EDF's NNB 2017 plans for assurance of C&I for their EPR reactor at the new Sizewell C and Hinkley C nuclear plants states: "*There is a statistical testing programme for the HPC EPR Protection System (covering the Protection System Core as well as the Protection System Interface Units) […] It is proposed to run 50,000 tests on the final software release, which will provide a 99% confidence level…*" [B]. Similar plans were created for assurance of Hitachi's protection system on their ABWR in 2015 [C], and Westinghouse's AP1000 reactor protection in 2017 [D].

b. <u>The widespread use of Smart Devices in new nuclear systems</u> – Research conducted by May and Kuball [3] [4] was the first to show how to perform ST on smart instruments, enabling the ONR to extend their guidance to include the use of ST on this type of equipment. This policy is a major increase in the scope of application of ST, to a significantly broader class of C&I systems. Although these devices were used occasionally in existing reactor designs (often to replace obsolescent sensors), they are used extensively in new reactor designs, including in safety functions. May's work has increased the scientific rigour used in the assurance of such equipment [A, B]. "*ONR now considers that the application of ST principles [forms] a key part of its regulatory expectations and is increasingly seen as relevant good practice in GB to provide assurance of safety for Class 1 (the most safety significant) and Class 2 safety systems that can incorporate smart devices in accordance with BS IEC 61226. The use of ST in this context also forms a part of the requirements within relevant international standards, including BS IEC 60080 and BS IEC 61508.*" [A]. In response to this impact on UK regulation, EDF have stated how May's research has informed their own policy on smart device testing in new plants and creation of a smart device test harness. *"[EDF] have invested considerable internal resource on this due to the importance we believe statistical testing will have in future reliability assurance of smart devices and thus in support to safety cases"*. [B]

ST has now appeared in assurance plans of nuclear companies [B, C and D], demonstrating the impact of ST on the nuclear industry.

4.2 ST incorporated in EDF infrastructure and expertise

On May's research, EDF state: "*The UK intends to maintain a significant level of nuclear production as a carbon neutral energy source, including construction of new plants and modifications to extend the life of existing plants. In both cases, software control & instrumentation (C&I) plays an important safety role. Statistical Testing (ST) has given us key additional evidence needed for regulatory approval of these increasingly complex systems.*" [B] As a result, EDF have engaged in development activities to support an increased role for ST.

May and Kuball's research on performing ST on smart devices [3] [4] motivated CINIF to sponsor the construction of a smart device test harness: "*New nuclear plants will make much greater use of smart devices in safety-related functions, compared to older generations of nuclear plants. […] smart device substantiation is expected to be required on a large scale into the future. Statistical testing is a technique that will be required to support this. The demonstration of feasibility of conducting ST on smart devices […] led CINIF to sponsor the construction of a smart device test harness. This is a tool designed for industry use, to make it easier to perform ST on a wide range of devices/systems. Development of this test rig has recently been completed and is intended to be made available to all CINIF licensees. […] EDF Energy NG took a leading role in this development and we have invested considerable internal resource on this due to the importance we believe statistical testing will have in future reliability assurance of smart devices and thus in support to safety cases. It is intended to interface with efforts underway at AMRC [Advanced Manufacturing Research Centre] on a smart device test rig, this was funded by BEIS and Dr. May contributed to the workstream associated with this*" [B]. EDF Energy also created a new role to lead ST activities within the company. The role is currently held by Dr Silke Kuball, previously a postdoctoral researcher in May's group at UoB [B]. ST also finds application in existing (cf. new) nuclear stock. In 2014, EDF Energy published a paper detailing their use of ST on their replacement for the obsolescent Data Processing System operator alarm system at the Dungeness AGR (Gough & Kuball 2014), part of plant life extension, which cites references [2] and [4] and formed part of the substantiation of this system [B].

4.3 Stimulated review of nuclear safety policy internationally

ST is a significant departure from traditional practice in the nuclear sector and is only recently establishing itself in the industry and with regulatory bodies. Governments and industries with nuclear capability are commissioning programmes of investigation into ST, to inform policy.

- The United States Nuclear Regulatory Commission (NRC) commissioned Brookhaven National Laboratory to evaluate "Development of Quantitative Software Reliability Models for Digital Protection Systems of Nuclear Power Plants" (2013; [E]). This cites [2] and [4], and led to further joint work by Brookhaven, Idaho National Laboratories, and the NRC that produced the US NRC report: "Development of A Statistical Testing Approach for Quantifying Safety-Related Digital System on Demand Failure Probability" (2017; [E]), referenced two of May's papers, including [1].
- In South Korea, the Korea Atomic Energy Research Institute, and Korea Hydro & Nuclear Power Co. are pursuing this field, e.g. [F] which progresses the topic of practical simulation-based test environments, citing [4].
- The Nuclear Power Institute of China, a subsidiary of China National Nuclear Corporation, started, in 2020, to investigate the possibility of using ST for assessing software reliability. [G]

In the UK, the ONR's concerns about smart device substantiation, and its subsequent decision to roll out ST on smart devices based on May's research [3] [4], contributed to NIRAB's conclusions in 2016 to include C&I and C&I safety in its list of areas identified for development to support the UK nuclear industry [A, H]. This led to ongoing advisory work (led by Dr May), commissioned for the BEIS Nuclear Innovation Programme, into future techniques for the sector, including platform simulation (digital twin) based testing [6], and diversity analysis for redundant channels [5] [I].

5. Sources to corroborate the impact

A. ONR – Corroborating statement (2020), Superintending Inspector for Nuclear Safety; SAPs and TAG-046, [Accessed 13/5/20]
B. EDF – Corroborating statement (2020), EDF Energy Fleet Lead C&I / C&I Nuclear Industry Forum (CINIF) Chair and Technical Lead Smart Device Testing and Qualification; HPC PCSR3: Chapter 7: Instrumentation and Control (2017; p.36) [Accessed 11/12/20]
C. Hitachi-GE Nuclear Energy Ltd, UK ABWR Generic Design Assessment (2015), see 'Action 2 – statistical testing', [Accessed 1/12/20]
D. ONR – Assessment Report of Westinghouse AP1000 (2017), see p.20, para 57, and search 'statistical testing' for more references [Accessed 1/12/20]
E. US NRC Reports: NUREG/CR-7044 (October 2013); and NUREG/CR-7234 (2017) [Accessed 11/12/20]
F. Lee et al. (2018). Development of simulation-based testing environment for safety-critical software, *Nuclear Engineering and Technology*, **50:4**, https://doi.org/10.1016/j.net.2018.02.007
G. Cai, Y et al. (2020). Quantitative software reliability assessment methodology based on Bayesian belief networks and statistical testing for safety-critical software, *Annals of Nuclear Energy*, **145**, https://doi.org/10.1016/j.anucene.2020.107593
H. NIRAB, UK Nuclear Innovation and Research Programme Recommendations (2016); see p.47, section R3.9, [Accessed 13/5/20]
I. Fraser Nash – Corroborating statement (2020), Group Leader, Electrical, Control and Instrumentation