European Commission

JRC CONFERENCE AND WORKSHOP REPORT

Advances in Reliability, Risk and Safety Analysis with Big Data: Proceedings of the 57th ESReDA Seminar

Hosted by the Technical University of Valencia, 23-24 October, 2019, Valencia, Spain

Ferreira, L. A., Simola, K., Kopustinskas, V

2020

ESReDA
European Safety, Reliability & Data Association

Joint Research Centre

EUR 30415 EN

EU Science Hub
https://ec.europa.eu/jrc

JRC122223

EUR 30415 EN

PDF	ISBN 978-92-76-23807-2	ISSN 1831-9424	doi:10.2760/836400

Luxembourg: Publications Office of the European Union, 2020.

How to cite this report: Ferreira, L.A., Simola, K. and Kopustinskas, V., Advances in Reliability, Risk and Safety Analysis with Big Data: Proceedings of the 57th ESReDA Seminar, EUR 30415 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-23807-2, doi:10.2760/836400, JRC122223.

European Safety, Reliability & Data Association

(ESReDA)

European Safety, Reliability & Data Association (ESReDA) is a European Association established in 1992 to promote research, application and training in Reliability, Availability, Maintainability and Safety (RAMS). The Association provides a forum for the exchange of information, data and current research in Safety and Reliability.

ESReDA membership is open to organisations, privates or governmental institutes, industry researchers and consultants, who are active in the field of Safety and Reliability. Membership fees are currently 1000 EURO for organisations and 500 EURO for universities and individual members. Special sponsoring or associate membership is also available.

For more information and available ESReDA proceedings please consult:
http://www.esreda.org/

Table of contents

Preface……………………………………………………………………..…..…………1

System health monitoring with deep learning: Is big data all we need?.................................2

Recommended practice for assurance of data-driven algorithms and models ……………..…4

Fleet-based remaining useful life prediction of safety-critical electronic devices ……...…...12

Assessing GB train accident risk using red aspect approaches to signal data………...……...20

Quality in big data for the forced unavailability of power plants……….……………..……..33

Mechanical seal failure prediction in an oil refinery: a first attempt to solve the problem using a data-driven approach (abstract) ...……………………………….……………….…...66

Trust in AI: The role of hybrid and private governance ..…………………………..…...…...68

Advanced analysis of reliability and risk of equipment subjected to degradation and obsolescence……………………………………...…………………………….…………….88

Robust statics for (big) Data Analytics (abstract)….………...……………..….……...……...96

A proposal of an algorithm to simulation right censored data type I in reliability field (abstract)……………………………………..…………………………………...……...99

IOT future in energy industries (abstract)…………………………………...………..…..…100

List of authors…………………………………………………………………………….101

Appendix 1: Slide presentations

Appendix 2: Seminar Programme

Preface

The 57th Seminar organized by ESReDA took place in the very beautiful city of Valencia, Spain. We were very kindly received by the Polytechnic University of Valencia/Universitat Politècnica de Valencia.

The Seminar was jointly organized by ESReDA and CMT Motores Termicos, a research unit at the Polytechnic University of Valencia. A sincere thanks is due to Professor Bernardo Tormos for the way he received us and allowed the Seminar to precede according to our best expectations.

In accordance with the theme proposed for the Seminar, communications were presented that made it possible to discuss and better understand the role of the latest big data, machine learning and artificial intelligence technologies in the development of reliability, risk and safety analyses for industrial systems.

The world is moving fast towards wide applications of big data techniques and artificial intelligence is considered to be the future of our societies. Rapid development of 5G telecommunications infrastructure would only speed up deployment of big data analytic tools. However, despite the recent advances in the these fields, there is still a long way to go for integrated applications of big data, machine learning and artificial intelligence tools in business practice.

We would like to express our gratitude to the authors and key note speakers in particular and to all those who shared with us these moments of discussion on subjects of great importance and topicality for the members of ESReDA.

The editorial work for this volume was supported by the Joint Research Centre of the European Commission in the frame of JRC support to ESReDA activities. A special thanks is due to Ariane Liessens (JRC) for the editorial work.

Prof. Luís AndradeFerreira
University of Porto, Portugal

Dr. Vytis Kopustinskas, Dr. Kaisa Simola
EC Joint Research Centre

System health monitoring with deep learning: Is big data all we need?

Olga Fink
ETH Zürich

Abstract

Deep learning and artificial intelligence have achieved some impressive successes in many different domains. These technologies applied in computer vision, natural language processing and speech recognition are gradually becoming an integral part of people's everyday lives. Concurrently, increasing amounts of condition monitoring data have been captured by complex industrial assets in many different industries, including energy and transportation. But what is the potential of these emerging technologies and "big data" for system health monitoring? Where could the development go and how could potential transformation look like? How could deep learning and artificial intelligence be integrated in decision making of asset managers supporting them to make optimal decisions?

While large amounts are captured on the condition of complex systems, faults in safety critical systems are rare. The diversity of the fault types and operating conditions makes it often impossible to extract and learn the fault patterns of all the possible fault types affecting a system. Consequently, faulty conditions cannot be used to learn patterns from. Supervised learning that has made most of the current success of AI possible is, therefore, often not applicable in the context of system health monitoring.

Even collecting a representative dataset with all possible operating conditions can be a challenging task since the operating conditions of complex industrial systems are highly varying. Therefore, training samples captured over short time periods may not be representative for the entire operating profile. Collecting datasets over longer time periods could capture more operating profiles and, thereby, improve the representativeness. However, that would also delay the point in time when the data-driven health monitoring systems could be taken into operation. Such delays are often not acceptable in real applications. We need, therefore, to rely on datasets that are not fully representative and have either none or very few faults and integrate the missing information either from other units of a fleet or from physical performance models.

The talk will give some insights into current challenges and highlight potential solutions for fault detection, isolation and the prediction of the remaining useful lifetime in the context of rare faults and a high variability of operating conditions. The talk will particularly present solutions in the field of domain adaption, transferring models between different units of a fleet and between

different operating conditions, and hybrid approaches fusing physical performance models and deep neural networks.

References:

Fink, Olga, Qin Wang, Markus Svensén, Pierre Dersin, Wan Jui Lee, and Melanie Ducoffe. 2020. "Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications." Engineering Applications of Artificial Intelligence 92 (June): 103678. https://doi.org/10.1016/j.engappai.2020.103678.

Recommended Practice for Assurance of Data-driven algorithms and models

Siegfried Eisinger, Jon Arne Glomsrud & Justin Fackrell
DNV GL. Norway

1. Introduction

Systems incorporating Machine Learning, and other data-driven techniques, have become increasingly widespread in recent years. Large organizations, such as Google, Alibaba and Amazon have capitalized on the capabilities of such techniques to do things better, do things faster, and/or do things previously impossible.

In industrial contexts too, data-driven techniques are being used in a variety of applications including:

- Early detection of failures (before they happen) and Condition-Based Maintenance (CBM)
- Semi- and fully-automated technical verification
- Simplifications and effectiveness improvements in class services (currently carried out manually by surveyors)

The goal of many data-driven applications is to facilitate decision support for an end user. End users are often people and/or organizations, but they can also be autonomous agents. The data from which data-driven models are derived is often very large and complex, and the models themselves are also complex. This complexity leads to challenges in establishing trust on the part of the user that a given data-driven system will perform as required. In a Recommended Practice (RP), DNV GL addresses these challenges by defining a framework for establishing trust in data driven models. The RP is currently in a draft stage and will be published end of 2019.

The goal of the RP is to help readers answer the following question: How can the user of an application which is based on a data-driven model be sure that the model is suitable for the intended use, and that the risk of failure and/or bad output is kept within tolerable limits?

To date no widely-recognized standard exists for assessment of a data-driven model. The current RP provides a structured approach for performing such an assessment. It makes use of CRISP-DM (The Cross-Industry Standard Process for Data Mining), a de-facto standard process for data-driven modelling. CRISP-DM was developed by an industry consortium in the ESPRIT European project.

Note that throughout the paper, we often use the word ‘algorithms’ as an abbreviation for ‘data- driven algorithms and models’.

2. Approach

Assurance of algorithms concerns the verification and validation that the algorithm is suitable for its intended use and in a wider context it also includes whether the intended use is according to rules, regulations and societal agreements. It should thus be possible to check any outcome of the algorithm with respect to whether it is acceptable or not. Moreover, it should be possible to predict all relevant algorithm outcomes, so that they can be checked. Obviously, full prediction and check of all possible outcomes is only possible for the simplest algorithms and for any real-world algorithm we are left with checking sub-sets. As algorithm complexity increases it gets more and more difficult to claim any sort of test completeness. Therefore, to base an assurance regime on the test of a finished algorithm becomes questionable. Even if it was possible to ensure a good test coverage, decisions during earlier stages in the lifecycle might have been flawed, leading to errors in the test specification and thus to faults which cannot be uncovered by the test. This leads to the statement

Assurance of complex systems cannot be based on verification and validation testing alone

Effective verification can only be performed if the scope is sufficiently limited. In addition, the stages of the lifecycle must be taken into account, ensuring that the output of one stage does not invalidate the results from previous phases, but supports the general goal. For example, even a perfect system which is put into operation might become flawed when maintenance is not performed in a correct way. Or, even a perfect design process will lead to a flawed product if the initial specification is flawed. This is the reason why many standards for assurance of complex systems adopt the strategy

Quality Assurance of complex systems should be done through quality assurance of all relevant lifecycle stages

This strategy has been implemented in for example IEC 61508 /1/ and other similar standards for reliability of critical systems.

One might object that some systems can be checked with 100% coverage through so-called Formal Proofs: Given that the basic specification is correct it can be checked that the system behaves completely according to that specification. It is true that such systems in principle can be approved through only testing, but systems which have a complete formal proof are seldom. Beyond that, performing a bad quality development process would most probably reveal a lot of errors through the formal proof, rendering the development process inefficient because of the error correction efforts they require. It would therefore most probably be most efficient to perform the suggested QA steps throughout the lifecycle and arrive at a work product with few errors, confirmed through formal proof.

Lifecycle schemes are often linked to Risk Analysis. Through the initial risk analysis the integrity requirements of the system are determined. Low integrity requirements would lead to less stringency with respect to QA rules than high integrity requirements. This approach has merits, rendering the life-cycle process into a ‘risk-based’ quality approach, focusing on critical issues and systems, while less critical issues and systems are treated in a more basic way. In IEC 61508 the first risk analysis is placed at the start of the lifecycle process after the concept is specified and it is used to perform the so-called Safety Integrity Level (SIL) allocation. The SIL is a four-level system linking the safety functions to the expected reliability with which it shall be delivered. The SIL is in turn linked to QA requirements throughout the lifecycle, their rigor increasing with increasing SIL. Table 1 quotes Table A.5 of /1/ for Software module testing and integration as one example, clearly demonstrating the basic requirement and the rigor increasing with the SIL.

Table 1 Copy of Table 4.5 of /1/. Software module testing and integration. R: 'Recommended', HR: 'Highly Recommended'

	Technique/Measure	SIL 1	SIL 2	SIL3	SIL 4
1	Probabilistic testing	-	R	R	R
2	Dynamic analysis and testing	R	HR	HR	HR
3	Data recording and analysis	HR	HR	HR	HR
4	Functional and black box testing	HR	HR	HR	HR
5	Performance testing	R	R	HR	HR
6	Model based testing	R	R	HR	HR
7	Interface testing	R	R	HR	HR
8	Test management and automation tools	R	HR	HR	HR
9	Forward traceability between the software design specification and the module and integration test specification	R	R	HR	HR
10	Formal verification	-	-	R	R

Accepting the two above statements, that assurance cannot be based on testing alone and that quality must be taken into account throughout the lifecycle our suggestion for an assurance regime for data-driven algorithms and models is based on a strategy similar to IEC 61508. The lifecycle model for data-driven algorithms and modules' is not identical to general software development, where often the V-model or a variant thereof is employed. For data mining projects the Cross-industry Standard Process for Data Mining (CRISP-DM) has been defined by an industry consortium in the ESPRIT European project (see /2/). This process has been accepted as a de-facto standard in the area, leading to the decision that it is also used in the present RP as basis.

3. The RP

Our framework for assurance of Data-driven Algorithms and models is due to be presented in full detail in the Recommended Practice DNVGL-RP-0510/3/, but the key concepts are presented here. The framework is structured around the CRISP-DM reference model/2/, as motivated in the previous section. The lifecycle model is structured along both a linear development process and explicit feedback stages both on a local level (e.g. between 'data preparation' and 'modelling') and on a more global level (e.g. between 'business understanding' and 'evaluation'). Compared to the software development V-model, the CRISP-DM model includes more relations to the proper understanding of the basics and explicit evaluation stages. The main reason for that might be that CRISP-DM was originally designed for business supporting tools, as opposed to critical applications. On the other hand, also for critical applications it is important to establish a correct 'problem understanding' and 'data understanding'. Before deployment a proper verification and validation phase must be included (i.e. 'evaluation'), demonstrating that the basic lifecycle steps seem to work for a wide area of applications.

The CRISP-DM process can be used in any assurance project regardless of the explicit usage of CRISP- DM during product development. The CRISP-DM workflow was developed in the late 1990's and has gained wide acceptance and popularity in the field of data science and machine learning, which was the main reason why we have decided to re-use the related lifecycle process.

Figure 1 Phases of the CRISP-DM reference model (ref. /2/, Figure 2)

Trust in an algorithm is established through the construction of an assurance case: this is a structured document containing detailed claims. Associated with each claim is one or more items of evidence which together support that claim. For example, in the activity data preparation, the claim "the data

is representative" may be supported by evidence showing that "the data is drawn from a group of 15 assets over 5 years, plus analysis confirming that this sample satisfactorily covers the feature space".

CRISP-DM offers more than the reference model of Figure 1. Each of the lifecycle stages is supported by sub-sections in two levels. In Figure 2 the sub-sections of Business Understanding are shown.

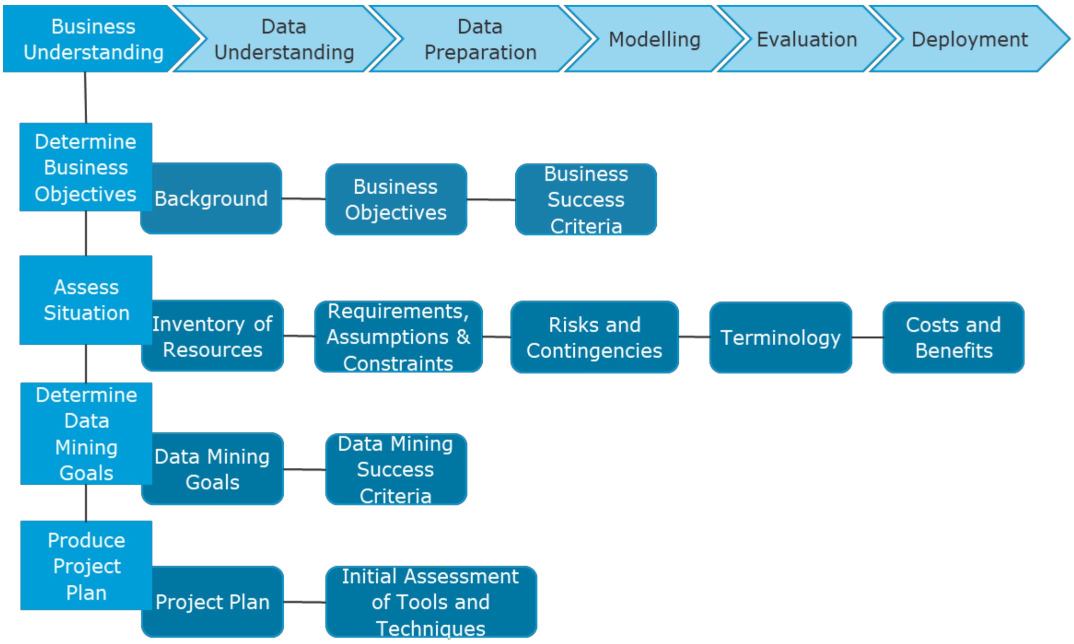


Figure 2 CRISP-DM Section on Business understanding with sub-sections

When constructing the claims, the sub-sections of CRISP-DM were re-used, giving our RP the same structure as CRISP-DM. In addition, Guidance notes were added helping the users to understand the meaning of the claim and giving clues on what evidence might be suitable. For the section on 'Determine Business Objectives', the emerging Claims and Guidance is shown in Figure 3.

3.1.1 Determine Business Objectives

1) The relevant background and business information are sufficiently documented.
2) Business Objectives are clearly defined and understood.
3) Business Success Criteria are sufficiently and objectively defined.

Guidance note:

Sufficient understanding of the business case, and clear and concise objectives and success criteria are important for a successful project. It may be helpful to formulate a use case, user stories and/or a value proposition. This helps focus attention on the project goal, and facilitates communication between the stakeholders.

In the first iterations, early in the development, it may not be possible to succinctly define what will be predicted, but completing a value proposition statement as early as possible in the project process will reduce misunderstanding and bring focus to the data preparation stage

An example of a value proposition (using Geoffrey Moore's suggested template, see Sec.1 [1.5.3]) is shown in Figure 2-1 A value proposition should address the following issues:

— who or what will benefit or make use of the result?
— what problem or need will be met?
— what should be developed to meet this need?

An alternative or supplement to a value proposition is a collection of one or more use cases which give concrete examples of the use of the model in practice. These bring focus to discussions concerning the business understanding and are useful aids in communication between the different participants and stakeholders in the development process.

---e-n-d---o-f---g-u-i-d-a-n-c-e---n-o-t-
e---

For <who?>
who <do what?>,
our product is a <what?>.
Unlike <competing solution>,
our product <does what?>

Figure 3-1 Value Proposition template suggested by Geofrey Moore

Figure 3 Claims and Guidance note for the sub-section 'Determine Business Objectives. Extract from /3/

The evidence must obviously be provided on a per -project basis. To facilitate this process, the content of the RP has been translated into the Argumentation tool REASON/4/. This tool can be used in an interactive fashion between all stakeholders. The provider of the algorithm can perform an initial self-assessment, entering all evidence which seems relevant. An independent assessor can then check the provided evidence and enter scores on whether the evidence is sufficient, or what other evidence is needed. Relevant parts of the assessment can be re-run when anything in the algorithm, the basic data or the usage area changes. Figure 4 shows an extract of the RP as it is implemented in REASON.

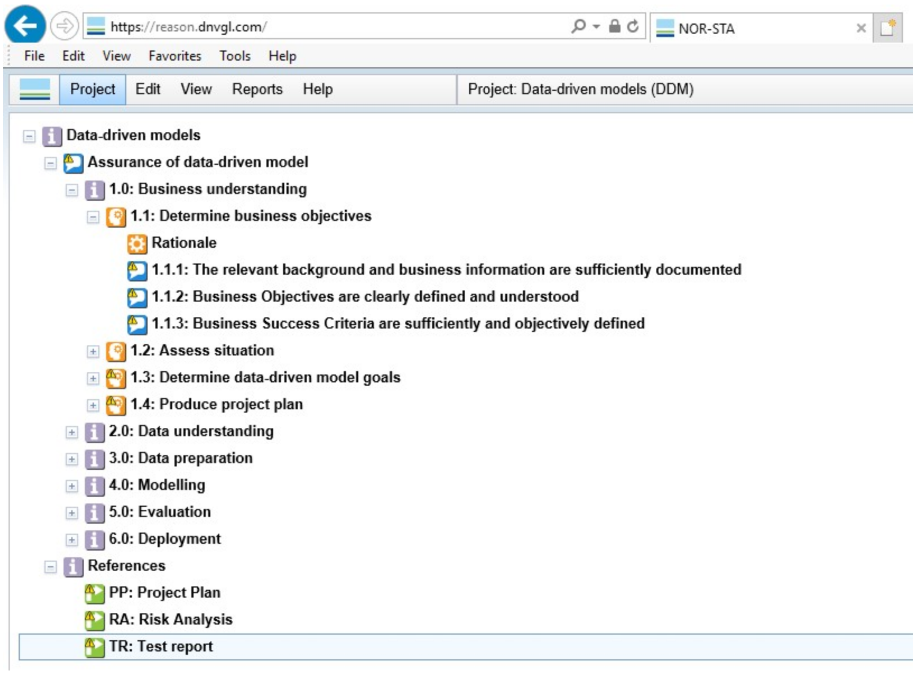


Figure 4 The RP implemented into the argumentation tool REASON

The stringency of the requirements on the evidence depends on the criticality and the uncertainty the application represents: higher consequence applications have a higher level of rigour, and thus more strict requirements on evidence. For example, an ML algorithm whose job is to suggest classifications of documents to a human user has a low criticality (there is no substantial risk to safety or asset integrity if a suggestion is wrong, since the human user can ignore the suggestion) but an ML algorithm which pre-sorts asset condition data into “OK”/“needs attention” where the human operator only looks at the “needs attention” cases has higher criticality and so has more extensive requirements on evidence. Similarly, if techniques are employed where there is little experience, or where the results are hard to check, the uncertainty increases, giving rise to a higher risk rating and thus more stringent evidence requirements. In a future version of the RP this might be elaborated into requirements tables like Table 1, but more research and development is needed for that.

4. Case Studies

Parallel with the development of the RP, a number of Use Cases have been run. The cases concern practical applications like mooring prediction, wear detection and remaining useful life estimation. The anonymized findings from the use cases are shown in Table 2. The findings demonstrate that Assurance of digital assets is needed and that the areas which are in focus in the upcoming Recommended Practice are relevant.

Table 2 Findings from the Use Cases run during the development of the Recommended Practice

Finding	Possible consequences	Use cases affected
No baseline model	• Model overengineered (a much simpler model would do)	A, B, D
No quantitative evaluation	• Unknown performance, impossible to use as basis of business decisions	A, E
Model only tested on simulation data not real data	• Jump to real world may include surprises	A, D
Model's computational feasibility is in question	• Cannot be used in practice	B
Not enough training cases	• Model is undertrained	B
Model does not generalize	• Model won't be able to deal with situations it hasn't seen before, i.e. it doesn't really *learn* anything	B
Intended use not properly understood	• Surprises at deployment time	C
Criticality not established	• Can't make business decision based on quantitative performance	C, D, B, A
Data available is out of date	• Even if project succeeds, not clear path to a future deployment	C
Quality of human-labelled data unknown	• Uncertainty in performance, and "how good is perfect?"	C
Evaluation metric is misleading	• "You fool yourself, and you are the easiest person to fool"	D
No held-out evaluation data	• Overfitting	E
Training data not representative	• Surprises at deployment time	E
Retraining plan is adhoc	• No systemetic quality	E
No continuous monitoring of performance	• Performance can degrade	E
Susceptibility to security attack	• Malicious attach, loss of data	E

5. Future Work

With respect to the future of this RP the following is foreseen

- Perform the formal review process for the RP (also including external reviews) prior to finalizing the RP and then publishing within the framework of rules.dnvgl.com
- Use results from the current research programme 'Assurance of Digital Assets' (ADA) to improve the RP with more recommendations on methods and techniques which should be used – in dependence on risk if relevant.
- So far, the assurance regime is process-centric, as motivated in Section 2. It does still make sense to add recommendations on how the model/algorithm itself should be validated. A Recommended Practice on Data Quality already exists (see /6/) and the two RPs are closely linked.
- CRISP-DM has been criticised for not including enough on the operational side, as data-driven models are not necessarily stable in this respect and need review and modification to stay useable. The RP should be improved in this respect.

References

[1] IEC 61508. Functional safety of electrical/electronic/programmable electronic safety-related Systems. IEC. Ed. 2. 2010

[2] CRISP-DM 1.0. Step-by-step data mining guide. SPSS. 199

[3] DNVGL-RP-0510. Recommended Practice for the Assurance of Data-driven Algorithms and Models. DNV GL. 2019. To be published.

[4] REASON. NOR-STA support system for achieving and assessing conformance to NORms and STAndards. See www.nor-sta.eu

[5] DNVGL-RP. Set of official Recommended Practice guidelines published by DNV GL. See https://rules.dnvgl.com

[6] DNV GL, DNVGL-RP-0497 Recommended practice: Data quality assessment framework. 2017. http://rules.dnvgl.com/docs/pdf/dnvgl/rp/2017-01/dnvgl-rp-0497.pdf.

Fleet-Based Remaining Useful Life Prediction of Safety-critical Electronic Devices

Mathias Verbeke, Alessandro Murgia, Tom Tourwé, Elena Tsiporkova
Data and AI Competence Lab, Sirris Bd. A. Reyerslaan 80, B-1030 Brussels BELGIUM

Abstract

Since ever more assets get connected to the Internet, an increasing amount of data is gathered on an asset's operating environment and its respective usage conditions. This enables one to closely monitor the behaviour of an asset during its lifetime and extract useful insights in a data-driven way. For complex assets operating in safety-critical environments, knowledge on their remaining useful lifetime is often invaluable, for example to proactively schedule maintenance interventions and avoid unwanted downtime. When a new asset type is introduced, however, limited to no historical data on other assets of this new type that were observed towards their end of life and can be used to learn a RUL prediction model is available. Consequently, traditional techniques to perform RUL prediction often fall short in this case. In this paper, a methodology that leverages the data of the entire fleet of assets to predict the RUL of new assets is presented, allowing to generalize across the behaviour of assets in different operating conditions. The methodology was validated on an industrial dataset of a fleet of electronic devices operating in safety-critical environments.

Keywords: fleet-based analytics, remaining useful life prediction, electronic devices, new asset types

1. Introduction

An important concept in reliability engineering is the asset's remaining useful lifetime (RUL), which is defined as the time the asset is likely to operate before it requires repair or replacement. Knowledge on the RUL allows avoiding unplanned downtime, proactively scheduling maintenance and optimizing the asset's operational efficiency. This is especially relevant for assets that are operating in safety-critical settings.

Due to the trend towards Internet-of-Things, an increasing amount of data is gathered on the asset's operating environment and the respective usage conditions. This has given rise to an extensive body of research in predicting the RUL of industrial assets in a data-driven way, using statistical or machine learning techniques. The approach to tackle this problem typically depends on the data that is available, for example using a threshold-based approach based on a set of inferred health indicators, or by extrapolating the gathered sensor values and comparing those to run-to-failure data from in-the-lab test setups gathered during product development.

Complex assets are characterized by a high amount of parameter settings and are used in a wide variety of operating conditions. This leads to a large number of degrees of freedom, due to which it becomes hard to determine fixed health indicator thresholds or generate run-to-failure data for each of these conditions. Furthermore, typically a range of different asset types exists. When a new asset type is introduced, limited to no historical data on other assets of this new type that were observed towards their end of life and can be used to learn a RUL prediction model is available. Consequently, techniques to perform RUL prediction often fall short in this case.

In order to meet this challenge, a methodology that leverages the data of the entire fleet of assets to predict the RUL is presented, allowing to generalize across the behaviour of assets in different operating conditions. The methodology was validated on an industrial dataset of a fleet of electronic devices operating in safety-critical environments.

2. Related work

Remaining useful life prediction is one of the activities that is enabled by monitoring product usage. Usage monitoring is a fundamental element in systems that assess the on-going health of a product or system, provide advance warning of failure, and provide information to improve the design and qualification of fielded and future products [1]. It is often denoted to Prognostics and Health Management (PHM) that links studies of failure mechanisms to system lifecycle management.

Usage monitoring has mainly been employed in the context of design and reliability tests, in which electronic devices such as desktops [2], notebooks [3], refrigerators [4], and game consoles [5] have been subjected to different types of loads (thermal, physical, electrical, etc.) to test their resilience. Most of these tests are performed by in-the-lab trials or small-scale field studies with actual users. More recently, Funk et al. [5] have proposed a methodology to build automatic observation modules into products, collect usage data, and analyze these data by means of process mining techniques in order to exploit in-the-field data by actual users and involve them in the development of such products earlier on. Next to (hardware) design and reliability, this also allows to more deeply study the interaction of the user with the product.

The main focus of this paper is on remaining useful lifetime prediction of (safety-critical) electronic devices, due to which the attention will be focused on related work in this subdomain. Okoh et al. [6] provide an overview of RUL techniques that are used in Through-life Engineering Services, i.e., services that aim to improve support services by providing run-to-failure information for better decision making. The authors present a classification of techniques used in RUL prediction for optimisation of products' future use based on current products in-service with regard to predictability, availability and reliability.

An increasing amount of research is also looking into hybrid approaches to predict the RUL. For example, Cheng and Pecht [7] propose a fusion prognostics method for RUL prediction of electronic products. In their approach, the PoF method is used to identify the critical parameters, identify and prioritize the potential failure mechanisms, identify the failure models, and define the failure criteria in terms of the isolated parameters, whereas the data-driven method is used to extract the features from the monitored parameters, create a healthy baseline, and compare the monitored parameters with the baseline to conduct anomaly detection and trend the isolated parameters. The approach is validated using a case study to predict the RUL of multilayer ceramic capacitors.

For a full overview of data-driven techniques for RUL prediction, we refer the interested reader to the recent book of Si, Zhang and Hu [8]. The book of Pecht and Kang [9] specifically focusses on prognostics and health management of electronics, including an overview of knowledge-based and data-driven techniques for remaining useful life prediction. To the best of our knowledge, however, no approaches exist to predict the remaining useful lifetime of new asset types on with limited information is available, based on the exploitation of the historic data of the fleet.

3. Case study: safety-critical electronic devices

The resulting methodology was validated on a real-world industrial case study. It involves a fleet of commercially available electronic devices that are used in safety-critical environments. From this product around 50 models are available, with a total installed base of around 100k devices. For each of the assets in the fleet, the power output as well as a number of device settings such as energy intensity are available over the lifetime of the device. A number of these device settings are set by the maintenance technician during fixed calibration interventions, as well as during finetuning by the end user. For the case study, we will focus solely on the power output of the asset which increases over time until the asset fails and reaches its end of life. This is illustrated in Figure 1, showing the power output evolution of a single functional unit until it reaches its end of life (marked in red).

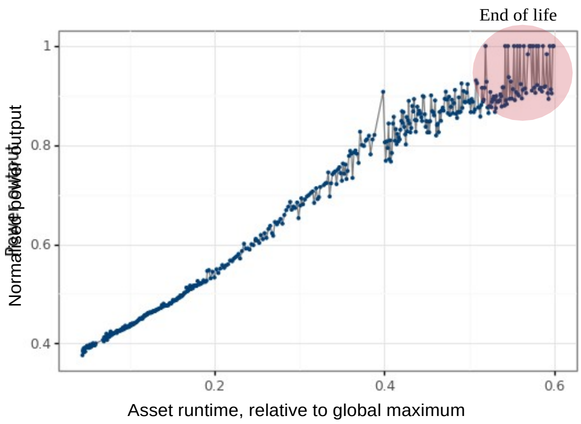


Figure 1: Power output evolution until the end of life of a single asset

4. Methodology

This section describes the methodology on a conceptual level. It consists of three main steps. First, assets operating in similar conditions are clustered together. Subsequently, for each of these groups, a so-called *reference curve* is calculated. This represents the prototypical behaviour of the assets in that particular group during their lifetime. To predict the RUL of a new asset, the top N most similar reference curves are identified and the average RUL of the assets in that cluster is determined. Finally, the RUL of the new asset is calculated as the weighted sum of these respective lifetimes. Each of these steps will now be discussed in more detail, taking the use case as illustrative example to demonstrate the approach.

Step 1: Clustering assets operating in similar conditions

Based on an extensive exploration of the data available on the entire fleet of assets, a clear difference in behaviour can be observed. Figure 2 depicts the energy intensity of the entire fleet of assets and clearly reveals a diverse behaviour across the fleet with several energy intensity modes.

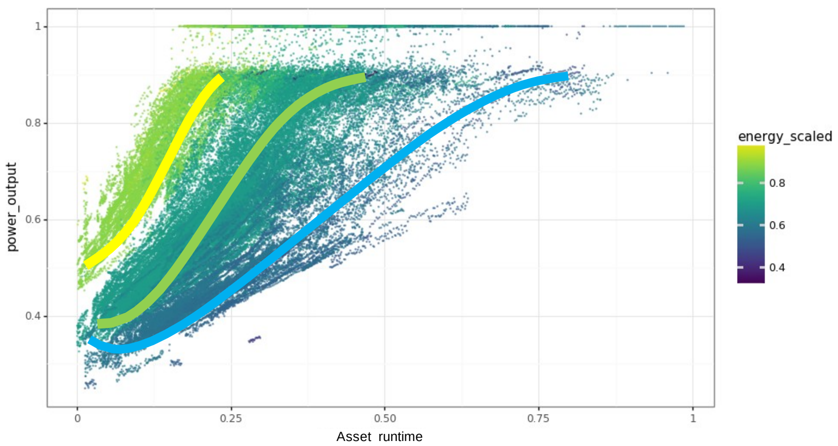


Figure 2: Energy intensity of the entire fleet of assets

To extract distinguishing behaviour across the fleet, the energy intensity curves of the entire fleet of assets are clustered using k-means clustering [1] 1 . This is an unsupervised machine learning algorithm which partitions n observations into k clusters. As the name indicates, each asset is assigned to the cluster with the nearest mean, serving as a prototype of the cluster.

The most important hyperparameter of the k-means algorithm is k, denoting the number of clusters to be formed. In the proposed approach, a silhouette cluster validation analysis is used in order to determine the optimal number of clusters. It studies the separation distance between the resulting clusters. The silhouette score ranges from −1 to +1, where a higher score indicates a better cluster separation.

For the case study, a silhouette analysis was performed in which cluster solutions with 4 up to 15 clusters were tested. The highest silhouette score was obtained using 5 clusters. The resulting clusters are depicted in Figure 3. Each of the clusters shows a different behaviour of the evolution of the energy intensity over the assets' lifetime. The slope of the curves is the main distinguishing characteristic, representing the speed at which the assets degrade over time.

[1] In the experimental validation, the implementation from tslearn [10] was used.

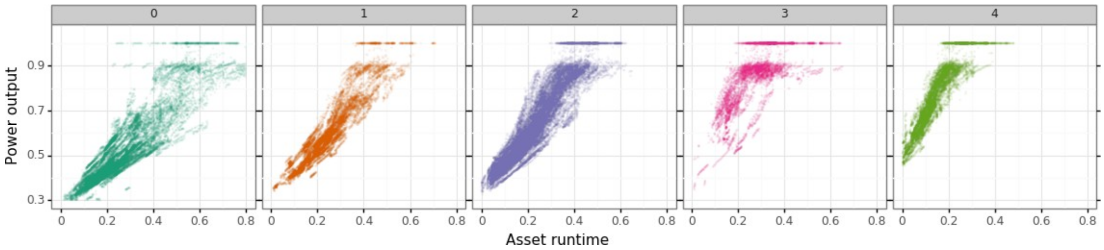


Figure 3: Clustering of the energy intensity curves with 5 clusters

Step 2: Extraction of a reference curve per cluster

In a second step, a reference curve per cluster is extracted. To calculate these reference curves, the median per timestamp across all energy intensity curves of the assets in the respective clusters are calculated.

The reference curves for the extracted clusters in the case study dataset (Figure 3) are shown in Figure 4. The region within one standard deviation (std) around the reference curves is also marked, showing that this limited boundary already captures quite some behavioural variation in the energy intensity, as these regions largely overlap with the point clouds from the clusters in Figure 3.

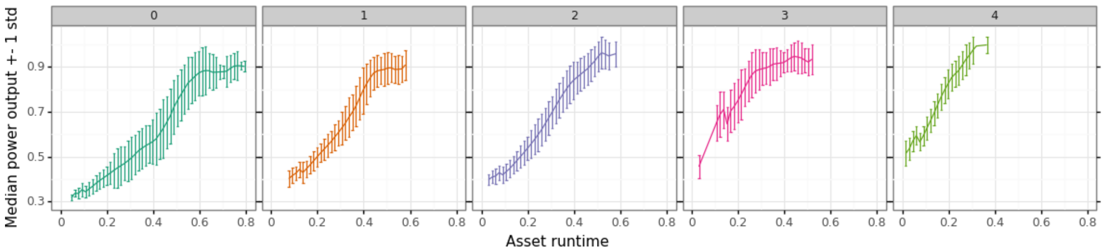


Figure 4: Median reference curves per energy intensity cluster

Step 3: Calculation of the RUL of the new asset

In the third and last step of the proposed approach, the remaining useful lifetime of an asset of a new type is calculated. This entails the following 4-step workflow, which will be illustrated by means of the prediction result on a prototypical asset as shown in Figure 5:

a. First, the available data on the new asset is compared to each of the reference curves. To this end, the coefficient of determination, denoted as R^2 is calculated. This coefficient measures the goodness-of-fit between the available new asset data and the corresponding part of each of the reference curves. More specifically, it provides a measure of how well the new asset data points are replicated by the respective reference curve, based on the proportion of total variation of new asset data points explained by the reference curve.

   For the example shown in Figure 5, the first part of the new asset data (thick black line at the start) is compared with the respective fragments of each of the reference curves. Note that the thin black line indicates the subsequent evolution of the power output until the asset reaches its end of life. This latter part is however treated as unseen data and is consequently not taken into account in the comparison with the reference curves.

b. Subsequently, the top *m* most similar reference curves are determined, which have an R^2 correlation metric above a fixed threshold. In the example, the top 3 reference curves with an R^2 score of above 50% are retained. These curves are shown in green on Figure 5, in which the color intensity indicates the degree of correlation (according to the color scale on the right).

c. In the third step, the end of lifetime of the retained top *m* reference curves is calculated.

d. Finally, the end of lifetime and the corresponding remaining useful life of the new asset is calculated as the weighted sum of the end of lifetime values of the retained top *m* reference curves as determined in the previous step. The weight factors are based on the normalized correlation strength. In the example plot, the ground truth end of lifetime is indicated with a vertical black dashed line, whereas the predicted end of lifetime is indicated with a vertical red dashed line.

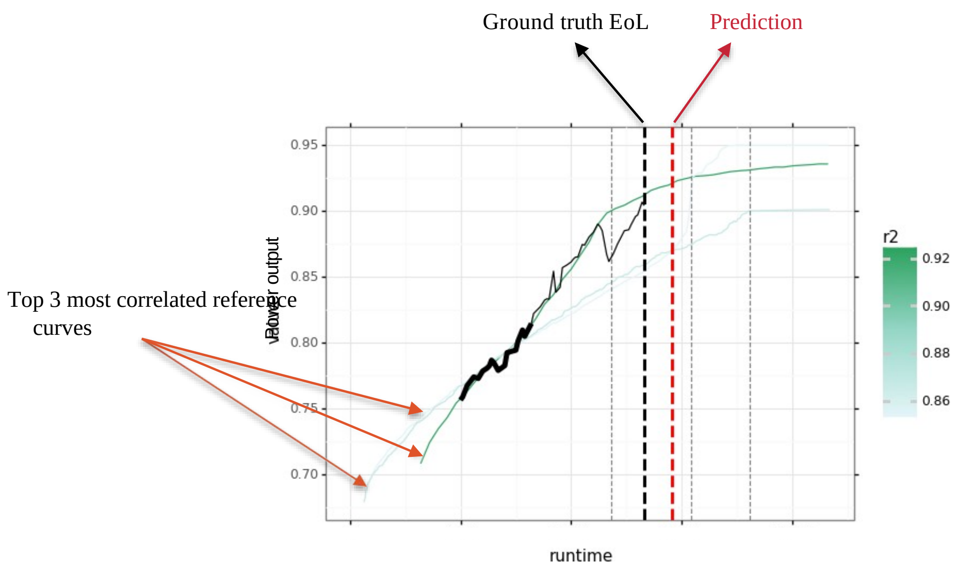


Figure 5: Prediction of the remaining useful lifetime of an asset of a new type

5. Validation

In order to validate the proposed approach, it was tested on two different subsets that differ in the amount of time within which the assets will reach their end of life. The first subset contains assets on which data is available for three quarters of their lifetime, and thus will reach their end of life in the remaining quarter. In the second subset, only data on half of the lifetime of the assets under investigation is available. The training set which was used to extract the reference curves consists of around 8.500 assets. Each of the test sets contains data of 50 new devices. For both the training and test sets, only assets of which the data was sufficiently qualitative (e.g., small percentage of missing data, no user interventions) were considered.

For each of the test sets, the end of lifetime and remaining useful life of each of the assets is calculated. For both datasets, an average accuracy of around 3-5% of the average lifetime over all assets in the fleet is obtained. The results for 4 example assets from the test set are depicted in Figure 6. While the current validation results look promising, a more extensive experimental validation will be performed in future work, considering a larger amount of assets from the full range of asset lifetimes.

Furthermore, also a comparison to an overall model that predicts the RUL without distinguishing between the different operating conditions of the assets in the fleet is planned.

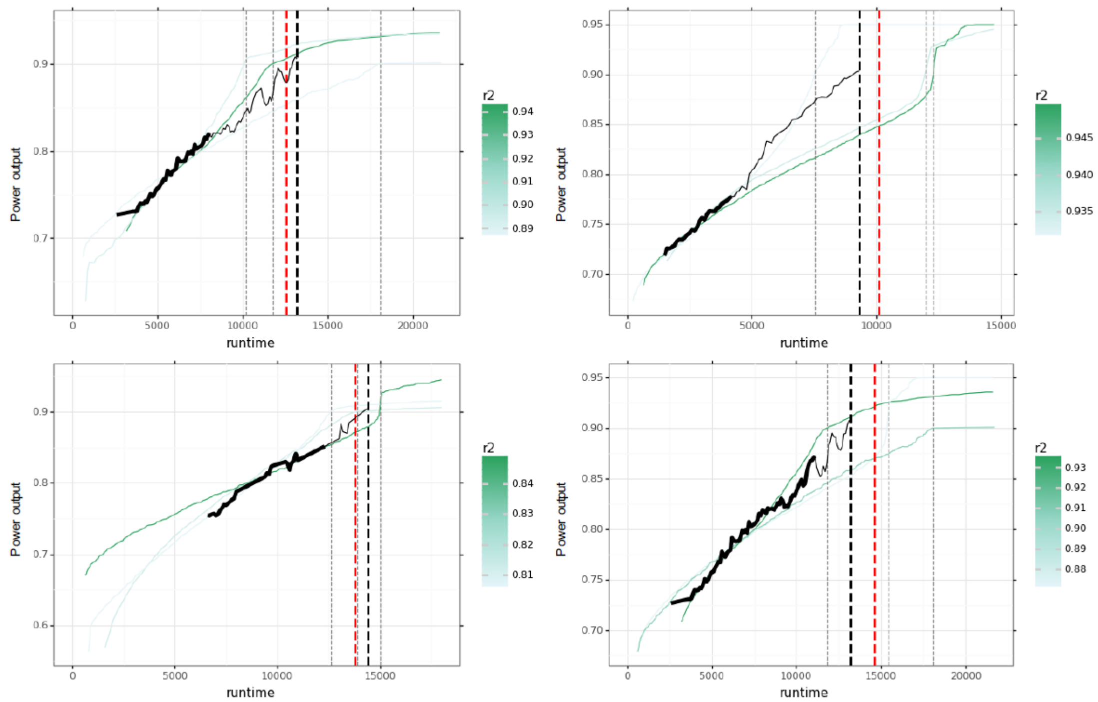


Figure 6: RUL prediction results on four example assets from the test set

6. Conclusions and Future Work

In this paper, a methodology was proposed to predict the remaining useful lifetime of new types of safety-critical assets on which little information is known at the start of the lifetime. This is done by leveraging the available historical data on the fleet of assets of former product types, which allows one to generalize across the behaviour of assets in different operating conditions. The preliminary validation experiments show that using the proposed approach, it becomes possible to predict the RUL with a reasonable accuracy.

In future work, it will be investigated if the accuracy can be further improved by exploring alternative (machine learning-based) RUL prediction models to complement or replace the current approach based on median reference curves per cluster. Also, the current methodology only takes the energy intensity into account to characterize the behaviour of assets in the fleet. In future work, the aim is to extend this with additional variables, e.g., to better take the variation in parameter settings into account throughout the lifetime of an asset. Finally, a more extensive error analysis will be performed to gain deeper insight into how to improve the weighted sum when calculating the remaining useful lifetime of new assets based on the end of lifetimes of the reference curves.

Acknowledgements

A substantial part of this work was performed by dr. Paul Teehan during his time at the Data and AI Competence Lab of Sirris. This work was supported by Flanders Innovation & Entrepreneurship (VLAIO) through the SBO project HYMOP (150033) and the Federal Public Service Economy through the project BitWind.

References

[1] N. Vichare, P. Rodgers, V. Eveloy, and M. Pecht, "Environment and Usage Monitoringof Electronic Products for Health Assessment and Product Design," *Qual. Technol. Quant. Manag.*, vol. 4, no. 2, pp. 235–250, Jan. 2007.

[2] D. Searls, T. Dishongh, and P. Dujari, "A Strategy for enabling datadriven product decisions through a comprehensive understanding ofthe usage environment," in *Proceedings of the Pacific RIM/ASME International Intersociety Electronic and Photonic Packaging Conference (IPACK'01),* Kauai, Hawaii, 2001, pp. 1279–1284.

[3] N. Vichare, P. Rodgers, V. Eveloy, and M. G. Pecht, "In Situ Temperature Measurement of a Notebook Computer—A Case Study in Health and Usage Monitoring of Electronics," *IEEE Trans. Device Mater. Reliab.*, vol. 4, no. 4, pp. 658–663, Dec. 2004.

[4] X. Yang, P. R. Moore, C.-B. Wong, J.-S. Pu, and S. K. Chong, "Product lifecycle information acquisition and management for consumer products," *Ind. Manag. Data Syst.*, vol. 107, no. 7, pp. 936–953, Aug. 2007.

[5] M. Funk, A. Rozinat, A. K. Alves de Medeiros, P. van der Putten, H. Corporaal, and W. van der Aalst, "Improving Product Usage Monitoring and Analysis with Semantic Concepts," in *Information Systems: Modeling, Development, and Integration*, vol. 20, J. Yang, A. Ginige, H. C. Mayr, and R.-D. Kutsche, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 190–201.

[6] C. Okoh, R. Roy, J. Mehnen, and L. Redding, "Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services," *Procedia CIRP*, vol. 16, pp. 158–163, 2014.

[7] S. Cheng and M. Pecht, "A fusion prognostics method for remaining useful life prediction of electronic products," in *2009 IEEE International Conference on Automation Science and Engineering*, Bangalore, India, 2009, pp. 102–107.

[8] X.-S. Si, Z.-X. Zhang, C.-H. Hu, *Data-Driven Remaining Useful Life Prognosis Techniques: stochastic models, methods and applications.* Springer, 2019.

[9] M. G. Pecht and M. Kang, Eds., *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet ofThings*. Chichester, UK: John Wiley and Sons Ltd, 2018.

[10] R. Tavenard, Faouzi, Johann, and Vandewiele, Gilles, "tslearn: A machine learning toolkit dedicated to time-series data." 2017.

Assessing GB Train Accident Risk Using Red Aspect Approaches to Signal Data

Chris Harrison[1], Xiaocheng Ge[2], Julian Stow [2]
[1]RSSB, United Kingdom
[2]University of Huddersfield, United Kingdom

Abstract

On the Great Britain (GB) railway, an event involving a signal passed at danger (SPAD) is one of the predominate hazards. It is important to have a better understanding of the circumstance under which this can happen and the red aspect approaches to signals (RAATS) project has been undertaken to do this. The result of this work is the development of a tool that can estimate the number of times a signal is approached at red by trains on the GB rail network by using a big data approach. This piece of extra information is crucial for understanding the likelihood of a SPAD at individual signals and also for normalisation of SPAD data at different levels, locally and nationally, for trending and benchmarking. The tool has been developed by mining different sources of data from Network Rail, recording and analysing millions of pieces of data from live operational feeds to update and summarise statistics from thousands of signal locations in GB on a daily basis. The tool has amassed several years of data and is now being used to better understand how SPAD risk arises on the GB network. There are also many other potential applications of the tool in understanding rail punctuality performance, timetable design, human reliability along with the refinement of other existing safety analysis tools.

Keywords: big data, human error, rail, red aspect approaches to signals (RAATS), risk, signal passed at danger (SPAD), train accident

1. Introduction

Monitoring is an essential part of any management system. In a railway safety context, such management systems are designed to keep all safety risks under control. In the EU the Common Safety Management (CSM) for Monitoring [1] requires the operators to monitor all the processes of the Management System and the Management System as a whole, which typically involves:

- Strategies and plans for monitoring
- Systems to collect data
- Processes to analyse data, turning it into information
- Use of the information to improve the processes and the management system

In Great Britain (GB), RSSB is one of the key players in terms of rail industry data collection and analytics. The primary aim of RSSB is to enable a better, safer railway and one of the ways we achieve this is by providing the GB rail industry information about safety performance and risk. Good health and safety management relies on good data and analysis. RSSB hold the central position within the industry where the data that underpins good health and safety management is defined, recorded, analysed and reported on. To do this effectively we are constantly evolving and developing new techniques and analysis tools to aid in monitoring and modelling railway safety. The Red Aspect Approaches to Signal (RAATS) tool is one of RSSB's (in collaboration with the University of Huddersfield) latest developments.

This paper starts by giving some background and context for the RAATS tool, briefly outlining the demand for it and why it is important. It then goes on to describe the development of it to date and presents some examples of the outputs and how they can be used in a risk and safety analysis context in the GB rail industry. The benefits of this approach are then evaluated explaining how doing this provides a more complete picture of the risk along with additional insights above and beyond just analysing incidents that have occurred. Finally the paper will outline the potential further developments and applications of the tool (and how it might interface with other data/tools) and how we see it and the underlying big data approach making further substantial improvements to the understanding, analysis and assessment of train accident risk on the GB railway.

2. Background

An event where a train passes a signal showing a stop aspect without authorisation is known as a 'signal passed at danger' (SPAD). SPADs can range from minor incidents where a signal is passed by only a few metres to serious incidents where longer overruns give rise to the chance of collision with other trains. The causes of SPADs can vary widely from driver error to degraded braking performance as a result of low adhesion [2]. Driver error is frequently cited as a primary cause, often described in terms of the failure to take sufficient action at preceding warning signals ('misread') or failure to control the train on the approach to the red signal ('misjudgement') [3]. However, it is recognised that there are many underlying technical, organisational and human factors related causes which can contribute to the eventual failure of a driver to stop at a red signal [4, 5]. An example of this is the accident at Ladbroke Grove, UK, in 1999 in which there were 31 fatalities. The accident report [6] identified key failings in the design of the signalling system, signal sighting and driver training as causes of the accident.

For a SPAD to occur, a train must approach the signal at red in the first place. It follows that knowing the number of trains that approach signals displaying a red aspect (the 'red approach rate') is fundamental to the understanding of SPAD risk at individual signals and the normalisation of SPAD data, both locally and nationally, for trending and benchmarking. SPAD risk has been rigorously studied previously using several techniques including fault and event tree analysis (e.g. the RSSB Safety Risk Model [7]) and Bayesian Belief Networks (e.g. by Marsh and Bearfield [8]). These techniques require a knowledge of the red approach rate to provide an accurate quantification of the resulting SPAD risk, but historically they haven’t been able to utilise such information due to the difficulties in obtaining it.

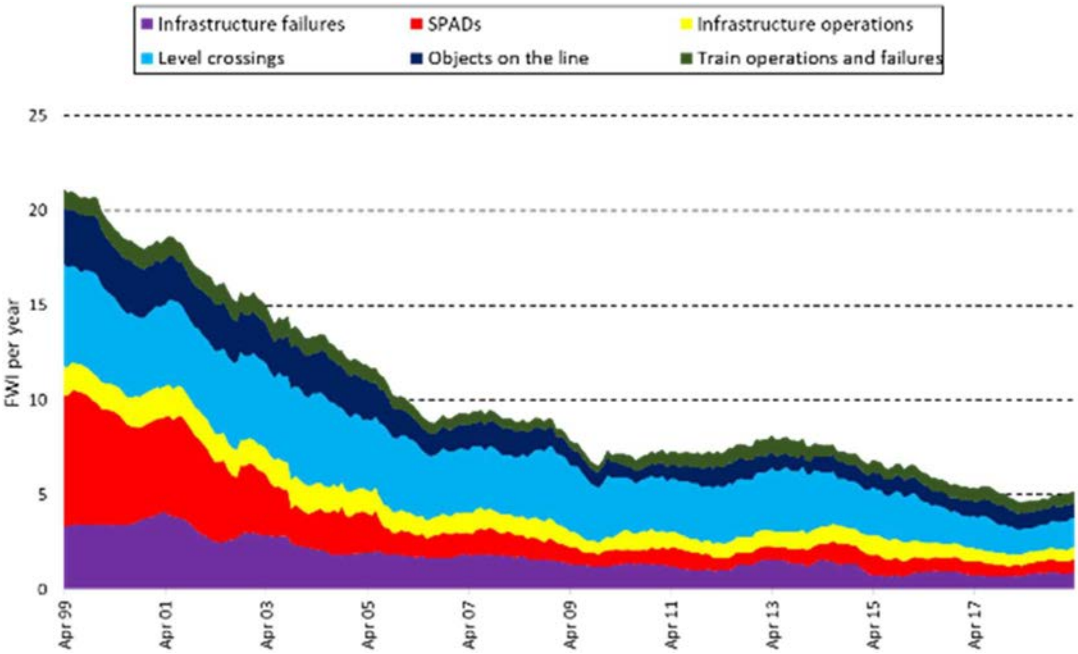


Figure 1: Risk as measured by the RSSB Precursor Indicator Model (PIM)

The chart in Figure 1 shows the current and historic risk from SPADs based on the RSSB Precursor Indicator Model (PIM) [9]. Since 1999 it can be seen that the GB rail industry has reduced SPAD risk from around 7 FWI[1]/year to the current (2019) value of less than 1 FWI/year. There are many good reasons for this, both from a technological and an operational management viewpoint. New safety intervention systems and engineering controls have been implemented on the network, along with improvements in the operating procedures and principles and better management of the workforce. However, there is still more that can be done and the GB rail industry continues to work to improve safety and reduce risk further. From the chart in Figure 1 it is also apparent that the risk level has plateaued, and the question is now what else can we do, what other tools and techniques can we develop to reduce the residual risk further? The RAATS tool has been developed to help answer this question.

[1] Fatalities and weighted injuries (FWI) is a measure of harm that allows different injuries to be aggregated together into a single composite metric. This is done by applying different weights to the different injury outcomes being considered. Further details can be found in reference [9].

3. Development of the RAATS tool

Network Rail (the GB mainline railway infrastructure manager) provides publicly available live data feeds which give various information on the movement of trains [10]. At the most fundamental level, the source of the information used by the RAATS tool is train describer (TD) data. A TD is an electronic device connected to each signalling panel which provides a description of each train (its 'headcode') and which section of track (or 'berth') it currently occupies. The TD is responsible for correctly displaying the train movements from berth-to-berth to the signaller and for ensuring that the train's identity is correctly passed to the next signaller's panel when it leaves the current signalling area.

RAATS uses two separate TD data feeds, termed C-class and S-class messages. C-class messages record train movements between individual track berths, whilst S-class messages record the times at which signal aspects change. Note that the S-class data only shows whether a signal is 'off', showing a red aspect or 'on', showing a proceed aspect (single or double yellow, or green). Both C-class and S-class messages are transmitted through the live feed with a combined total of approximately million messages being sent per day. As such, it has many of the characteristics normally associated with big data such as high volume, high velocity and significant value as specified by Attoh-Okine in [11].

Big data is the new frontier for collecting and analysing data and for transforming it into usable information. Although it is difficult to have a clear definition of big data applications, in the literature there are several tentative ways to define this new technology, and all of them rely on the capability of the big data implementations to handle, at high speed, a big volume of data, coming from various sources. Those three elements are summarised using the three V's approach, namely: Volume, Variety and Velocity:

- Volume is the size of the data sets: the magnitude order can be from Terabytes to Petabytes.
- Variety means that big data is capable of dealing with data coming from different sources and having different or no structure.
- Velocity can be understood as the capability to quickly handle input data (the speed at which data arrives) or to provide real-time information as the output (the speed at which some meaning can be extracted and disseminated).

The first iteration of the RAATS analysis model used the time at which a train enters and leaves each berth and the times at which the corresponding signals change state to classify every train approach to every signal into different categories. The model is designed as a decision tree and further details can be found in reference [12].

In the second iteration the methodology was refined by developing a speed profile algorithm to identify if a train stopped at a red aspect based on the timings of trains entering and leaving each signalling berth, the timings of signal changes, berth lengths and berth-platform association. For train approaches at berths with known lengths and calculable entrance speed values, the red aspect approaches can be measured using this speed profile method. For train approaches where the berth length was unknown an alternative approach was developed using a classification tree based on the methodology developed in the first iteration mentioned previously. Further details of the approach can be found in reference [13].

The tool is currently undergoing refinement and preparation for the third iteration. This will result in a tool that uses real time data, and which can process and make available the statistics of red aspect approaches for a day within 24 hours. The algorithms have undergone further refinement and optimisation to run in such an environment.

4. Overview of the RAATS tool

The RAATS tool receives real, live data of the movements of every single train in the GB rail network along with information on the status of signals where possible[2]. There are thousands of records an hour that the tool has to capture, process, analyse and then collate with historic information in a format that can be interrogated.

There are two layers in the kernel of the RAATS Tool: a modelling layer that models the scenario of a train approach towards a signal; and a classification layer that classifies the approaches into different classes (see Fig. 2).

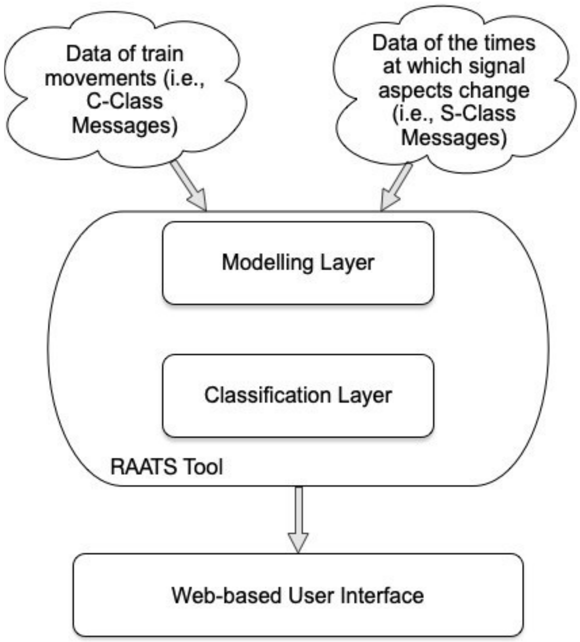


Figure 2: Architecture of the RAATS tool

[2] Not all signals are currently capable of providing this information.

At the model layer, millions of records are processed and those which are related are selected and used to describe the circumstances of a train travelling through three sequential berths. The results from the first layer provide a foundation for the process in the classification layer. In the classification layer, each case of a train approaching a signal is classified into one of several classes based on the timing information of the train's movement and the status of the corresponding signal.

By processing live data streams from Network Rail, we are able to know each train's movement at the berth level in terms of entry and exit timings. Figure 3 is an illustration of a typical scenario. A train travels into a berth of interest (Berth1 at LocA) at the time T_a. It is a red approach if the signal (Sig2) at the end of Berth1 is displaying a red aspect at time T_a. At time T_{S1}, Signal Sig2 clears and the train moves towards the signal and leaves Berth 1 at LocB at time T_b. A short while later signal Sig2 changes back to red at time T_{S0}.

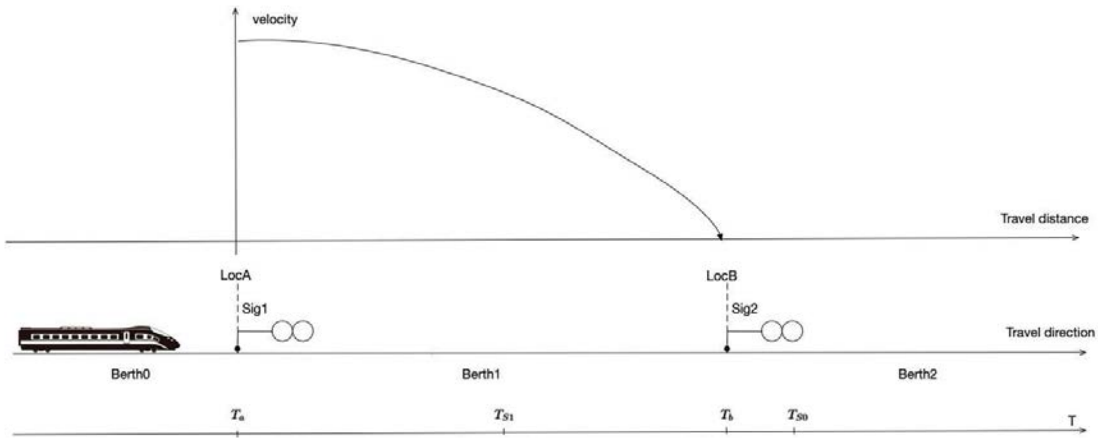


Figure 3: A typical model of a train moving towards a signal with a red aspect

Based on the knowledge of typical approaches to red signals on the GB rail network and the timings T_a, T_b, T_{S1} and T_{S0}, the model can estimate a train's speed when entering the berth (Berth 1) and classify the train's approach to a red signal (Sig2) into one of the following scenarios:

- NRA (Non-Red Approach): When the train enters the berth, the signal aspect at the end of the berth is showing a proceed (i.e., non-red) aspect.
- CSS (Cleared after Stopping at Signal): Signal clears after the train has come to a stand at the signal and the train departs immediately.
- CBD (Cleared Before Departure): Train enters the platform and stops at the red signal. This is a subset of the CSS category and is where a CSS approach occurs at a platform.
- CAS (Cleared on Approach to Signal): The signal is red when the train enters the berth but it clears (i.e. changes to a non-red aspect) before the train comes to a stand.

- PSS (Possibly Stopped at Signal): The signal possibly clears after the train has come to a stand and the train departs immediately. This class is the 'grey area' between the CSS and CAS classes and covers those train approaches which cannot be confidently classified as CSS or CAS using the currently available data.

The tool currently exists as web-based application. Figure 4 shows an example of the current interface and some of the results that can be displayed on a signal basis.

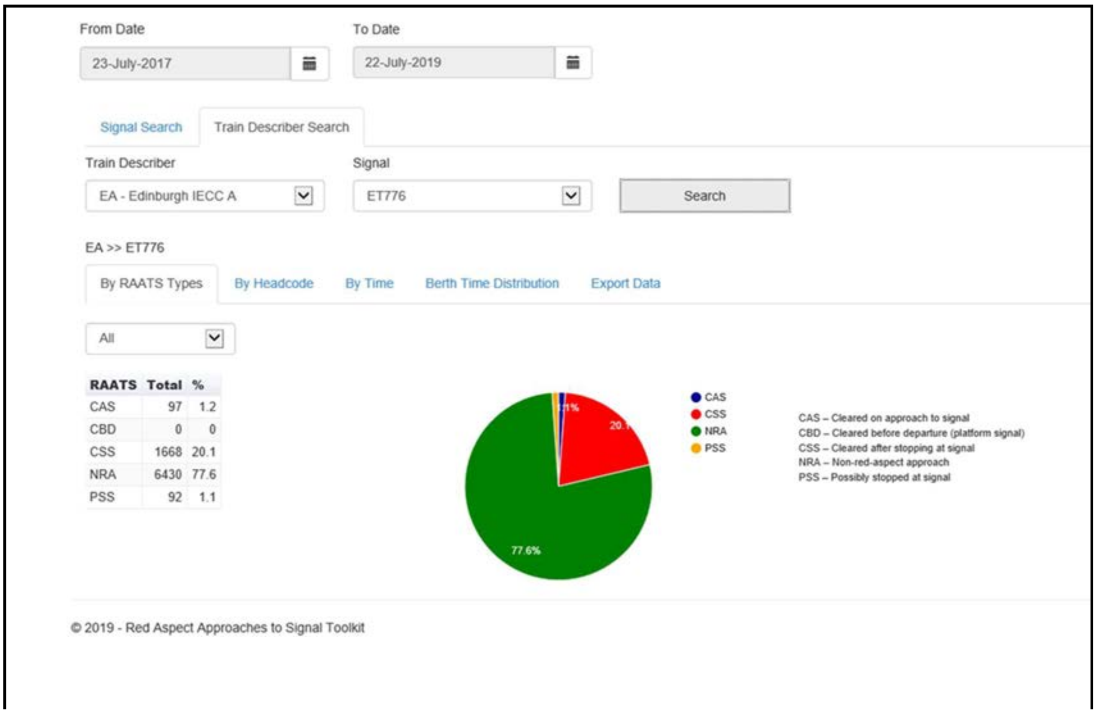


Figure 4: RAATS tool interface

5. Current applications of RAATS

A. Using RAATS data to understand risk

The outputs from RAATS are vitally important in risk and safety analysis. They allow a better indication of the chance for a signal to be passed at red than any other normaliser (e.g., train miles, total number of trains passing a signal) that has been used to date. In addition, there are many risk models and tools currently in existence, not to mention the many assessments that have been undertaken that have utilised, or attempted to incorporate within them, parameters related to the red aspect approach to a signal. For such a key parameter, it is important to note that until recently only crude estimates of the rates or proxies for them have been used as understandably no tool (until RAATS) was available to obtain them. This situation has mainly driven the development of the tool, but note that knowledge of red aspects is not limited to just risk and safety analysis, they also play an integral part in understanding train performance (punctuality and delay). This section of the paper will look at some of these applications in more detail and illustrate how the RAATS data is envisaged to be used and developed in the future.

B. Risk Modelling

Network Rail assesses the SPAD risk associated with every signal on the network using a process which examines the frequency and the potential consequences of passing that signal at danger. A tool has been developed, the Signal Overrun Risk Assessment Tool (SORAT), to undertake signal risk assessments. The risk assessment considers factors such as distance to a conflict point (such as a junction), train speed and passenger loading. This process does not currently incorporate any estimates for the number of red aspect approaches to a signal within the risk assessment; however, it aspires to do so if such information was readily available. Enabling this will improve the granularity of the modelling that the tool undertakes and allow better localisation and estimates of the relative level of SPAD risk at signals to be made.

Another application that will benefit from RAATS data is the RSSB Safety Risk Model (SRM) [7]. The SRM is a mathematical model of the risk arising from the operation and maintenance of the GB railway. It provides the GB rail industry with national estimates of risk through a series of models, some of which consider SPADs and the consequences that can ensue from them. Research is currently underway by RSSB (R&D project T1136 [14]) to look at redeveloping the model and investigating new techniques and ideas that can help improve and refine the risk estimates the model makes. One of the ideas that has been looked at is developing a SPAD Bayesian Network model that will enable a framework of causes, factors, attributes and conditions to be modelled. A demo model has been produced and RAATS data has been used as one of the inputs.

C. Human reliability quantification

A previous study [3] made an observation that signals with a high proportion of red aspect approaches actually have a relatively low chance of a SPAD occurring for each red aspect approach because drivers are accustomed to approaching the signal at red. Conversely, signals with a low proportion of red aspect approaches have a relatively high chance of a SPAD occurring when the signal is approached at red because drivers are not expecting a red signal.

There is also the question of SPAD potential based on the frequency of red aspect approaches. Are signals with more red aspect approaches proportionately more likely to have a SPAD event? It is likely to be a combination of the two factors, the frequency of red aspect approaches and the proportion of approaches to the signal that are red aspect approaches that influence the chance of a SPAD at the signal.

Initial analysis of the RAATS data has also highlighted that the type of train (e.g., express, stopping, empty coaching stock, freight etc) is also a factor that needs to be considered when calculating SPAD probability. The work raises the question of what is meant by SPAD probability, as it really depends on the context in which this is answered. On the one hand we can look at SPAD probabilities associated with different types of train approaches to a single signal, which could be further broken down by operator. Alternatively, looking at it another way, SPAD probability could be calculated from an operator's perspective along a route, taking into consideration all of the train approaches to signals on a journey across all of the operator's services.

6. Further applications of RAATS

There are a number of other potential uses of the underlying train TD data that RAATS is based on. The analysis has the potential to assist with understanding performance and capacity constraints on the network. This could be achieved by comparing the theoretical timetable against what actually occurs and assessing if there are areas where it could be optimised or designed better. TD data could also be used to identify changes, locally and nationally, in red aspect demands as a result of timetable changes or as a result of incidents causing disruption. This is particularly important when large scale changes are being made and an example of what happens when it goes wrong was widely reported in the British press over 2018 and led to an inquiry by the Office of Rail and Road [15].

Another use could be in understanding the routing of trains, the positions of points at junctions along with the frequency of their use. This is particularly useful in the context of a signal risk assessment where there are multiple conflict paths that a train could take in the event of passing it at danger. The RAATS data provides a way of calculating the probability that each route could be taken. It also provides a way of counting trains as they pass through the rail network and these statistics have a wide use in understanding usage and normalisation when making comparisons.

Further applications of the data might involve further algorithm development, particularly with regards the braking and acceleration profile of a train as it passes through the berth. The model is relatively simple at the moment, however it could be developed further, possibly by linking to other data on the train or external to understand where and for how long traction is taken or braking is made and then infer from this other metrics like track degradation, energy usage or emissions.

7. Future Development

D. Using RAATS in further big data analyses

RAATS data in itself provides useful information to understand the number of times a signal is approached at red. The main benefit of this data though will be in combination with other data sets, some of which are potentially very large (e.g. train recorder information, weather conditions, delay statistics etc). To make the sort of gains in safety improvement outlined at the start of the paper is likely to require ever more sophisticated and granular modelling techniques. The reason for this lies in the fact that economically viable and justifiable safety improvements tend to need to be identified at quite specific locations (e.g. a particular signal) rather than as a national policy (e.g. all signals). There are of course exceptions, and large-scale fitment of say a new train protection system or equivalent on the GB railway could potentially make a substantial difference, but it would be at cost and perhaps justified in terms of both safety improvement and other wider economic benefits. In terms of identifying improvements solely on a safety basis (and for which there is a legal requirement to do so), there is a need and an emphasis now to model at very specific locations, which in turn increases the amount of data that needs to be collected and the demands on the modelling framework to make sense of it. Utilising the principles of big data analysis and creating a model that has the capability of combining diverse and large data sets is at the forefront of the GB rail industry's efforts to better understand SPAD risk.

In [16] a first attempt was made at such an approach, bringing together SPAD and RAATS data with signal asset data to assess the characteristics of the GB rail network potentially give rise to SPADs. This method offers some huge opportunity, as it facilitates an analysis framework to be set up where information at signals that have experienced a SPAD can be used to infer information about signals that have yet to experience a SPAD. On the GB rail network there are approximately 30,000 signals and each year around 300 SPADs occur [9]. A lot of these are at signals that have never seen a SPAD before and in some cases upon investigation some additional controls and risk mitigation can be identified to help prevent further occurrences. In the future the aim is to be able to do this pre-incident and use a big data approach to help prevent incidents before they occur by identifying classes of signal based on their characteristics that warrant some sort of intervention.

Some progress has been made in attempting to do exactly that. In the Tavison project
[17] the aim was to try and combine several related data sources (one of which was RAATS) to better understand SPAD risk. Other sources of data included the on-train data recorder (OTDR) and close call incident data within a graph database to assess where the network was vulnerable to SPADs manifesting. The work showed that it was indeed possible and developed a working framework, however it also highlighted that there are several obstacles to

overcome in doing this longer term. Notably that such frameworks are very data intensive, particularly with obtaining and processing the OTDR data in real time to be able to undertake the necessary analysis.

E. Algorithm and tool development

RAATS is an ongoing project. As discussed in previous sections, there are several potential usages of RAATS data and some of these require further algorithm development. One of the most promising ideas is being researched under a project called Red Approaches to Signals by Train Journeys (RABYTS). Analysing the situations of red approaches along a train journey will give us a richer picture and information to analyse the circumstances and factors of a SPAD event. It will also help the industry to evaluate and monitor the performance and safety management of individual train operating company (TOC) by establishing a more accurate national benchmark of red approach rate. By linking the red approach rate data with train service data there is also the potential to help the improvement of rail service performance. The RAATS data by train services can be visualised as an actual "running" map which can then be compared with the service timetable. Achieving this will likely be done through further application of big data principles, in particular the linking together of large datasets and the creation of a framework to extract useful safety intelligence.

One of the key limitations at the moment of the RAATS data is the coverage. This is due to the fact that not every signal on the GB network is currently capable of providing the information necessary for the RAATS algorithm. There is a research project underway at the University of Huddersfield to look at overcoming this drawback by developing a statistical algorithm that can be applied to those locations where the signal status is not provided. This will be based on the train movement data alone and if successful will enable RAATS information to be available for the whole of the GB rail network.

8. Summary & Conclusion

This paper has provided an overview of the RAATS tool and its development to date. The outputs and applications of the tool have been presented along with some ideas for future development. The purpose of the tool is to enable better understanding of the factors and influences that lead to SPADs which can be used to help identify vulnerable signals and where the red aspect approach profile is changing that may indicate where action needs to be taken. There are potentially many other applications and big data approaches that cab be explored.

So why would doing all of this be different to what the GB rail industry currently does in this area and what benefit is it likely to bring? The main difference and perhaps where the most benefit is likely to be gained is in shifting the industry focus from a less reactive to a more predictive and proactive risk management regime. What this means is trying to use the SPAD data and the RAATS data (in combination with other data sources through big data principles) to try and identify high risk signals and intervene before SPADs occur at them rather than reactively address issues emerging as SPADs occur.

The RAATS project and the tool are the first attempt to apply a big data approach to improve the safety analysis and monitoring of the GB rail operation. The lessons and experience of the RAATS project, and in more general terms the experience of safety monitoring by big data approaches, can be shared among other nations in Europe.

References

[1] Common Safety Method for Monitoring, European Commission Regulation 1078/2012

[2] RIAC Human Factors Working Group: 'Signals Passed at Danger (SPADs) and human factors', Office of Rail and Road, 2014

[3] Nikandros, G., Tombs, D.: 'Measuring railway signals passed at danger'. 12th Australian Conf. on Safety Critical Systems and Software, Adelaide. Conf. in Research and Practice in Information Technology (CRPIT), vol. 86, Tony Cant, Ed

[4] Pasquini, A., Rizzo, A., Save, L.: 'A methodology for the analysis of SPAD', Safety Science, 2004, 42, (5), pp. 437–455

[5] RSSB: 'Common factors in SPADs (T070)' (RSSB, 2004)

[6] HSE Books: 'Ladbroke Grove Rail Enquiry, Part 1 – Report'. Her Majesties Stationary Office, 2000, p. iii

[7] RSSB Safety Risk Model: https://www.rssb.co.uk/Standards-and-Safety/Improving-Safety-Health--Wellbeing/Analysing-and-modelling-risk accessed September 18 2019

[8] Marsh, W., Bearfield, G.: 'Using Bayesian networks to model accident causation in the UK railway industry' Probabilistic Safety Assessment and Management (Springer, London, 2004), pp. 3597–3602

[9] RSSB Annual Health and Safety Report 2018/19: https://www.rssb.co.uk/-

[10] /media/Project/RSSB/Platform/Documents/Registered/Registered-content/Improving-Safety-and-Health/annual-health-and-safety-report-2018-19.pdf, accessed 18 September 2019

[11] Network Rail Operational Data Feeds: www.networkrail.co.uk/data-

feeds/, accessed 27 February 2015

[12] Attoh-Okine, N.: 'Big data challenges in railway engineering'. Proc. 2014 IEEE Int. Conf. on Big Data, Washington DC, USA, October 2014, pp. 7–9

[13] Zhao, Y., Stow, J., Harrison, C.: 'Estimating the frequency of trains approaching red signals: a case study for improving the understanding of SPAD risk', IET Intelligent Transport Systems, Vol. 10, Iss. 9, 2016, pp. 579-586

[14] Zhao, Y., Stow, J., Harrison, C.: 'A method for classifying red signal approaches using train operational data', Safety Science, Vol. 110, Part B, 2018, Pages 67-74,

[15] RSSB R&D project T1136, 'Development of a new Safety Risk Model (SRM)': https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=25124, accessed 18 September 2019

[16] 'Inquiry into May 2018 timetable Disruption' Office of Rail and Road, December 2018. Available at ORR https://orr.gov.uk/data/assets/pdf_file/~~0010/39916/inquiry-~~ ~~into-may-2018-timetable-~~disruption-december-2018-report.pdf, accessed 12

[17] September 2019

[18] Zhao, Y., 'Use of Red Aspect Approach Rates to Improve Understanding of SPADS', Institute of Rail Research, University of Huddersfield, IRR 110/137, 2016

[19] van Gulijk, C., Figueres, M., Hughes, P., El-Rashidy, R., 'Tavison – The Augmented SPAD BowTie', Institute of Rail Research, University of Huddersfield, IRR 110/174, 2018

Quality in big data for the forced unavailability of power plants

Henk Wels
Retired engineer
hc-wels@outlook.com

Abstract

Big data are currently considered in many places as solutions to many problems. Indeed, big data offers many exciting opportunities for analysis but it may miss sufficient quality. Data quality is an essential aspect in big data analysis. For power plants e.g. unavailability data, a low quality may lead to either over- or underinvesting in the grid reserve margin as well as to incorrect judgement of a plant when compared to its peers in benchmarking. A too large reserve margin is a waste of money, too little increases the risk on blackouts. Incorrect judging in benchmarking is detrimental to plant personnel motivation, which motivation is necessary for improvement.

In the paper quality of forced unavailability data is discussed from a historical perspective starting with Dutch data originating from a German data system, the successor of this German system being the international VGB KISSY database and the so called Transparency data from EMFIP, EEX and other platforms. Wind data are touched upon.

All data mentioned can be used to measure and analyze forced unavailability of power plants using typical non-expert analysis software such as Excel with varying effort doing so. However the amount of detail in the data seems to be inversely proportional to the amount of data. Too much detail without using it has its dangers: too little detail bears the risk of big data being a total black box without any technical explanation on variations in forced unavailability.

Keywords: Reliability Power Data Quality

1. Introduction

In the paper data quality aspects are reported in a Dutch system no longer present to gather reliability data to improve power plants, in an existing German system (KISSY) to gather such data for various reasons and applications and in the new so-called ENTSO-E Transparency data which are meant to bring power to places under commercial trading boundary conditions. All data sets can be considered big data when they were started. As the author recently retired, it is good to look back on items already realized using the data as well as looking forward to new and exciting analyses using high quality data.

In the Dutch centrally gathered data system, we did not think much on data quality but it came almost automatically as power plants were visited to discuss the centrally gathered failures. A yearly feedback took care of definition questions and showed what the data could be and were used for. Regrettably this system was stopped due to liberalization and corresponding reduction of personnel costs.

In the VGB KISSY database data quality is the result of strict definitions and user involvement. The number of data is typically higher compared to the Dutch system. Nevertheless, there are differences in data quality between power plants of different operators. This is logical as the people that fill in the data are not flawless machines. Feedback to people filling in the data is organized ad-hoc when it is felt necessary and there is continuous feedback from working groups. The KISSY data are available on-line to data contributors and have been used for example to derive subsystem RAM data from. They further have been used in life extension studies and are now being used for analysis of cycling of power plants.

The Transparency data also contains unavailability of power plants, however its use for operational purposes appears to be limited up to now. This is modern big data. However, it is easy to show that data quality can be improved when some transparency data with data from Dutch plants are compared. Nevertheless, this cannot be generalized as due to the legal requirements the transparency data should be flawless and data quality is dependent on the personnel filling in the data. The limits for Transparency data use as well as some new options for usage are discussed. Wind data are touched upon.

2. Visiting power plants to improve planned and unplanned unavailability

Dr. J. van Liere, division head of KEMA (now merged into DNV-GL), in 1987 held a presentation for the Commission on Steam Technical Questions (Commissie Stoomtechnische Vraagstukken, CSV) on causes of forced unavailability of power plants. Based on SEP (Joint Electricity Producers, Samenwerkende Electriciteitsproducenten) and KEMA damage investigations it was shown that, in accordance with the Pareto principle, about 80 % of the forced unavailability (FOR) was caused by only 20 % of the components whereas half of the FOR was contributed to only 12 "components", see table I.

It was concluded in this presentation that a thorough analysis of failures for these 12 components followed by feedback to the utilities involved and measures with regard to these components, would lower the FOR substantially.

Table I. Early (1987) Pareto analysis showing components contributing to 50 % of forced unavailability.

Component code (BES)	Description
009	general (no details)
100	boiler (no details)
112	evaporator
113	superheater
121	drum
122	headers
300	steam turbine (no details)
311	HP turbine house and inlet section
315	HP turbine rotor
344	LP rotating blades
371	turbine bearings
500	generator (no details)

BES = Bau Element Schlüssel

KEMA as supported by the CSV and the Group of Directors of Dutch Utilities (VDEN = Vereniging van Directeuren van Energiebedrijven Nederland, with KEMA director Mr. van Erpers Royaards as a secretary) decided to regularly visit power plants by a KEMA investigator that would discuss in detail failures and damages, especially for the 12 components mentioned before. This was to be a 3 year project, with a yearly feedback to CSV, after which a decision for continuation would be made.

The project was duly set up with 6 investigators (so called Storingsbezoekers) that mid 1988 started to visit the Dutch plants, in total about 50 production units. The background of the Storingsbezoekers varied from damage investigators, a reliability analyst (the writer of this paper) and some chemists. From meeting notes at the end of 1989:

- Redundant systems: It was agreed at the meeting that the Storingsbezoekers would ask if the failures of some redundant systems such as coal mills, feedwater pumps, etc. could be reported to KEMA
- Coding of gasturbines: No coding existed yet while the sector was moving to hot wind box reporting of plants. A German code (BES for gasturbines) was in progress of development and could be used.
- Details of failures: An extension of the DBASE3 database was constructed and slowly being filled
- Memo on procedures: As two of the Storingsbezoekers were not present, this was shifted to the next meeting
- Coding for causes: To be discussed further, a coding would certainly not replace free text and explanation of the failure

- Calculation spare HP turbine: After a memo on the subject for one of the plants, this was converted into an advice (later followed by a similar advice on a spare transformer for that plant)
- Model for the costs of FOR: A memo was discussed with SEP, found to be mainly correct, but it was further detailed for the bonus-malus system as cycling and reserve plants had a day-night tariff in this system. It was found that reserve plants were paid per calendar hour rather than per operating hour.
- The number of failures over 1989 was somewhat lower than expected, possibly due to holidays and Storingsbezoekers were asked to activate their plant contacts.

The general status of contacts was a discussion point at each monthly meeting. It appeared from the contact at one of the plants that the company involved was thinking about a new-to-build plant with 5 – 6 large gasturbines (Eems 95-96 CCGT). Mr. van Otterloo, head of the Risk & Reliability section of KEMA, was to visit the newbuilding organization. At another plant, discussion of failures was lagging behind as the contacts had insufficient time, there was disagreement about information exchange, etc. It was agreed that higher management would send a formal note to this plant. At another plant, discussing failures had just started and the Storingsbezoeker was confident that details would be available. Another Storingsbezoeker noted that his contact would leave the plant and he therefore had to start contacting again. The contribution to the database was “voluntary” and sometimes actively discouraged by management however being too late to keep the ghost in the bottle.

It was agreed to organize for the people at the plants filling in the forms to have a feedback day again, similar to the one early in 1989 (presentations, feedback on failures, etc.)[1].

As the project leader had a family reunion at the USA, this was also used to visit EPRI and it was found that EPRI had the same problems, being how to motivate personnel at plants, feedback and quality of information in the database. In order to have specific problems at components solved, EPRI made the Boiler Tube Handbook.

<u>Does this sound familiar?</u> The components involved for conventional power plants such as evaporators and superheaters are still causing problems, motivation can only be improved by showing interest in helping the plant solving failures, knowledge of the plant characteristics from an initial plant walk down (not to be repeated, only when necessary) and feedback on what is being done with the data. Gathering maintenance information as well as detailed component information for every component failed threatens to enlarge the amount of information causing too much effort by all involved.

<u>Is this big data?</u> At that time it was, as the database soon could no longer be handled on floppies or other digital information bearers at that time. In the end, 27.000 failures were documented over the period 1976 – 1993 which is still accessible today using Excel.

[1] This was continued every year, resulting in the need to actually produce results. Evidently this day was used for information exchange between utility experts as well, ranging from maintenance people to former operators filling in the forms and newbuilding experts

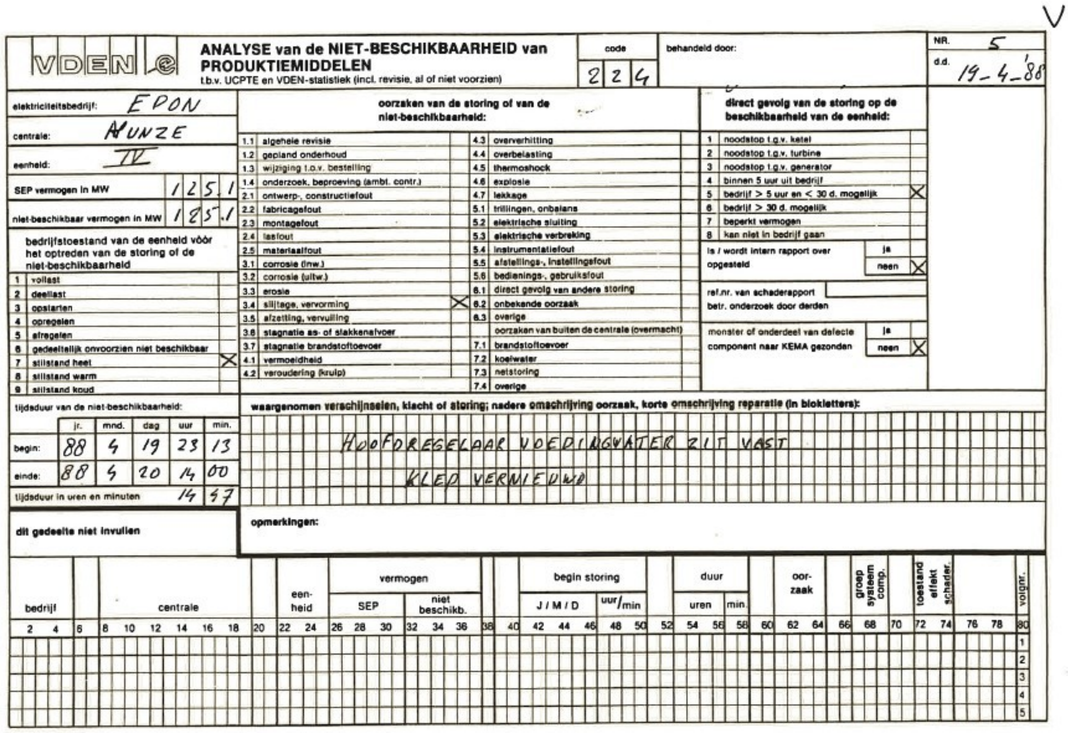
Was the "new" database replacing other databases? No, the VDEN database and failure notes to SEP and internal in the company were still to be filled in, therefore soon KEMA started to use the VDEN coding forms except for the cause & effect coding for reasons explained next. As the VDEN data had already started in 1976 based on the VGB coding and data gathering whereas KEMA started in 1988, why not use the "old" data also? A secretary was hired to type in the forms before 1988 up to 1976 as the free text on the forms was NOT gathered as there was insufficient space on the mainframe for this text. When asked for the main direct cause of failures at power plants, the secretary answered "leakage" immediately, having typed that in a multiple of times. Human errors as a direct cause for failures were only scantly filled in, for evident reasons. As the resources for analysis of the VDEN data at SEP's main frame (Univac) were dwindling, KEMA was asked to yearly report to VDEN also.

Internal electricity producer forms were not discontinued as is the case at present, with several power plants reporting in Excel to the main office. The main reason? Excel is simple to fill in. A thorough analysis apart from a Pareto type of analysis is somewhat more difficult!

Were the earlier databases sufficient? By no means as SEP was not particularly interested in improving the plants by adding detailed information as this was and is the task of the utilities. Therefore in fact it was registered only if the generator would provide power when needed. Therefore no action was taken for instance on a large number of forms with the text "VUK tijdens VIK", translated as "Fire out of Boiler when Fired applied to Boiler", the result of a multiple of failures at a plant where the operators & administrators did not see any follow up on their messages and their burden in filling in the data.

Were the earlier databases specifically made for 1 task only? No, the SEP database on registration of failures also was used for the Bonus Malus system based on colleague averages. At one plant this led to not to report on fan failures, this could be repaired during the night as the plant would not be asked by the system. However, one smart manager decided for his plant already scoring good in the Bonus Malus system to report EVERY failure as the plant would be punished the next year as the average would go down. Plants that did not score well had, as numerous information had to be supplied, problems with the amount of hours needed for Storingsbezoekers.

An example of a VDEN form is given in figure 2.1. It is a typical description with sufficient quality for a discussion at the plant ("Hoofdregelaar voedingwater zit vast, klep vernieuwen" meaning "Main controller feedwater is stuck, valve replacement"). Yet it is insufficient to find out how this failure could be been prevented. At this very moment we still don't know how much on how critical components are maintained, the effectiveness of such maintenance and the decisions for design reviews to lower the amount of failures. Including detailed maintenance for a large number of components would cause TOO MUCH big data and TOO MUCH effort on maintenance people / engineers whereas design reviews would need interaction with the designers on details. This strikes the typical gap between OEMs not disclosing details for commercial reasons and the operators not having sufficient details and being reluctant to share information and not receiving much back. The author had only seldom the opportunity to improve this with boiler companies Schelde and Stork on the so-called Kolendag (Coal plant day) and during a Design Review at the new Eems 95-96 CCGT and the Magnum plants, in which RAM Guarantees were incorporated. This also involved a FMECA by the OEM together with the maintenance experts / operators. A research project called Black Book, with constructions and operation not-to-be-repeated was unsuccessful as resources were lacking and students were unable to realize the Book due to insufficient expert knowledge. An example of such a construction is given in figures 2.2 and 2.3.



VDEN
ANALYSE van de NIET-BESCHIKBAARHEID van PRODUKTIEMIDDELEN
t.b.v. UCPTE en VDEN-statistiek (incl. revisie, al of niet voorzien)
code: 2 2 4
behandeld door:
NR. 5
d.d. 19-4-'88

elektriciteitsbedrijf: EPON
centrale: HUNZE
eenheid: IV
SEP vermogen in MW: 125.1
niet-beschikbaar vermogen in MW: 125.1

bedrijfstoestand van de eenheid vóór het optreden van de storing of de niet-beschikbaarheid

1	vollast	
2	deellast	
3	opstarten	
4	opregelen	
5	afregelen	
6	gedeeltelijk onvoorzien niet beschikbaar	
7	stilstand heet	X
8	stilstand warm	
9	stilstand koud	

oorzaken van de storing of van de niet-beschikbaarheid:

1.1	algehele revisie		4.3	oververhitting	
1.2	gepland onderhoud		4.4	overbelasting	
1.3	wijziging t.o.v. bestelling		4.5	thermoshock	
1.4	onderzoek, beproeving (ambt. contr.)		4.6	explosie	
2.1	ontwerp-, constructiefout		4.7	lekkage	
2.2	fabricagefout		5.1	trillingen, onbalans	
2.3	montagefout		5.2	elektrische sluiting	
2.4	lasfout		5.3	elektrische verbreking	
2.5	materiaalfout		5.4	instrumentatiefout	
3.1	corrosie (inw.)		5.5	afstellings-, instellingsfout	
3.2	corrosie (uitw.)		5.6	bedienings-, gebruiksfout	
3.3	erosie		6.1	direct gevolg van andere storing	
3.4	slijtage, vervorming	X	6.2	onbekende oorzaak	
3.5	afzetting, vervuiling		6.3	overige	
3.6	stagnatie as- of slakkenafvoer			oorzaken van buiten de centrale (overmacht)	
3.7	stagnatie brandstoftoevoer		7.1	brandstoftoevoer	
4.1	vermoeidheid		7.2	koelwater	
4.2	veroudering (kruip)		7.3	netstoring	
			7.4	overige	

direct gevolg van de storing op de beschikbaarheid van de eenheid:

1	noodstop t.g.v. ketel	
2	noodstop t.g.v. turbine	
3	noodstop t.g.v. generator	
4	binnen 5 uur uit bedrijf	
5	bedrijf > 5 uur en < 30 d. mogelijk	X
6	bedrijf > 30 d. mogelijk	
7	beperkt vermogen	
8	kan niet in bedrijf gaan	

is / wordt intern rapport over opgesteld: ja / neen X
ref.nr. van schaderapport
betr. onderzoek door derden
monster of onderdeel van defecte component naar KEMA gezonden: ja / neen X

tijdsduur van de niet-beschikbaarheid:

	jr.	mnd.	dag	uur	min.
begin:	88	4	19	23	13
einde:	88	4	20	14	00
tijdsduur in uren en minuten				14	47

waargenomen verschijnselen, klacht of storing; nadere omschrijving oorzaak, korte omschrijving reparatie (in blokletters):
HOOFDREGELAAR VOEDINGWATER ZIT VAST
KLEP VERNIEUWD

opmerkingen:

dit gedeelte niet invullen

bedrijf | centrale | een-heid | vermogen: SEP, niet beschikb. | begin storing: J / M / D, uur/min | duur: uren, min. | oor-zaak | groep systeem comp. | toestand effekt schader. | volgnr.
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80
1 2 3 4 5

Figure 2.1 Example of a VDEN unavailability registration form.[2]

[2] Regrettably the gentleman who filled in the form is no longer with us. When the Storingsbezoeker saw an OEM Maintenance Manual Handbook and asked for permission to study this, he was OK-ed however it was mentioned that the plant no longer carried out maintenance in such a way.

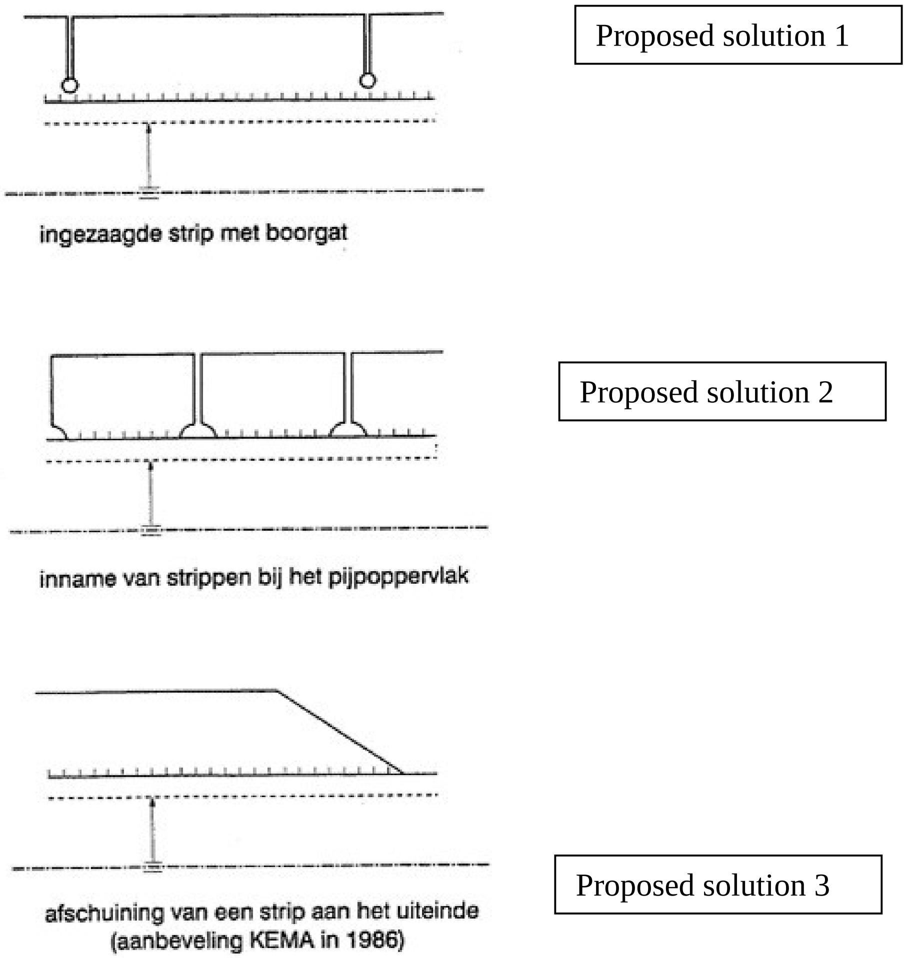


Figure 2.2 Example of recommendations to lower stresses at damage investigation 1986.

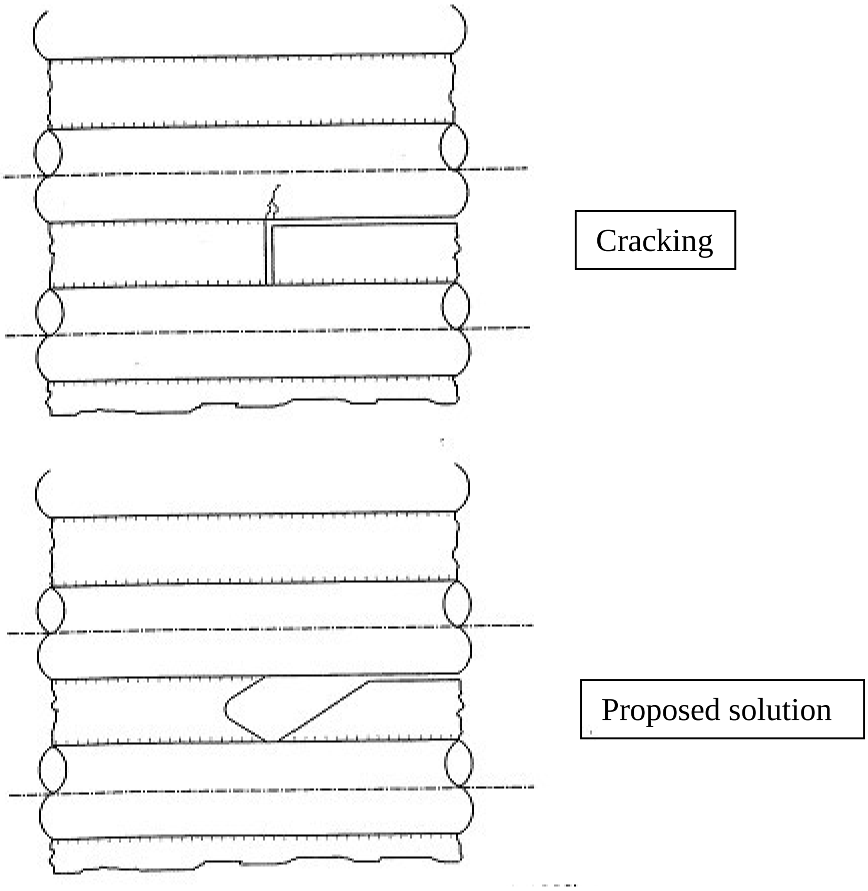


Figure 2.3 Cracking of evaporator wall in a new plant (1993).

From a project perspective, the Storingsbezoekers were successful. Many reliability analysis results were applied at Dutch power plants, carrying out a FMECA as well as a Pareto analysis is now standard, new power plants were better than their predecessors, etc. RAM models were generated for most Dutch plants much appreciated by newbuilding departments (optimum number of coal mills, feedwater pumps, estimates for new constructions, ageing investigations) and questioned by some experts at the utilities ("we already know the problems in our power plants, what does this model bring").

Sometimes the author wonders if the FOR of the plants was bettered due to investigators as well as management being interested in FOR itself causing more motivation at the plants. Such motivation and corresponding efforts were surely present at Maasvlakte MV-1 and MV-2 causing 2 years at 0 % FOR. This could not be held as external and internal boundary conditions changed. One cannot help noting the reduction in total FOR after 1988 (start of the project) and increases in 1994-1995 (start of liberalization) in figure 2.4.

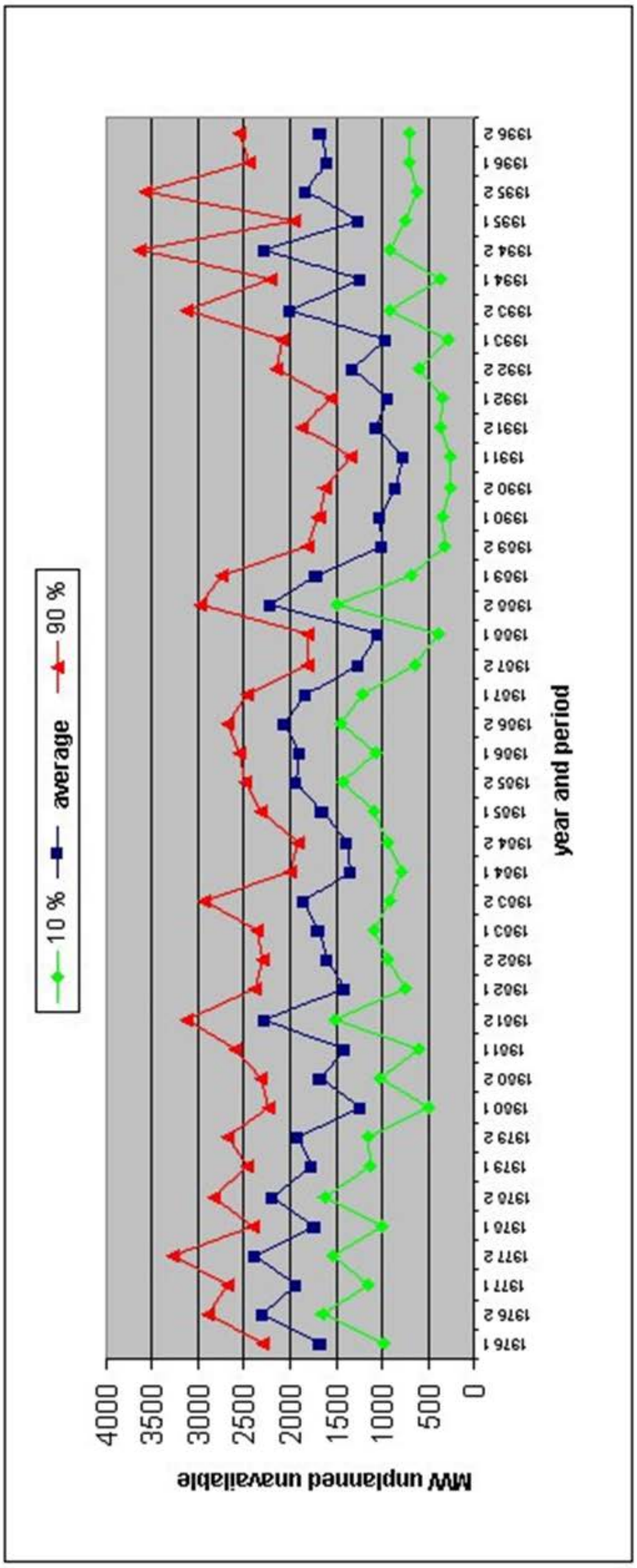


Figure 2.4 Development of MW unplanned unavailable.

Another success was that KEMA was ordered to aid in a new SEP database that was realized in 1993, targeted to newbuilding, operation and maintenance departments. Each SEP plant could fill in the data as well as analyze and generate reports using separate communication computers and a statistics computer, for digital safety reasons physically separated from SEP's mainframes. The database was realized by an external software company at substantial costs. Users had access to detailed information at the plants within the company as well as aggregated information within all SEP plants. This system was in operation for about 2 years until SEP was disbanded due to liberalization. It worked properly however the effect of a failure on the plant being a time series was separated from a record for full description of the failure to be consistent with the overall MWhrs etc. not delivered. Therefore filling in the database was somewhat cumbersome. Such issues are still valid today. At that time an automatic coupling with SCADA could not yet be implemented.

3 VGB KISSY database

Data for assessing the Reliability Availability Maintainability (RAM) of power plants have already been gathered for a long time at VGB. Due to precise definitions at VGB, availability assessed from plants around 1975 can still be compared with availabilities of power plants now. For a full description of KISSY[3], reference is given to [1] and [2].

In the Sixties as a result of the favorable economy, power plants could hardly be built in time to supply the load and the question was how with an existing set of plants, maximum power could be drawn from these plants. This both lead to early benchmarking efforts for the "best" power plant as well as investigations what was causing forced unavailability. In terms of VGB statistics this still has as a result a double line for analysis and reporting: data for the availability of plants as a whole as well as for the unavailability of components within the plant. Especially for the analysis of unavailability of components, it is necessary to pinpoint the causes of unavailability. This resulted about 1980 in a coding system (Schadens-Merkmal-Schlüssel (SMS)) defining the amount of hindrance from events resulting in unavailability (trip, power curtailment, etc.), the causes of the problem, etc. However, due to the extensive codes as well as the large amount of work involved at that time, an improved coding system (Ereignis-Merkmal-Schlüsselsystems (EMS)) was defined and became operational in 2000. At about the same time, as a next step, in definitions the point of view of the grid operator was incorporated with respect to dependable power from a plant with external influences acting on that plant. This point of view was incorporated in the 1995 definitions. For definitions on the coding systems, reference is given to [3].

The KISSY database of VGB was realized in the MESAP software from the firm Seven2one Informationssysteme für Energie- und Unweltplanung GmbH and is now one-and-a-half decade in operation. A working group at VGB and some major utilities are continuously working to keep KISSY up to standards using a new database background, introducing new parameters for analysis, using commercial indicators, on-line reports rather than spreadsheets send over mail, etc.

[3] KISSY = Kraftwerks-InformationS-SYstem

New data are easy to input both from the Internet as well as by sending Excel blocks of data to VGB Essen. All input formats are available in the German, English, French and 4 other languages. Part of the input sheets, which are available at the Web (www.vgb.org, search for KISSY), is shown in figure 3.1.

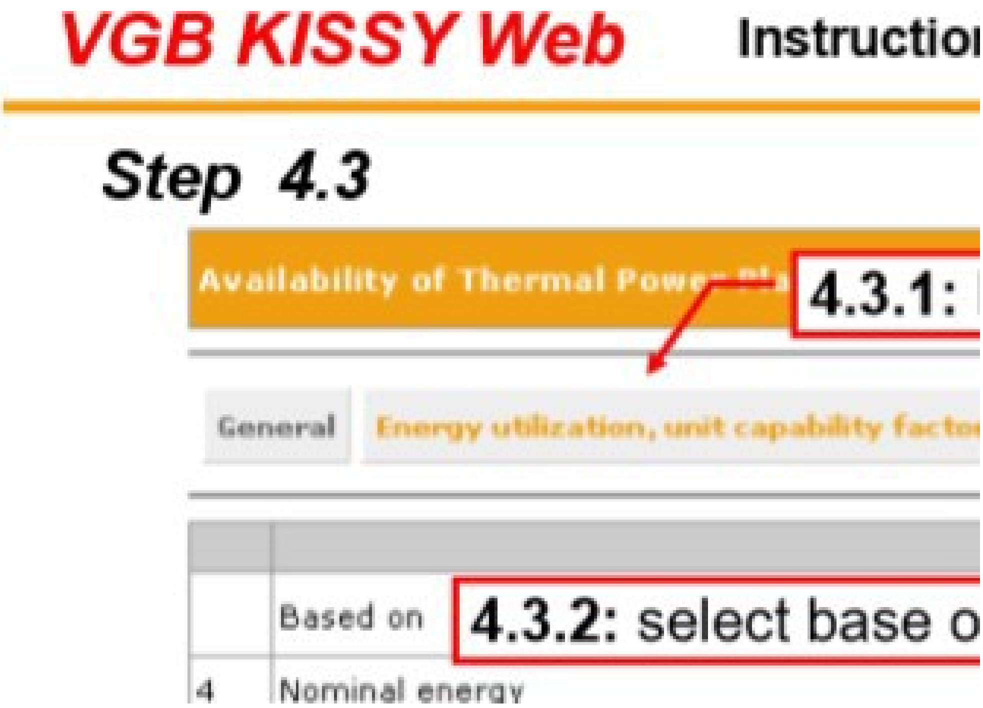


Figure 3.1 Web input form for availability data.

Targets that have been reached by KISSY are:

- Central, powerful database system
- Relational database with input directly from Internet
- Systematic check on input data to improve quality
- Access security separately configured
- Full history of own data-inputs available for companies that have inputted in KISSY
- Conversion without loss of information of the old data
- Fast workflow from data input to reports

Examples for both parts (availability as a whole and unavailability due to subsystem failures) of the database are given in figure 3.2 and 3.3.

Reports per plant can be generated on line and are sent to the requestor in spreadsheet format, however only for those users that contribute data to KISSY with a comparison to other plants (peer group). Each year, standardized reports on performance indicators and analysis of non-availability of power plant components over a specific year as well as during10 year periods are available from VGB. At this moment, some 300 power plants (with international users also from outside Germany) are present. EURELECTRIC has decided to incorporate VGB for all technical information on power plants, with the result that the Therperf data on availability are also gathered and analyzed parallel to KISSY.

The operating aspects are coded in accordance to the VGB event characteristics key (EMS). 11,452 unavailability incidents were recorded in the year 2017, i.e. during the period under review, 2008 to 2017, a total of about 84,000 unavailability incidents was evaluated. The evaluation of unavailability of thermal power plants at hand is covering the operating period 2008 to 2017 with operating parameters of a total of 303 power plant units, incorporating conventional hard-coal and lignite fired plants, CCGTs, open cycle GTs, nuclear plants and, recently, hydro storage plants.

An example of the availability (on energy basis) is given in figure 3.2, showing a rise in energy unavailability from about 5 % in 2013 to over 10 % in 2018 to the main building blocks in several grids being hard coal fossil fired units with a capacity of 500-1000 MW and Combined Cycle units with a capacity of 350 – 500 MW.

Some explanations of the rise are discussed in the paragraph on cycling. However the data in the database are in no way homogeneous as utilities enter or leave, plants are commissioned and decommissioned, economic circumstances in a country may change, etc. Therefore it is important to carry out a proper statistical analysis taking such factors into account. This is discussed later in the paper.

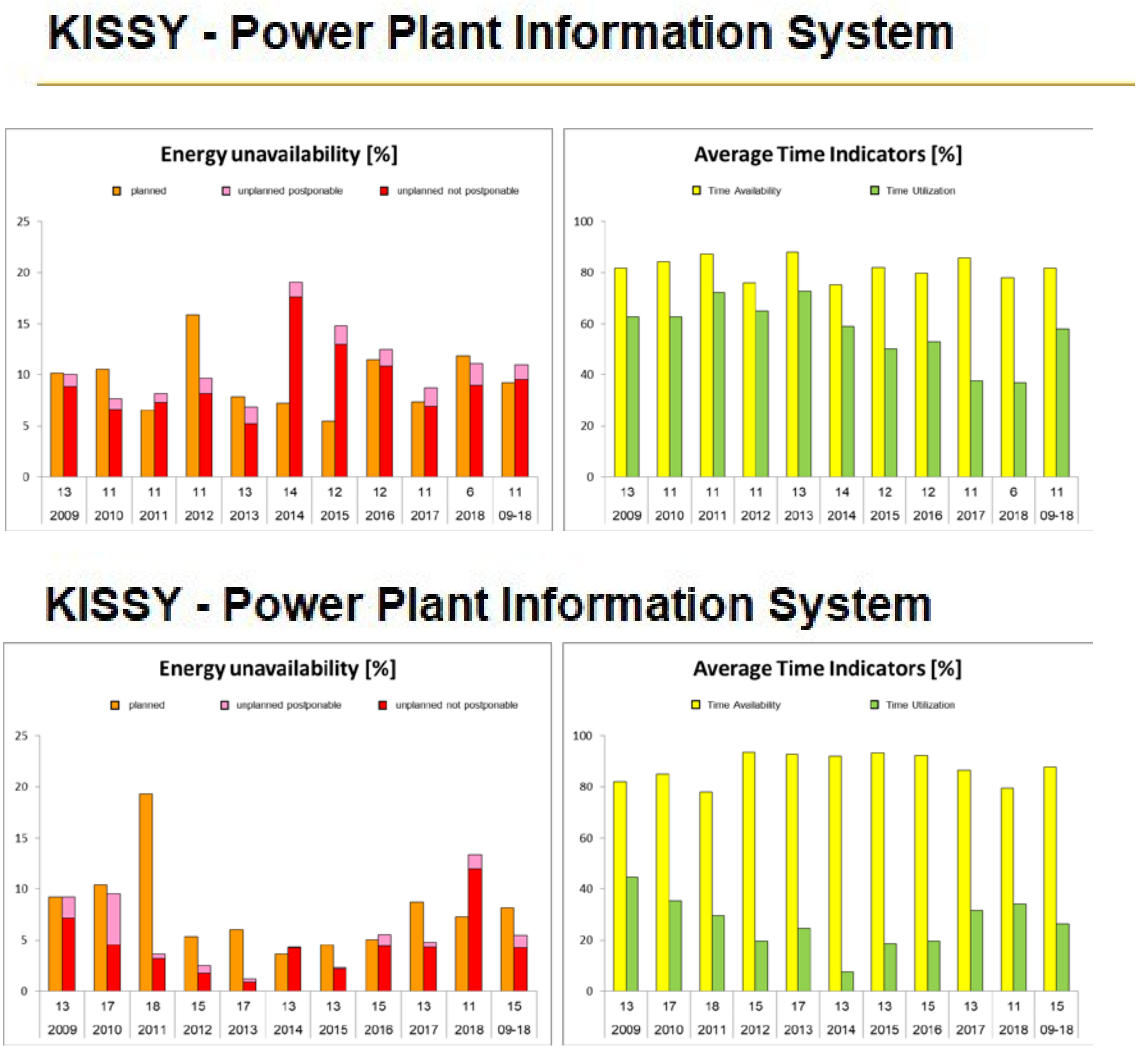


Figure 3.2 Energy unavailability of hard-coal fired plants and combined cycle units.

The rising forced unavailability should be further explained, as not all plants contributing to the availability part are also contributing to the unavailability part. For instance not all plants are being KKS-coded and are using country dependent coding for subsystems. While conversions are possible with the majority of coding automatically, still some manual handiwork is needed.

Another issue is with the components contributing to forced unavailability. Figure 3.3 shows that still components such as HAD = evaporator and HAH = superheater cause much unavailability. The number of failures for coal fired plants is dominated by HF = coal milling equipment. Now, not every plant has the same number and size of mills, mills may be redundant, some plants fire biomass using the coal milling equipment and/or separate biomass mills. Similarly, for combined cycles the number of generators or gasturbines may vary per plant. In the yearly report this is not (yet) taken into account.

KISSY - Power Plant Infor

Analysis of unplanned un

Unavailability incidents per u			
not postponable			po
KKS	Count	%	KKS
Σ H	31.33	47.69	Σ H
Σ HF	22.17	33.74	Σ HF
HFC	5.82	8.87	HFC
HFB	2.19	3.34	HFB
HFE	0.59	0.89	Σ HA
Σ HA	4.02	6.11	HAD
HAD	1.18	1.80	HAH
HAH	0.79	1.20	Σ HT
HAJ	0.55	0.84	Σ HN
Σ HT	1.04	1.58	HNC
Σ HD	0.81	1.23	Σ HL

Figure 3.3 Major subsystems causing unavailability.

Yet, progress is being made on this subject as well as the subject of adding other types of plant into the database. Successfully, storage hydro plants were incorporated recently. The KISSY group is for a long time discussing input of wind turbines. Evidently, as discussed later in this paper for wind turbines, the number of records will increase substantially, wind effects (fuel!) have to be taken into account, etc.

Other databases in which power plant unavailability data are gathered are NERC and ORAP. NERC overall reports are accessible to the general public, however coding systems and definitions differ from VGB KISSY. Failure details are not accessible to third parties. The commercial ORAP database is financially supported by OEMs and can be very well used for benchmarking, however one has to contribute to the database in order to compare with monthly fleet reports.

By using both the availability data in the VGB database for plants as a whole and the data for its components, in the following fields of application benefits have been derived:

- Planning new power plants and deciding for plant concepts
- Deriving reference values when buying power plants and /or placing contracts
- Analyzing production for both an existing plant and a plant with limited operational experience using a detailed reliability reference model for plant unavailability, occurrence and length of full outages etc.
- Targets setting, best to be set using values from a reliability model in combination with statistical best-of-peer-group plants.
- Analysis for betterment of weak points and analysis of specific constructions that show a higher than average unavailability
- Use in Reliability Block diagrams (RBD), fault trees or equivalent reliability models to predict the failure characteristics of power plants. Components and systems defined by a 3 digit KKS-code allow modelling the dominant items in a power plant. When used for new plants, these RBDs allow optimizing for example the amount of redundancy or provide comparison material for choosing single components.
- Use in modification and/or life extension of existing plants. A database like VGB in combination with plant specific information shows both the dominant failures to improve on as well as what can be reached by careful re-engineering when carrying out life extension work.
- Estimating the effect of operating regime on a power plant: base load, cycling, reserve, seasonal load. Such analysis should be accompanied by life calculations of for instance steam chest, turbine housings, etc.
- Optimum spare parts policy by balancing the costs of unavailability and the gain in repair time by strategic spares, either hold singly or in a pool.
- Maintenance optimization, for instance judging the effect of overhaul frequency and actions taken on the forced unavailability. Is the maintenance strategy really effective in preventing plant failures?
- Dispatch of present and future plant. Failure data for a single plant are insufficient. When for instance a plant did not have a failure of its steam turbine yet during its operational life, one should still assess the probability and use this in dispatch portfolio optimization as well as for insurance purposes.

Experience has shown that benefits are a multiple of the costs connected to gathering data, contributing to databases such as VGB KISSY and analyzing both one's own plant and the peer plants from VGB.

Yet, according to the author, there are some issues with KISSY from a user point of view that are slowly being solved. As discussed before, one issue is the averaging over all blocks as if 1 super-component would be present. This is evidently not the case for duo-blocks (2 boilers for 1 steam turbine), combined cycles (for instance the MD-1 block in the Netherlands, recently taken out of operation, had 3 GTs, 3 HRSGs, 1 ST and 4 generators). Yet, by setting standard configurations being checked by the companies inputting the data, reliability parameters in VGB project 361 such as per figure 3.4 could be calculated. In figure 3.5 based on the same data, ageing parameters could be calculated. Now, it is a major effort in having the characteristics of each of 300 power plants in the database. This can only be realized by direct contact with the companies and/or the plant locations. We reached that in the Netherlands, however in an international context this is a different problem.

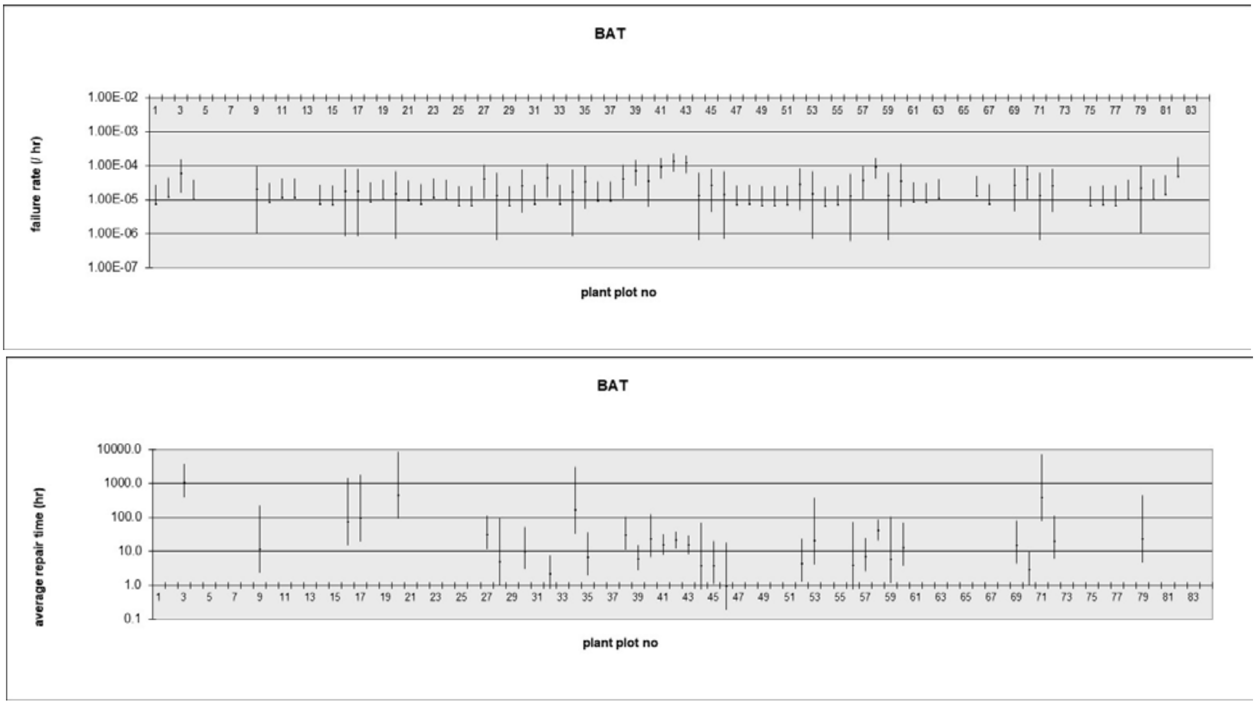


Figure 3.4 Failure rate and average repair time for BAT = step-up transformer.

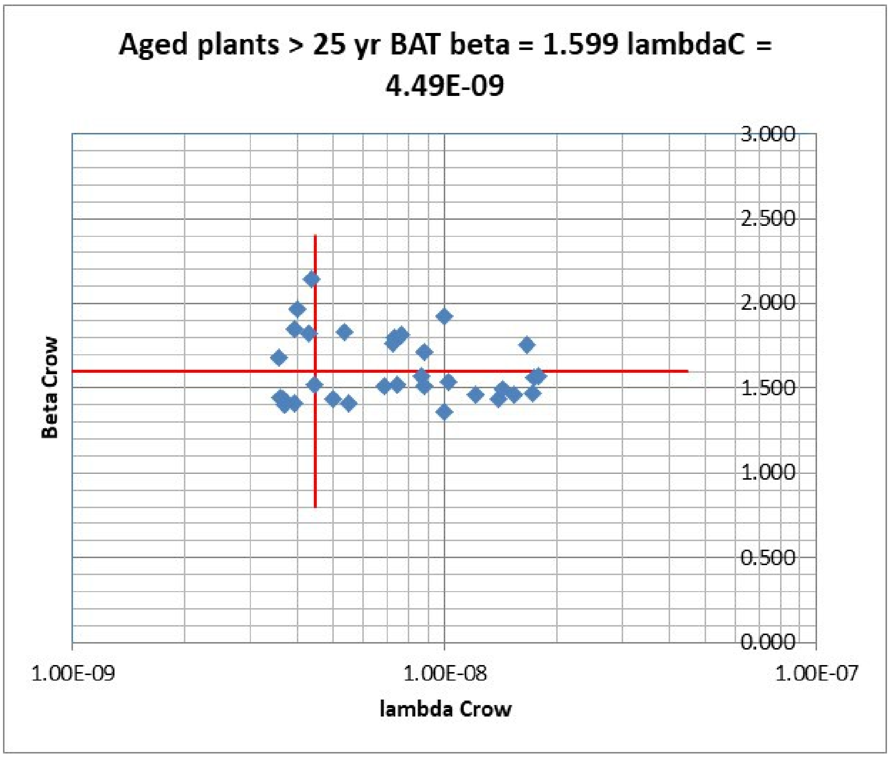


Figure 3.5 Coefficients Beta and Lambda describing ageing in a Crow model for BAT = step-up transformer.

Another problem is that there is yet no accepted model to explain forced unavailability, let alone explain as a function of planned unavailability, maintenance efforts and operational regime. In the VGB 361 project, the author simply assumed all failures being a function of operating time per year. Evidently, reserve plants show less failures than plants in daily operation. However, for a thorough analysis, failures of subsystems should be related to starts, operating hours, hours without operation (corrosion!), age of the plant, specific troublesome constructions present at a plant, etc. Interestingly, so called Causal Models are also deemed necessary in combination with big data in The Book of Why [5] from Judea Pearl.

Recently, in a VGB working group on cycling using the VGB KISSY data, the author showed that on average an increasing yearly unavailability could be explained by (more) starts, (less) operating hours per year and (less) planned availability for a specific group of coal fired plants. Yet, proper statistical methods such as factor analysis, variance analysis and regression should be applied to raw data rather than plotting averages. This can be done, although in a simplified way, in Excel. Together with such statistical methods, differences should be explained by technical reasoning checked with the personnel at the plants. For instance a backup plant in the Netherlands was found (for good reasons) to have typically other components being dominant compared to a base load plant. Aged power plants with a decommissioning date known some years in advance will not invest in maintenance, but will carry out minimal maintenance leading to overall forced unavailability of over 20 %. Examples of such maintenance were found when modelling evaporators and superheaters in the KISSY 361 project as minimal maintenance was indicated by the free text. Please note that manual analysis of free text is cumbersome, let alone having several

European languages and different amount of detail in the free text. Hopefully, big data analysis methods will help in incorporating text results for explanatory purposes. Some efforts by utilities applying black-box deep-learning models have not been successful. Minimal maintenance was also known to be present in some plants in the Dutch system.

Analysis of sister plants, having very similar maintenance schedules applied, might further help in analyze the effects of maintenance, at least in quantifying the uncertainty and variation in its effects.

Some examples of data quality are given in figure 3.6. Time aspects are well recorded, however KKS component coding is always not to the maximum depth (record 831-833, LA is coded while LAC would be more appropriate) and the detail in comments naturally varies per person filling in the data.

	A	B	C	Q	S	T
1	Unit Name	Begin[Dat	Begin[Tim	Duratic	Capacity Power plant unit	Unavail Capacity
1	Unit Name	Begin[Dat	Begin[Tim	Duratic	Capacity Power plant unit	Unavail Capacity
3	1	5/6/2009	0:00	516.80	3: 200 - 399 MW	
8	1	5/23/2003	15:29	3.43	3: 200 - 399 MW	

Figure 3.6 Example of data quality in VGB's KISSY database.

A special consideration must be given to the relation between maintenance and planned unavailability. When carrying out an analysis with KISSY data, it was asked by the plant maintenance department involved why some plants shows LESS unavailability rather than LARGE unavailability (which is the more easy to explain). This could only be established by direct contact with the plants abroad, which was not realized due to insufficient budget as well as insufficient contact with parties to provide such information. Yet forced unavailability data do show however that a surprisingly large number of failures is repeated in a week and some subsystems are being dominant over decades, indicating that maintenance should be improved (when cost effective). Maintenance is a large cost driver for many utilities and more effort should be taken to analyze maintenance for dominant subsystems.

Given some preliminary inputs a Decision Analysis model was constructed to estimate the value of KISSY to utilities. Its spreadsheet showed a Base Case value of 155 MEUR in total over a 10 year window with a total demand of 12 GW and with initially 21 * 600 MW plants. Evidently, the amount of additional work to supply data to KISSY instead of reporting the unavailability data within the company itself is minor when compared to this value.

It is interesting to see the main drivers of this value as per figure 3.7. This so-called Tornado diagram is derived from setting all inputs at their Base Case value and varying each input to Low and High respectively. By sorting those input values that have the largest effect, the typical Tornado form appears.

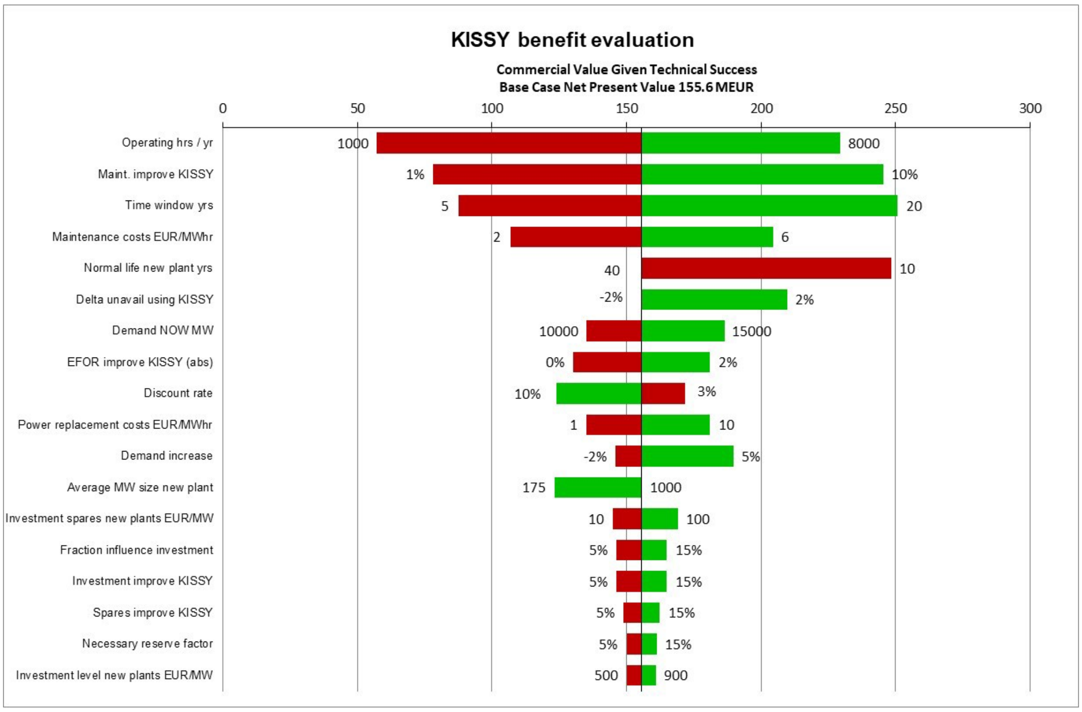


Figure 3.7 Tornado diagram showing main drivers for benefit due to KISSY.

The Tornado diagram shows that application to plants having low operation hours on a yearly basis evidently delivers less value than applying data to plants having high operation hours. Furthermore it shows that if the KISSY data indeed could be applied to reduce maintenance costs further, this is an important driver. If normal plant life is low, the value becomes higher as more new plants are necessary to replace older plant with good reasons for newbuilding departments to apply KISSY data. Interestingly, the amount of forced unavailability improvement by KISSY data etc. is NOT the main uncertainty.

If we assign a 50 % probability to each Base Case input and a 10 % - 90 % interval probability (weighing factor 25 % each) to the Low and High input values AND calculate all possible combinations, a cumulative probability function for the benefit shows up. The 10%-90 % interval for value appears to be 121 MEUR – 545 MEUR with the average being 387 MEUR. This is the case if all KISSY data are applied in the fields mentioned to the fullest. Evidently, without application, no value at all occurs.

As maintenance records in databases such as SAP are generally scarcely filled with detailed information, they do not help much in explaining the effects of maintenance on forced unavailability. They were not designed to do so, as one of the targets is to supply maintenance persons with maintenance tasks and record costs. Yet, they can be analyzed for exploratory purposes. In that way, a recent example of exploratory analysis is given by a thesis from Mr. Lany Slobbe, his doctor title given postumely, as he suddenly passed away. Mr. Slobbe analyzed and combined large routine administrative databases on patients, generally not designed for such analysis [6]. Yet, he was able to show that patients with stroke problems had a higher chance on living longer when treated in a hospital having more routine (volume of patients). Apparently in 2002 the rise in life expectancy in the Netherlands suddenly accelerated. It was found that Health Care expenditure rose rapidly after 2001, and was accompanied by a sharp rise of specialist visits, drug prescriptions, hospital admissions and surgical procedures among the elderly. His findings are consistent with the idea that the sharp upturn of life expectancy in the Netherlands was at least partly due to a sharp increase in health care for the elderly, and has been facilitated by a relaxation of budgetary constraints in the health care system. The author finds it amusing that recent discussions on how to pay for pensions on elderly people living longer may have been caused by government decisions in the past!

4 Transparency data

Data to the European electricity market, enabling proper trading at a level playing field, are available to the general public as well as to traders and electricity production companies. For example the ENTSO-E Transparency Platform provides data from 2015 on with respect to unavailability of production units, high voltage lines, etc. , hourly production data for all plants > 100 MW and data for day ahead prices and imbalance costs (on a 15 minute basis), etc..

The analysis of unavailability data copied from the ENTSOE platform as per figure 4.1 is straightforward. Also monthly outage spreadsheets can be coupled to monthly spreadsheets giving the reasons for outages. As usual, complementary information is sometimes given (in various languages which is typically European, see figure 4.2), sometimes missing. In the monthly unavailability spreadsheets some plants appeared not to be present, however the Helpdesk showed where to find them.

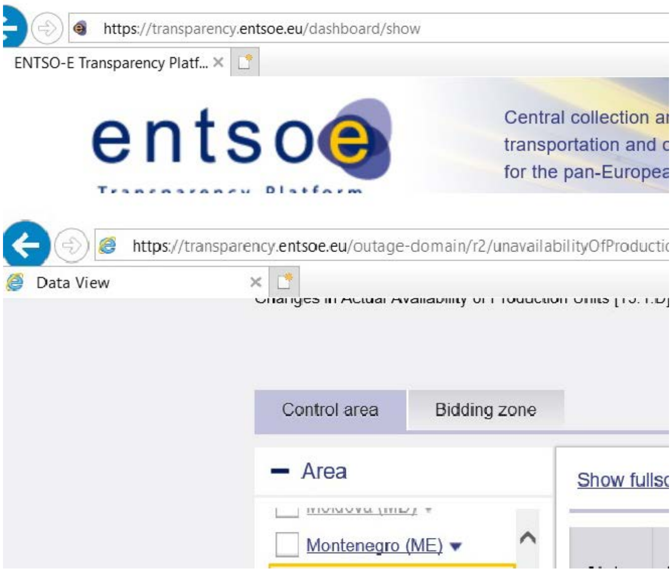


Figure 4.1 Basic Transparency data for unavailability.

	A	B	C
1			Drop Report Filter Fields Here
2			
3			
4	UnitName	reason	text
5	⊟ ABTHB	⊟ Complementary Information	Other
6	⊟ Aura	⊟ Complementary Information	Due to work in station G2 and G1 is unavailable.
7			Fault on G2. G1 already out for revision.
8			G2 failure. G1 is out for revision.
9			G2: fault on regulator. G1 in already unavailable due to revision.
10	⊟ Bakonyi Gázturbinás Erőmű	⊟ Foreseen Maintenance	
11	⊟ BATHIE	⊟ Complementary Information	Des variations de puissance pour essais sont possibles --- Power variations are expected due to tests
12	⊟ Blaiken	⊟ Complementary Information	Technical unavailability
13	⊟ Bőny Szélerőműpark	⊟ Foreseen Maintenance	
14	⊟ Borkum Riffgrund II	⊟ Failure	
15			Other
16		⊟ Foreseen Maintenance	
18	⊟ Brokke	⊟ Complementary Information	Empty sand traps
19	⊟ BROMMAT	⊟ Complementary Information	Des variations de puissance pour essais sont possibles --- Power variations are expected due to tests
20		⊟ Failure	
21	⊟ CHEYLAS	⊟ Complementary Information	Des variations de puissance pour essais sont possibles --- Power variations are expected due to tests
22	⊟ COCHE	⊟ Complementary Information	Des variations de puissance pour essais sont possibles --- Power variations are expected due to tests
23		⊟ Foreseen Maintenance	

Figure 4.2 Reasons for outages with varying amount of detail and different languages.

The overall unavailability results for the Dutch plants are given in figure 4.3. While at first sight the information seems to be correct, specific plant knowledge shows this only partly to be the case. Some remarks for the various plants, coded anonymously, are given below:

- Unavailability results for plant C are unrealistic, C being out of operation end of 2015 and prematurely closed due to some incidents. At least the full month of December 2015 would need to be marked as forced unavailable.
- Plant D is unrealistic as evidently a PWR has a yearly fuel stop.
- Plant E is unrealistic, being taken out of operation end of 2015.
- Plant F is unrealistic, having been stopped without extensive mothballing measures. It is evident that it will not be started again, therefore why 100 % planned unavailable in 2015, 100 % unplanned in 2016?
- While plant G has been extensively mothballed, why 100 % planned unavailable in 2015, 100 % unplanned in 2016? Surely as it is 100 % planned unavailable for 2017 and 2018, the coding for 2016 must be considered a mistake.
- The large forced unavailability for plant H is intriguing. However, by checking newspapers, Tennet reports, etc. it was found that the plant had a (large) transformer problem and otherwise was unable to produce (external failure) as the nearby substation had a large unplanned outage.
- Plants K – N are sister plants. The records are inconsistent for operation in 2015 – 2018 with one or more of the plants missing. Yet these plants are interesting given the large number of starts (283 for plant K, 341 for N in 2018). One of the plants might have had a generator failure as public information from its owner seems to confirm, however one needs further information to be certain.
- While plants U and V are present in the unavailability data, no operation hours were found in the monthly spreadsheets.

	C	D	E	
2	Unit Name	Unit Name	Begin year	Ca
3	A	A	2015	
4	B	B	2015	
5		B	2016	
6		B	2017	
7		B	2018	
8	C	C	2015	
9	D	D	2016	

	C	D	E	
2	Unit Name	Unit Name	Begin year	Ca
24	I	I	2015	
25		I	2016	
26		I	2017	
27		I	2018	
28	J	J	2015	
29		J	2016	

2	Unit Name	Unit Name	Begin year	Ca
44	P	P	2015	
45		P	2016	
46		P	2017	
47		P	2018	
48	Q	Q	2015	
49		Q	2016	
50		Q	2017	

Figure 4.3 Dutch results for forced unavailability.

- Plants AC and AE have finished operation in 2018. They seem to have had a very reasonable unavailability up to the last year operating, contrary to the usual maintenance (neglect) savings in the last years.
- Plants AF and AG have illogical gaps in operation at about 6500 hrs into 2018.
- The large planned unavailability of 2018 for AH can simply be explained as AH was taken out of operation early 2018.
- Plant AK was only found once (2018) and both a planned and unplanned unavailability of 0 % seems questionable.
- Plant AO was present in the 2018 operation data but not in the unavailability data
- As plants AQ and AR are well known to the author, their unavailability records can be explained (as shown later) in the sense that AQ is backup for the AR plant, both firing blast furnace gas.

A direct comparison can be made when plant unavailability records are present per event from projects etc. at the plant. Such a comparison is shown in figure 4.4 for plant A. The comparison clearly shows that all larger outages are duly reported. Some smaller outages are not present in the Transparency data. From the outage description, it appears that some these outages have to do with coal quality issues and some have to do with firing biomass. Such small outages do not have to be reported in the Transparency data but are important for plant improvement. In KISSY these records therefore are present.

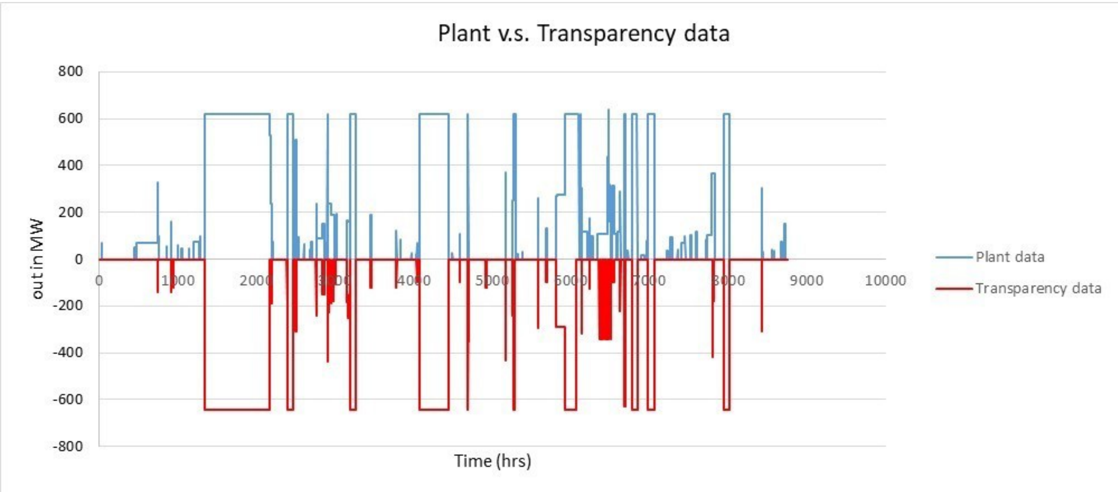


Figure 4.4 Comparison Transparency data with plant data.

Other Transparency plant unavailability records confirm for example the operation of plant AQ as a backup for AR (figure 4.5). This type of operation may be important for the future as conventional plants may be backup for renewables. The backup was reasonably successful except for plant AR tripping during an overhaul of plant AQ or vice versa or cases where AQ, as it is older and less automated, cannot be started fast enough to prevent both AQ and AR out of operation. The latter shows the importance of handling degradation failures compared to failures due to tripping of the plant. Regrettably, for instance KISSY data seem to indicate that more modern plants (having more instrumentation) when having an outage seem to trip more often compared to older plants. One wonders if all such trips are necessary when weighing protection of equipment against the costs of having the plant fully out of operation.

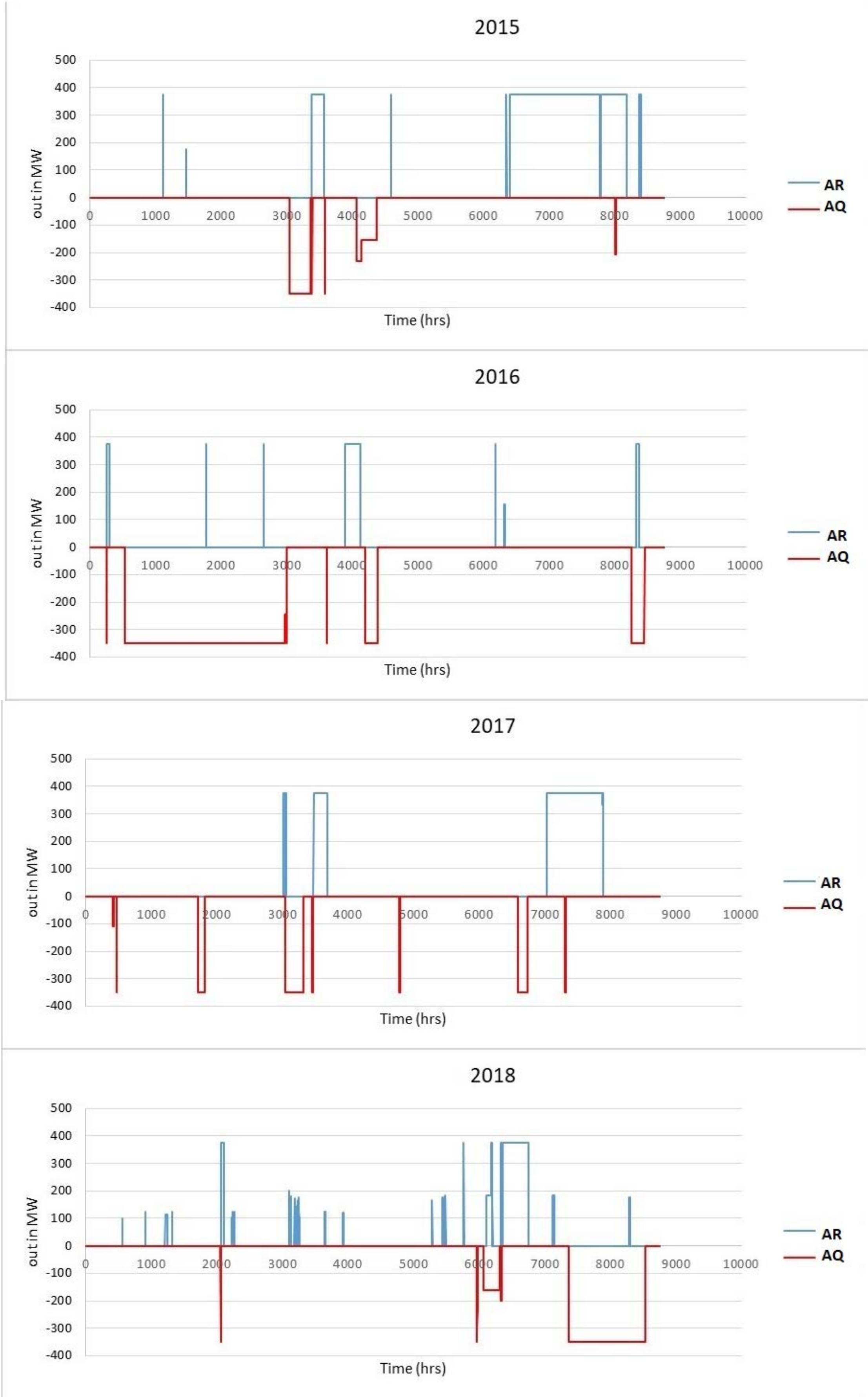


Figure 4.5 Plant AQ operation as a backup for plant AR.

Analysis of plant operation is less straightforward. While simple daily overviews are present, they cannot be used to download a set of plants over a prolonged period. The amount of data is staggering and causes the need of special tools to handle these. While analysis started using manual downloads from the Platform, the ENTSO-E team kindly mentioned download of monthly production data for power plants. The author has used FileZilla and the free ad-in Power Query to handle large monthly .CSV files. The data could be downloaded fairly quickly. Yet the author wanted to combine 5 years of outage data with these monthly data which was found to be quite tedious. With some effort, the years 2015 - 2018 were constructed for all Dutch (> 100 MW) plants. In the spreadsheet it is easy to access generation data for any other European plant also on the list of 1656 plants. Evidently the analysis spreadsheet coupled to 6*12 times 2015 - 2018 monthly operating data spreadsheets is somewhat slow, as it means handling about 60 million records...

A very interesting exercise using this spreadsheet is to compare operation of a plant against day-ahead and imbalance prices. This was carried out by checking on an hourly basis if the price was higher than a target with the plant being actually in operation. Some figures on operation of plants and prices are shown below (figures 4.6 and 4.7).

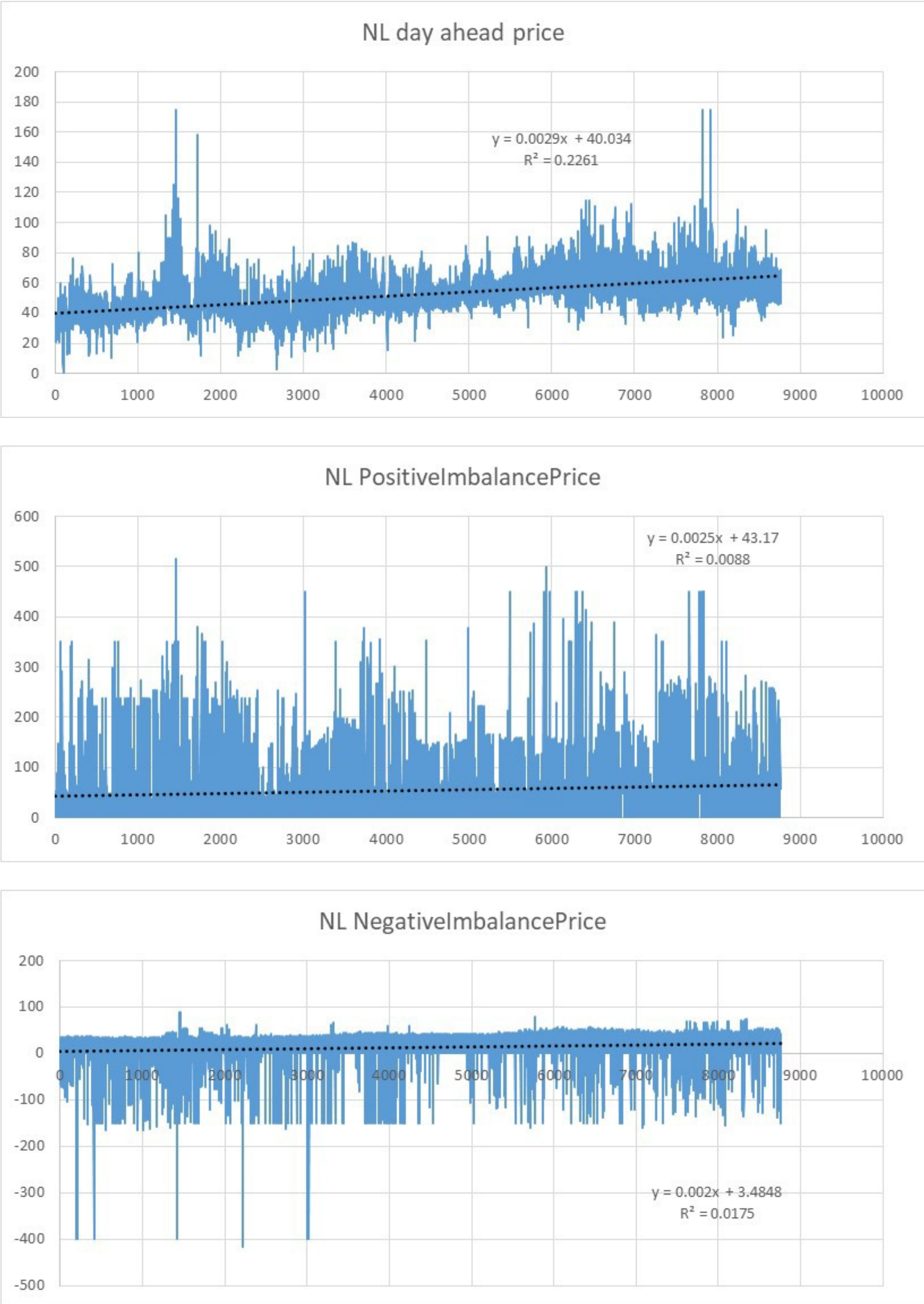
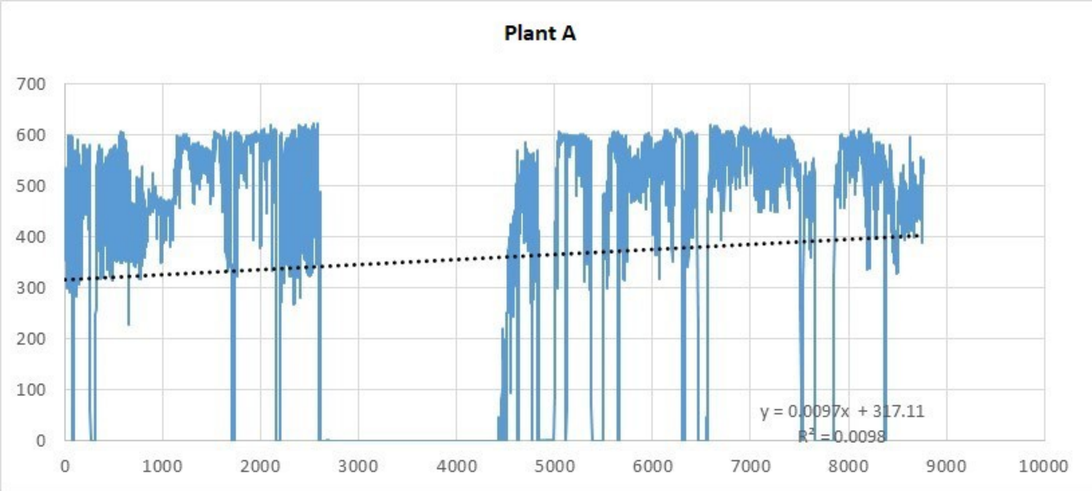
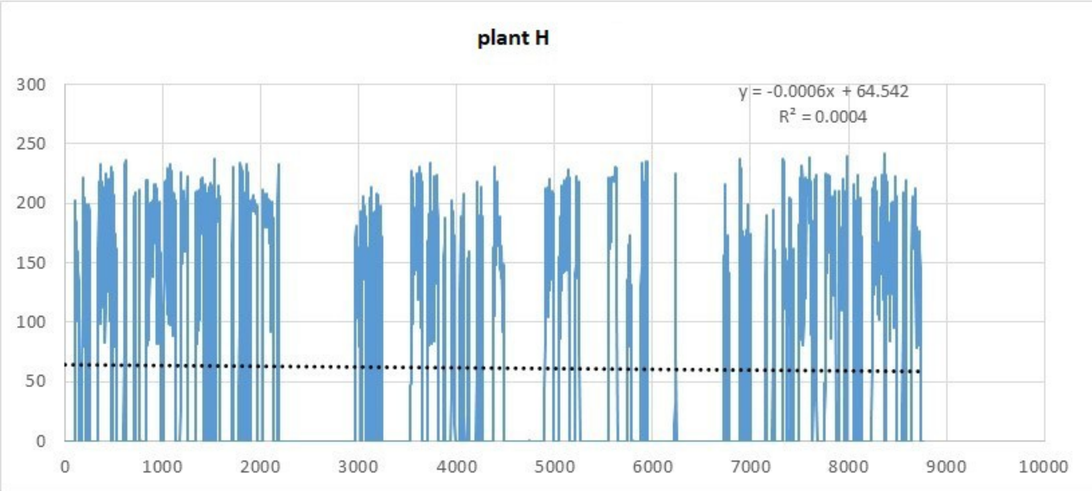


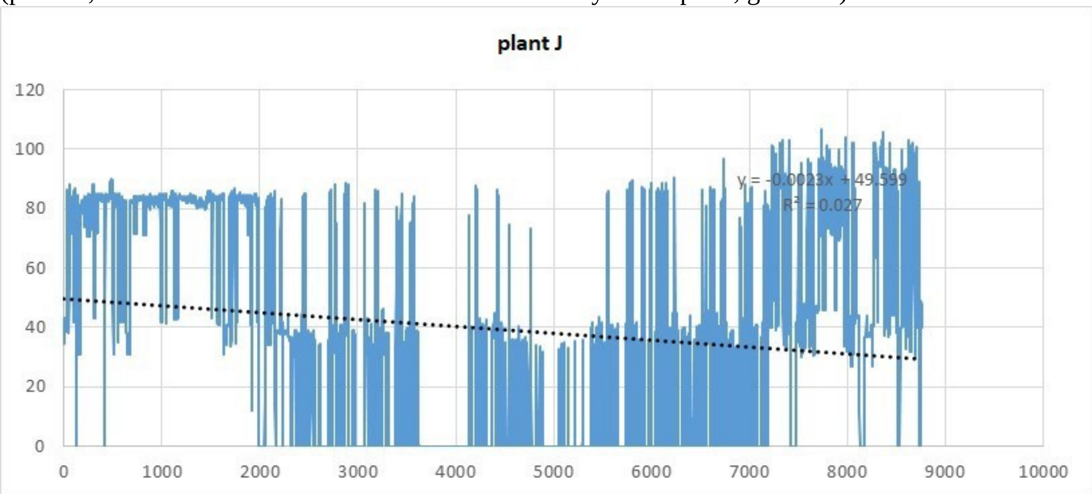
Figure 4.6 Basic Transparency prices 2018 in the Netherlands.



(plant A, no correlation with day ahead price, coal fired)

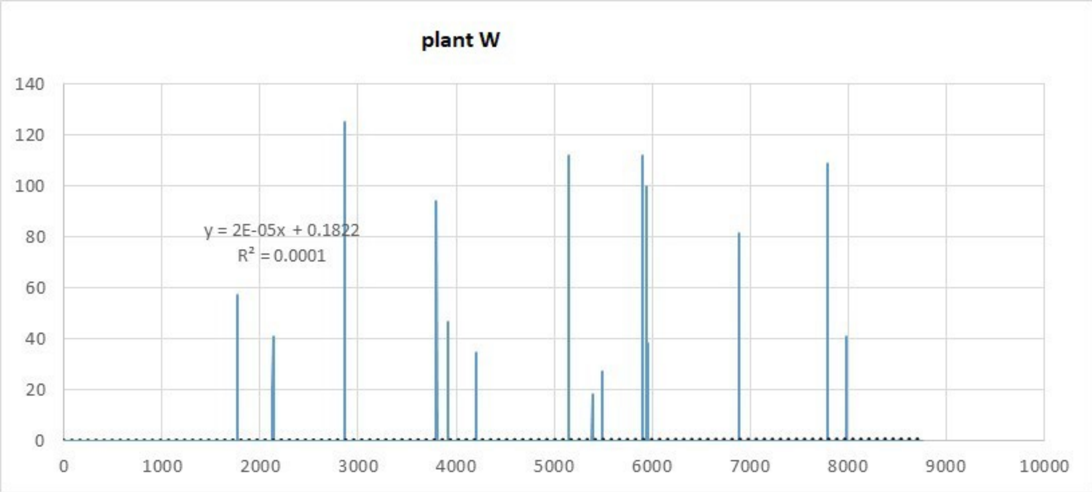


(plant H, correlation 18 % with over 48 EUR/MWhr day ahead price, gas fired)

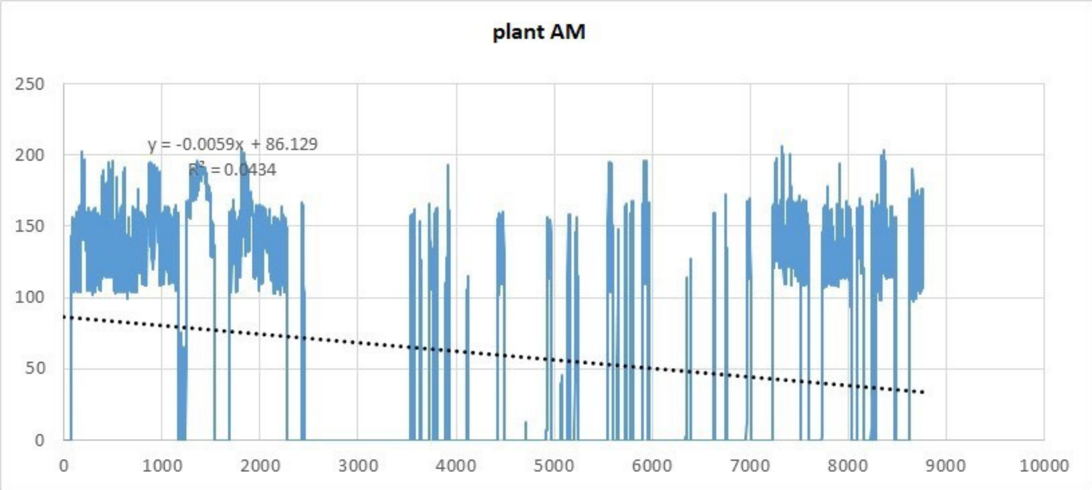


(plant J, 4% correlation with over 24 EUR/MWhr day ahead price, gas fired with 2GTs, district heating, in summer fired with 1 GT)

Figure 4.7 Some examples of plant operation from Transparency data.



(plant W, 11% correlation with Pos Imbalance Price, GT)



(plant AM, no correlation with day ahead price, gas fired district heating)

Figure 4.7 (cont.) Some examples of plant operation from Transparency data.

Some remarks on the operation and prices, as derived from the Transparency data, are given below:

- Average Day Ahead prices in 2018 are typically increasing from about 40 EUR/MWhr to 60 EUR/MWhr, resulting in more operation of CCGTs that were not “in the money” otherwise.
- Imbalance prices for supplying additional load are typically in the 50 – 100 EUR/MWhr range with extremes of over 300 EUR/MWhr. Imbalance prices to decrease load are about 20 EUR/MWhr, with a large number of negative prices at -150 EUR/MWhr and sometimes one is “given” 400 EUR/MWhr to take the plant off the grid, possibly due to (over) production of renewables.
- Typical base load plants, mainly coal fired, show no correlation at all with Day Ahead prices. Evidently similar for the NPP plant. They may have fixed contracts.
- Gas fired CCGT operation such as plant H and I do show correlation with Day Ahead prices. A correlation coefficient was maximized by increasing the price stepwise. Yet, this correlation is far from perfect as sometimes during planned

overhauls the price is high. Yet, when comparing the time series most CCGT operation seems to follow high prices, needing to be above 40 EUR/MWhr.

- Differences in price behavior occur for plant J , which supplies district heating and which has 2 GTs, 2 HRSGs and 1 ST. Apparently, in summer 1 GT is kept operating with the plant at half load. The operation of J shows almost no correlation at all with Day Ahead prices.
- Plants K – N are interesting sister plants which apparently are unable to operate all (5) of them fully for economic reasons. Plant N, being the youngest, shows the highest correlation at 50 EUR/MWhr, plant K being the oldest showing maximum (but low) correlation at 30 EUR/MWhr. All plants have a large number of starts (150 – 350) on a yearly basis however it is unclear why one of the K-N units shows an operation pattern typically having less starts than its sisters.
- Differences in pattern also occur at plant U, which may have a minimum load strategy rather than starting and stopping the plant.
- Plant W operation, being an old OCGT as remnant from a hot-box conversion of a conventional plant, shows no correlation with Day Ahead prices but indeed some correlation with Imbalance price. Yet, the OCGT seems to be a bit slow to start for "sudden" price peaks.
- Plant AM operation has no correlation with Day Ahead prices which is logical as the plant supplied district heating early 2018 (when the price was low) as well as late 2018 (when the price was high).

To conclude:

- One can learn much from the operation of individual plants from the Transparency data. However, similar to any unavailability data set encountered by the author, the data quality varies with the plant and company. It is unclear whether there is quality control in the unavailability records. If there is not, it should be improved to have a consistent set of information together with the operation data, which are not always over the same time period.
- Apart from some plants providing complementary (technical) information, in general the reasons for outages are a black box. Large duration planned outages are likely to be overhauls, however one cannot be certain. Stops with the plant out of operation can especially for CCGTs be the result of economic reasons or plants can be stopped in summer when no district heating is required (seasonal stop). In those cases, the operators are "at home", the plant cannot be started and is therefore planned unavailable. Access to other sources (the staff at the plant or occasional newspaper and yearly reports) are necessary if one is interested in the technical background, necessary for improvement.
- Unavailability records appear to be inconsistent with operating records, some values for unavailability are unrealistic, definitions for planned versus unplanned are unclear (ex-ante being the only difference is ambiguous as the time interval to declare an outage ex-ante is not well defined).

5 Wind energy

Conventional generation is and will be further replaced by wind energy. Only wind parks in the 100-200 MW have similar capacity as conventional generation and as only large turbines have a MW size (>8 MW is reached), the number of data is enormous if one wants to follow individual turbines to derive failure data.

A limited number wind park production data can be found on the ENTSO-E Transparency platform, for instance planned maintenance for the wind park or some larger grid connection problems. Yet, either no information on outages or a very limited number of outages was found (f.i. full production loss of a wind park). It is evident that to note outages of individual turbines is a lot of work which is also confirmed by the discussion on NERC wind generation, already taking place some years. While notifying wind turbine outages appear to be mandatory in the US, production companies found this difficult. Commercial firms aid meanwhile in having these data generated.

Discussion in the KISSY Working Group also has been going on for some years to introduce wind. Evidently such outages have to be generated by turbine and wind park SCADA systems, wind turbines do not have KKS-codes (a new coding system RDS-PP was developed) and lack of wind and too much wind (cut-in and cut-out of a turbine) should be taken into account. Interaction between turbines, for instance due to wake, is an issue that should be thought about.

The outage data are aggregated data for 2 wind parks (Baltic 1 consisting of 21 turbines in total 48 MW, Baltic 2 consisting of 80 turbines in total 288 MW). As according to literature one may expect 1 – 2 failures per turbine per year, unavailability records showing 2 – 5 MW losses should be abundant. They are not. This is consistent with only reporting large (order of magnitude 100 MW) outage data.

Advanced modelling (as wind data must be local) is probably used in the ENTSOE Transparency Platform wind data. Figure 5.1 shows the amount of variation in wind production (forecasted) for December 2018. It would be interesting to see if typical patterns would be present (storm season, low production, variation over the day with land warming up, etc.). Evidently, in order to keep the grid frequency constant, such variations must be balanced by conventional generation. Evidently, this will cause more wear and tear on such power plants and reserve (for the in Germany so called Dunkelflaute, meaning a period without renewable generation from sun and wind) must be kept ready and paid for. The reserve plants market is still being discussed by major players.

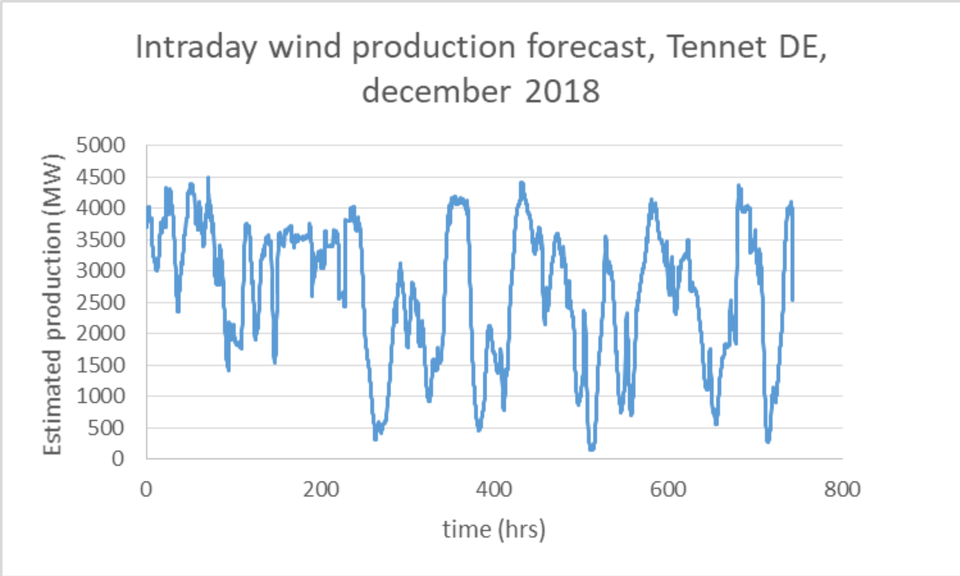


Figure 5.1 Example of Intraday wind production forecast.

The author has helped to carry out a study for the production situation on a Caribbean Island, having diesels, electricity production from a refinery as well as small size wind turbine parks. It was found that hourly wind production records were held manually and publicly assessable wind information from an airfield nearby was available. While not part of the original study, results were derived such as shown in figure 5.2 below. The results are interesting as the island demand is 100 – 120 MW with during the months analyzed 0 – 40 % wind production. Some wind production data are shown in figure 5.3. The wind data of the wind parks were correlated to wind measured about 1 km further away at the airport. Evidently, wind production by a single array of turbines is dependent on wind direction. Part of the spread may be explained by the difference between measurement point of wind and having wind production by the hour while wind speed is averaged over the hour. At least 15 Minute wind speeds, as local as possible, should help to reduce the spread further.

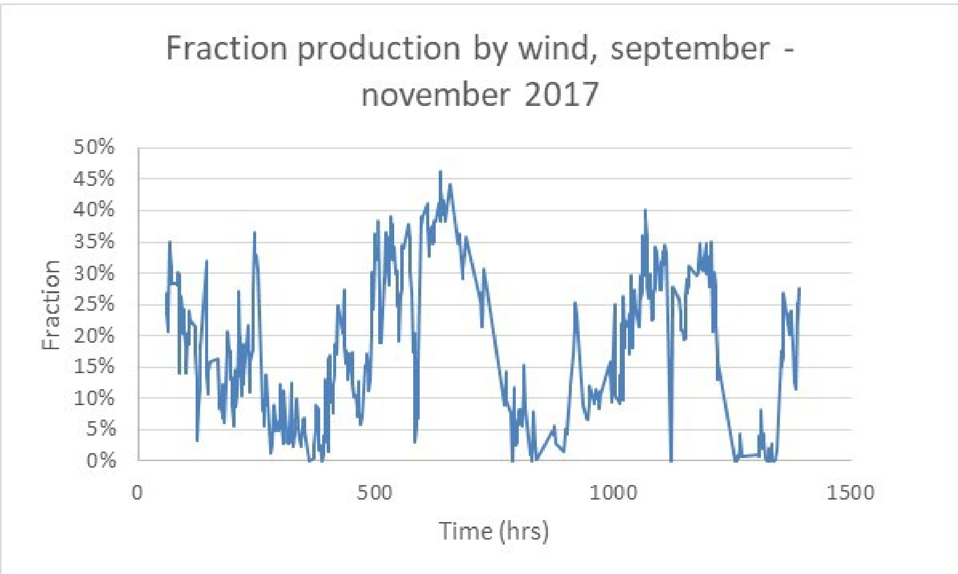


Figure 5.2 Wind production at a Caribbean island.

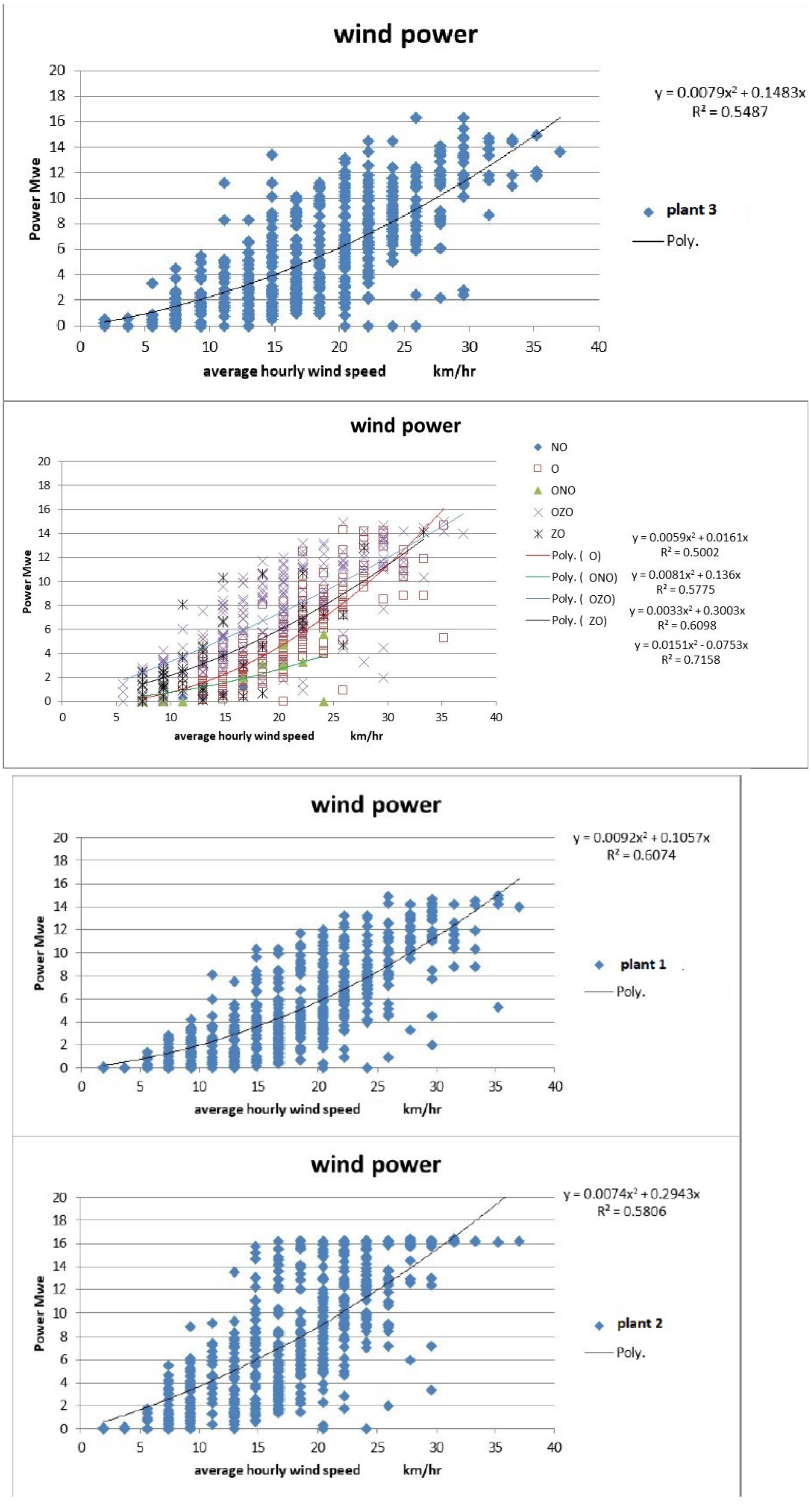


Figure 5.3 Wind production as a function of wind park and direction.

The problem for the utility company is NOT a period with much wind, as diesel generation is relatively easy to stop, decrease and increase (by varying the amount of diesels in operation). The problem is to predict wind production to such an extent that the utility is able to plan and carry out maintenance during day hours at acceptable costs and to have sufficient reserve power for months without wind. Similarly, electricity production on this island is dependent on industry (a refinery) and therefore on economic conditions.

The situation, probably less extreme, might be similar for the Netherlands in the future with a large amount of renewables in the grid and short term decision making on generation and reserves by politics. Please note the discussion in the Netherlands on coal production from relatively young large coal fired power plants, to be closed. Yet, technically this is an interesting situation as by using the characteristics of the wind turbines and local wind conditions (both on direction and wind speed) one should be able to predict wind generation and, given a sufficient window for decision making, solve the above questions for reserve power and plant maintenance.

Similar types of analysis are also possible for North Sea windfarms. One should use:

- The production curve per turbine type as given by the manufacturer
- Transparency data for total park production
- A correction for wind at hub height versus metrological (10 m) reference height
- Actual wind conditions including gustiness, if unavailable one can use for example the KNMI North Sea wind data or computer generated GRIB files
- As turbines influence another being in each other's wake, some analysis of array effects and/or wind direction

As an example see figure 5.4, which is comparable to 5.3 except for the amount of power. Also for North Sea windfarms, the amount of data necessary for analysis is substantial but doable.

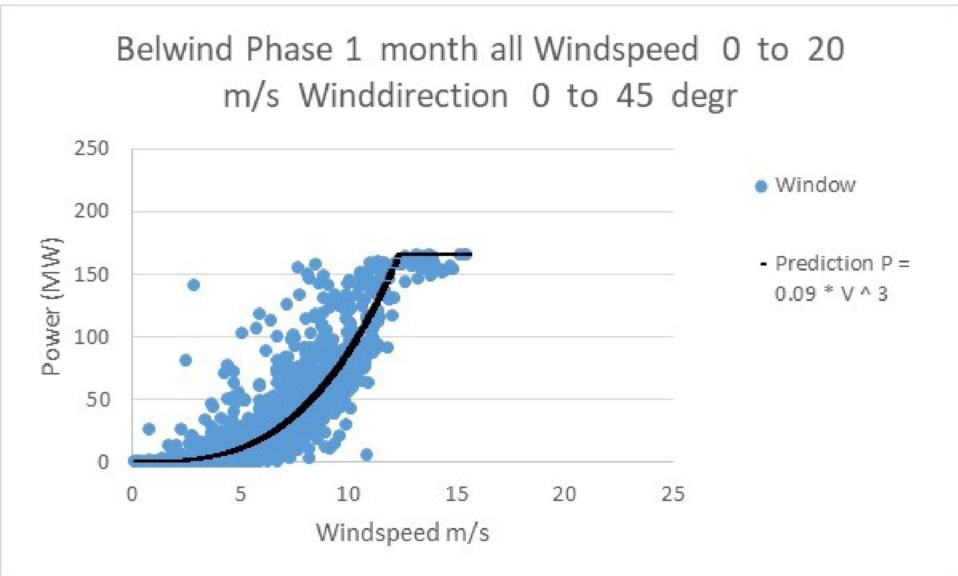


Figure 5.4 Example of production from a North sea windfarm.

To conclude:
Renewables such as wind power require high quality big data for analysis purposes. Yet the physics of wind turbines are known, prediction of wind based on GRIB files is possible and it should be feasible to have a few days ahead power predictions in order to match generation resources. To derive failure characteristics of single wind turbines in large wind parks is a major amount of effort shown as per [7].

Acknowledgements

Discussions and help was much appreciated from Mr. Stefan Prost at VGB, Dr. Ralf Uttich at RWE and the Helpdesk for the ENTSO-E Transparency data. Statements in the paper not necessarily coincide with their opinion. The same is valid for any of the companies involved or mentioned in this paper.

References

[1] Aydt, J., Lehougre, J., Uttich, R., Prost, S., VGB-database supports performance analysis VGB PowerTech 9/2017.
[2] Aydt, J., Meier, H., Bauer, F., Prost, S., Lehougre, J. Technical Benchmarking – A Tool for Better Performance, VGB PowerTech 9/2017.
[3] VGB-Richtlinie Verfügbarkeit von Wärmekraftwerken, VGB-RV 808, 6. Ausgabe 1999, as well as VGB-Powertech Terms of Utility Industry, Booklet 3, Fundamentals and systematics of availability determination for Thermal Power Plants, 7th Edition 2008
[4] Wels, H., Forecasting the reliability of components in thermal power plants using the VGB database KISSY VGB PowerTech 10/2015.
[5] Pearl, J. , Mackenzie, D. The Book of Why, The New Science of Cause and Effect, Penguin, ISBN 978-0-141-98241-0, 2019.
[6] Slobbe, L., Working with Administrative Health Data, finding solid ground in the data morass, ISBN/EAN 978-94-028-1452-1, 2018.
[7] Reliawind, European Commission project ID 212966.

Mechanical seal failure prediction in an oil refinery: a first attempt to solve the problem using a data-driven approach

L. Pereira, L. Andrade Ferreira
University of Porto, Faculty of Engineering, FEUP/DEMec, R. Roberto Frias, 4200-465 Porto, Portugal

Abstract

Despite the massive amount of data available in today's plants, maintenance teams, in an attempt to understand and predict failures, are still using rudimentary tools that, typically, can only address the influence of each influencing variable separately.

The growing gathering data capability asks for more sophisticated tools. These tools are currently available and have been highly exploited by the major technology companies in recent years [T. Żabińskia et al.]. In academia and, in particular, engineering and physics fields, the use of data-hungry tools is in high-demand and several examples can be found, from the prediction of superconductivity temperatures [V. Stanev et al.] to material discovery and design [Y. Liua et al.]. Nevertheless, the successful application of machine learning in maintenance is hindered by the messy, noisy and incomplete data commonly found in industrial environments. In this work, a first attempt to predict the failure of mechanical seals using a data-driven approach and real data collected in an oil refinery is presented.

A mechanical seal is a device used to control the fluid leakage between a rotating shaft and the housing of a dynamic turbomachine, e.g. centrifugal pump. Due to its improved performance in comparison with traditional sealing systems, such as packing (regarding safety, power losses, shaft wearing, water consumption, leakage and maintenance time), ANSI/API 682 mechanical seals are commonly used in oil refinery's centrifugal pumps that work with dangerous fluids, e.g. vacuum residue at 350ºC. Despite its advantages, this sealing solution is prone to fail due the contact between the very smooth and flat rotating and stationary rings face (two contacts of this kind by device), causing the reduction of centrifugal pump's reliability and making the system highly-sensitive to maintenance operations. Therefore, intervention decisions must be based on robust information and the device must be intervened only when it is strictly required.

For this devices, two types of data can be collected: real-time data and maintenance operations information. The first is highly-reliable (assuming proper functioning of the sensory systems) and its analyses can be easily automated. Nevertheless, autoregressive models or other strategy to reconstruct

missing data is required due the expected malfunctioning of the sensoring systems during short periods of time. Maintenance operation information is harder to analyse due its dependency on human intervention. In order to homogenize internal maintenance databases (and making it easier to perform automated analyses), effort has to be put in the developing of internal standards and formation of human resources. The combination of both data types can lead to more robust analyses and ensures that all available data is completely exploited.

In this work, a complete analysis, from data gathering to the application of machine learning algorithms to predict mechanical seal failures in an oil refinery, is performed. The main goal is clearly defined and a quantitative definition of mechanical seal failure, taking into account the current maintenance strategy, is established. Data gathering, combination and cleaning is performed and new information is created, in order to ensure the proper feeding of machine learning algorithms. Different machine learning algorithms, e.g. decision trees and naïve Bayes, are trained and evaluated. At this stage, only classification is considered due its simplicity (though the problem is inherently continuous). Feature selection and hyperparameter tuning is executed and physical interpretation of the results, whenever possible, is made. The results, although not good enough for right-way implementation of the developed system in the plant (which may be attributed to the randomness of mechanical seal failure and lack of relevant information, e.g. vibration measures), show the potential of these type of techniques.

References:

T. Żabińskia, T. Mączkab, J. Kluskac, M. Maderad, J. Sęp, Condition monitoring in Industry 4.0 production systems - the idea of computational intelligent methods application, Procedia CIRP 79 (2019) 63–67

V. Stanev, C. Oses, A. G., E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, npj Computational Materials npj 4, Article number: 29 (2018)

Y. Liua, T. Zhaoa, W. Jua, S. Shibc, Materials discovery and design using machine learning, Journal of Materiomics, Volume 3, Issue 3, September 2017, Pages 159-177

See slide presentation in Appendix 1.

Trust in AI: The role of hybrid and private governance

Asun Lera St.Clair
DNV, GTR- Digital Assurance

Executive summary

The fast advancement of AI and its deployment in many different social contexts and industries calls for a meaningful governance system that can prevent harmful applications, ensure robustness, prevent bias and violations of privacy, and provide trust without stalling the growth potential of these technologies and their capabilities to support the common good. The paper outlines some emergent perspectives on AI governance and discusses how hybrid and private governance should be a key element. Business transactions of any type are layered and mediated by both hybrid and private accountability mechanisms that self-regulate the behavior of companies, align business practices with widely held ethical principles, and enable societal trust. Assurance providers (accountant firms, auditors, verifiers, or certifiers) are guarantors that these systems and technologies work as intended. These types of governance systems are commonly referred to as private or hybrid governance, and they are very common ways to regulate industries.

This paper contributes to the literature on ethics, AI and governance, by unpacking the role of private governance could play in regulating the process of building and deploying safe and responsible AI. We argue that traditional private governance mechanisms are a key component of the needed regulatory and accountability landscape for AI, and that the role of private governance is not yet adequately addressed in the current debate on AI ethics and AI governance. Business transactions of any type are layered and mediated by both hybrid and private accountability mechanisms that self-regulate the behaviour of companies, align business practices with widely held ethical principles, and enable societal trust. Assurance providers (accountant firms, auditors, verifiers, certifiers etc..) are guarantors that these systems and technologies work as intended. These types of governance systems are commonly referred to as private or hybrid governance, and they are very common ways to regulate many industries. We focus on the role of assurance mechanisms such as standards, recommended practices and certification, which are well-established mechanisms for providing trust in technologies. These create layers of trust (e.g., environmental standards and assurance methods to assess industry compliance with such standards). They also ensure products (e.g., a ship, a food product, or a management system) are safe, do not cause unacceptable harm to society and the environment, and respect established ethical principles such as human rights. At the same time, they generate trust and enable the well-functioning of business-to-business partnerships and transactions while providing trust to society.

Understanding how private governance works to reduce risk, cope with uncertainties and enhance trust is an important component of the debate on AI and ethics. In addition, understanding and addressing the disruptions created by AI applications to these governance mechanisms and their capacity to help coping with new ethical and societal challenges is of utmost importance. This is critical not only because AI *applications are seemingly driven by efficiency gains and cost savings rather than safety or the common good, or because public governance and social norms lag behind AI research. It is also crucial because AI systems change well-established risk profiles of industries and create new ethical challenges. This may change the nature and the boundaries of assurance while at the same time making the role of assurance as a provider of trust even more important. We argue that enabling trust in AI is a matter of generating appropriate processes that facilitate deployment of AI applications in a safe and responsible manner.*

1. Introduction

When the information age's most enthusiastic apostles celebrate the breakdown of hierarchy and authority, they neglect one critical factor: trust and the shared ethical norms that underlie it *(Francis Fukuyama 1996)*

We may not have evidence for a crisis of trust; but we have massive evidence for a culture of suspicion *(Onora O'Neill 2002)*

We need to cultivate in ourselves, collectively, a special kind of moral virtue, one that expresses what I will call the technomoral virtues *(Shannon Vallor 2016)*

The exponential advance of AI technologies and their application across various social arenas, from driving to medical diagnosis and finance, is reshaping our personal and professional lives, yet it is seen by many with suspicion. The transformational potential of these digital technologies is immense; they may in fact be the determining factor for the solution of many existing global challenges, such as food and water security, sustainable value chains, less road accidents, or the decarbonization of the global economy. Yet, the power of AI can also be misused for less worthy goals, such as war, discrimination, the perpetuation of inequality or to enhance nationalism and unfair competition. There are many historical cases showing that technological advances have not been matched by societal acceptance, and that governance mechanisms lag behind the rapidly increasing digital capabilities. Moreover, there are fears that AI technologies may acquire consciousness and become our ultimate invention, perhaps substituting the primacy of human beings, although that is very far from current technological capabilities. Perhaps more importantly, there is a well-founded mistrust that a hurried push to develop and deploy AI technologies is driven by efficiency and cost savings by a handful of companies, bypassing concerns for safety, ethics or the common good. Considering that AI technologies are still in their infancy, yet deployed at a rapid pace, a key question is: What is the best way forward to ensure that AI technologies serve human goals, abide by widely shared ethical principles, and consider impacts on society? The discussion on trust in AI is a discussion about generating appropriate and meaningful governance and accountability processes, to ensure that we enable the benefits of AI technologies while preventing misuse and harmful applications.

Digital technology industry giants, consumer and interest groups, governments and academics all agree on the urgent need to create suitable governance mechanisms built on shared values for the further maturing and deployment of AI technologies

(Dafoe 2018; Microsoft 2018; AI Now 2018, EU AI HLEG 2019). Some argue for the need to imbue ethical reasoning into the construction of algorithms and to develop computational ethics as a new engineering discipline (Kaplan 2016). Others call for the actual auditing of algorithm performance (O'Neill 2017, 2018). Some authors tackle the issues from the perspective that it is us, human beings, that need to mature our moral computation. Shannon Vallor (2016) wishes a future ruled by human beings with mature technomoral virtues to protect the good life in an age of disruptive technological advance. Some ongoing discussions follow this perspective arguing for the use of AI to augment our capabilities but morality as an arena for human beings (Spohrer 2019).

Rachel Botsman (2017) provides an insightful analysis on the contradiction that, while digital technologies enable trust among strangers, they may also have negative consequences, such as the breakdown of societal cohesion and the further individualization of people. Clearly, there is a central role for different forms of governance to regulate the further development and deployment of AI technologies, even though one of the key characteristics of the digital age is an increase in transparency and, in many instances the removal of intermediaries in many socio-economic transactions.

This paper contributes to the literature on ethics, AI and governance, by unpacking the role of private governance could play in regulating the process of building and deploying safe and responsible AI. We argue that traditional private governance mechanisms are a key component of the needed regulatory and accountability landscape for AI, and that the role of private governance is not yet adequately addressed in the current debate on AI ethics and AI governance. Business transactions of any type are layered and mediated by both hybrid and private accountability mechanisms that self-regulate the behaviour of companies, align business practices with widely held ethical principles, and enable societal trust. Assurance providers (accountant firms, auditors, verifiers, certifiers etc..) are guarantors that these systems and technologies work as intended. These types of governance systems are commonly referred to as private or hybrid governance, and they are very common ways to regulate many industries. We focus on the role of assurance mechanisms such as standards, recommended practices and certification, which are well-established mechanisms for providing trust in technologies. These create layers of trust (e.g., environmental standards and assurance methods to assess industry compliance with such standards). They also ensure products (e.g., a ship, a food product, or a management system) are safe, do not cause unacceptable harm to society and the environment, and respect established ethical principles such as human rights. At the same time, they generate trust and enable the well-functioning of business-to-business partnerships and transactions while providing trust to society.

Understanding how private governance works to reduce risk, cope with uncertainties and enhance trust is an important component of the debate on AI and ethics. In addition, understanding and addressing the disruptions created by AI applications to these governance mechanisms and their capacity to help coping with new ethical and societal challenges is of utmost importance. This is critical not only because AI applications are seemingly driven by efficiency gains and cost savings rather than safety or the common good, or because public governance and social norms lag behind AI research. It is also crucial because AI systems change well-established risk profiles of industries and create new ethical challenges. This may change the nature and the boundaries of assurance while at the same time making the role of assurance as a provider of trust even more important. We argue that enabling trust in AI is a matter of generating appropriate processes that facilitate deployment of AI applications in a safe and responsible manner.

2. On trust

Trust underlies all social and economic relations. Trust is the firm belief in the reliability, truth, or ability of someone or something[1]1. We participate in public life, drive our cars, enter into contracts and engage in economic transactions, relate to government organisations or with customers because there is trust underlying all these actions. Behind this trust, there is a set of shared ethical norms, tacit or explicit principles as to what is, and what is not, appropriate behaviour by all parties involved, how things should work, or how things should be. Trust is a fundamental characteristic underlying all human relations. Francis Fukuyama (1995) studied in depth the role of trust in economic life and unpacked how different industrial structures and societies are glued together through diverse forms of trust mechanisms, some cultural, some institutional, others simply the result of centuries-long social practices. Even in an age of a truly globalized economy, there exist different trust cultures, each evolving and generating different types of mechanisms to ensure there is a minimum amount of social trust involved in any kind of transaction. Thus, trust is one of the most pervasive characteristics of social life. Whether in the form of personal relations embedded in social institutions like the family in many Mediterranean and Asian cultures, or in the form of formal governance structures in more rule-oriented cultures, trust enables the fluidity of social and economic transactions and creates the basis for prosperity and market transactions (Fukuyama 1995). It is so pervasive, that “a complete absence of trust would prevent [one] even getting up in the morning (Luhmann 1979)”.

Trust facilitates interactions among people but also interactions between people and technologies, including relations “among the members of a system, whether these be human agents, artificial agents or a combination of both (a hybrid system) (Taddeo 2017:565). This ability to facilitate interactions is what makes trust a valuable asset. AI adds new dimensions to the interactions between humans and technology because they perform intellectual tasks and, in many applications, have decision making power. As we delegate to digital technologies cognitive tasks that were earlier performed by humans, AI technologies become more than a mere facilitator, given our decisions become dependent on the technology. This may require new forms of trust. Trust and the dependence of users becomes entangled. Taddeo (2010) argues for three dimensions in the interface between digital technologies and trust: the occurrence of trust in digital environments, the nature of trust in technology, and the relation between trust, technology and design (Taddeo 2010: 284). The implication of this analysis is that we are facing a new kind of technological mediation altogether, one that raises substantive societal and ethical issues.

[1] Definition given in Oxford dictionary https://www.oxforddictionaries.com/

Another author, Botsman (2017), explores the many ways in which digital technologies affect us and re- organize us and our work. We have created trust-building mechanisms that enable us to share our homes with strangers through platforms such as Airbnb, and we trust our personal lives and opinions to be showcased in public domains in ways that would have been unthinkable just a few years back. Yet, this same system "makes it easy for almost a third of the world's population to gossip and gripe, share and like, even if the content is false, and without proper checks and balances or any real redress (Botsman 2017: 106)." Some may say these concerns with trust apply only for issues related to digitally mediated interactions in the public arena or relations between companies and customers. However, the absence of proper checks and balances and institutional mechanisms enabling trust prevents the formation of judgments of trustworthiness for any kind of interaction. In short, ubiquitous digital systems and the speed of their implementation leads very often to a *fragile trust*, trust that is not underpinned by shared values and is not supported by an established regulatory regime; it is a trust simply based on the dependency we have on those technologies, nothing more. In other instances, the speed of technology adoption simply leads to *blind trust*, that is, we accept their use in our daily life and deployment in industries without the necessary assurance that the technologies are fit for purpose, safe and reliable. It is a misplaced trust, often lacking sufficient knowledge and sufficient reflection on the possible consequences of technology on society. In short, trust is a complex issue, including but not limited to trust that a particular technology is sound. Like the dwindling trust in science, trust in technology is not only about rational and robust quantitative, mathematical or technical representations of the world. Trust in technology encompasses many other dimensions, including political, ideological, ethical, social and psychological (refs).

2.1 Specific trust issues raised by AI

There are many different perspectives as to what the key pillars for trusted AI are, but there seems to be wide agreement on the need for certain key technical specifications. The head of Ai at the World Economic Forum states the following as the key issues to demonstrate the trustworthiness of AI (Firth- Butterfield 2018):

- Bias
- Transparency
- Accountability, and
- Privacy

IBM researchers define the pillars to trusted AI as meeting the following criteria (Hind et al 2018):

- Fairness
- Robustness (safety and reliability)
- Explainability, and
- Lineage

In a similar manner, Microsoft articulates the characteristics of trustworthy AI as meeting the following characteristics (Microsoft 2018):

- Fairness
- Reliability and safety
- Privacy and security, and
- Inclusiveness

All these criteria are fundamentally presented as technical issues, that is, issues that can be solved in the process of design by AI experts. The expected end result is an AI application that is transparent, able to be understood by humans, and able to explain how it has made a decision or prediction. This is the goal of what is often referred to as *explainable AI*. However, meeting the criteria outlined above requires acknowledgment of another set of trust issues which are often underplayed. We will need to have trust in the data, trust in the models that generate predictions, and trust not only on algorithmic agency but on the contexts in which this algorithmic agency operates. In addition, we need trust in domain knowledge required by a particular AI application, as well as a clear understanding of the impact that the AI in question may have on the autonomy of human users.

Trust in data	Trust in models	Understanding the context as well as algorithmic agency	Roles of AI & Levels of autonomy
The origin of many AI applications is data. Thus, trusting data and ensuring its quality is a point of departure for any attempts to regulate AI. The EU AI HLEG considers that trustworthy data should address bias, fairness, transparency and discrimination (EU HLEG 2019). But one also needs to trust that the data used is relevant to the problem to be solved.	Models are a set of structured assumptions (Ghahramani 2015). We tend to believe human assumptions as correct, while AI techniques generate their own assumptions, and this leads to a lack of trust given that the AI may use unknown or unfamiliar forms of rationality. But one cannot assess the assumptions contained in a model without unpacking the assumptions made by the humans that created the model (Miller 2018).	Many ethical and societal questions are raised by the agency aspect of AI (Taddeo and Floridi 2018). Given the agency of an algorithm still interacts both with physical systems and with humans, it is a pre- requisite to understand cyber-human interactions. In this regard, it is important to know which types of human expertise participate in these complex interactions. in their entirety.	AI agents can act as independent agents and play different roles in decision processes. They can act as consultants, decision aids, managers, or delegates. Each role can have different levels of autonomy, with different levels of responsibility ascribed to the AI and to humans. Also, AI affects the autonomy of the human user

To build explainable AI, these four issues are [important/necessary/critical/essential] to address. A perspective that looks only at the technologies themselves (no matter how explainable those technologies are) is insufficient. Two of the four issues listed above are critically important. First, recognition that the context or wider system in which the AI is deployed matters as much as the AI itself. This wider system may include other technologies, or become a cyber-physical system, introducing further complexity to the search for AI explainability. Second, with ever higher levels of autonomy, it is difficult to abstract the AI from its interactions with humans. Not only are humans (so far) those who design AIs, and thus bring their own cognitive biases into the process; humans also play a role in defining the context in which AIs are deployed and how they are used, which may reflect the human biases and understanding of the context and algorithms rather than the capabilities of the AI.

How much responsibility we delegate and how much freedom or autonomy we grant to an employee, a child, a pet, an institution, or a contractor depends on our level of trust. The same goes for AI. For instance, a *Consultant AI* gives advice to humans, but it does not independently make decisions, as the humans involved in the process also make independent judgements. An example would be medical expert systems. Consultant AI is typically used in cases where bad predictions may have serious consequences, and the trust in the system is not deemed sufficiently high to let it make decisions autonomously (Xiang 2002). *Assistant AIs* are used to help humans to collect and analyse relevant information to reach decisions, e.g. by analysing large amounts of data that humans would not be able to process themselves and propose recommended actions. In such cases, the analysis of the AI agent is trusted to a large degree, but humans take responsibility for concluding on appropriate actions. An *AI Delegate* is trusted to act on behalf of humans, e.g. an AI agent controlling an autonomous robot or vehicle. This, however, does not mean human agency is completely removed when we include the context in which an AI is deployed.

This means trust in AI is not a binary yes/no question, but rather a continuous scale of capability levels, and types of interactions. Each type of interaction or role has associated levels of trust. The Society of Automotive Engineers (SAE 2018) applies this principle to define a scale of six autonomy levels for cars: 0) No automation; 1) Driver assistance; 2) Partial automation; 3) Conditional automation; 4) High automation, and; 5) Full automation. Such taxonomies not only facilitate a common understanding among actors creating autonomous systems, but can also form the basis of governance frameworks by stipulating the level of trust required from a system depending on how it is operated or used.

A large amount of literature has tackled the issue of trust in AI by attempting to imbue ethics into code. *Computational ethics* seeks the integration of ethical thinking into the coding and design of machines. But the ethical challenges raised by AI are more numerous and complex. While it is important to ensure that biases are not automated, or that an AI application does not violate people’s right to privacy, it is a simply inappropriate to claim we can design moral machines or that this should be our goal in order to generate trust in AI. Firstly, there is not one single global and general moral code in this world. Secondly, many people would argue that morality is a trait of human beings, a trait so complex that it is impossible to capture by binary code (Hao 2018). A lot of the experiments we see emerging in embedding ethics into machines are no more than the integration of the values and worldviews of a particular group of people into the design of AI applications (Hao 2018). This is a misrepresentation and simplification of ethics. We tend to make moral calculations after an event has occurred. We ask ourselves, was this the right thing to do? Moral reasoning is driven by moral intuitions and interpretations of the facts, rather than by applying a particular ethical code or metaethical system. Our ethical intuitions are the result of embedded ethics in the social, cultural and normative fabric where we are situated. But we do not hold values in theory; we hold values as people and express them with our actions. Ethical examples, such as the trolley problem serve to illustrate different ways to reason ethically, but the examples are not the source of ethical behaviour just a mere meta-analysis illustrating how one could provide an ethical rationale to action after the action has been taken. This is one of the fundamental reasons why the morality of a technology is not a characteristic of the technology, but an attribute of the humans that create, interact, and are impacted by that technology, as argued by Shannon Vallor (Vallor 2017). If we want a world of moral machines, we need a world where human virtues regulate the design and deployment of these technologies.

In summary, trust in AI depends on more than explainability. We must analyse and understand the context and goals of the application, the role of algorithmic agents, and the roles of human beings, with their own assumptions of how the world works. At the core lies the need to have a very clear view on where humans are involved in the overall processes of designing and deploying a particular AI solution, and what level of responsibility is delegated to an application.

An additional critical aspect of trustworthy AI is the need to know the expectations of those who are supposed to trust the AI application. So far we are still quite in the dark, with little research offering empirical evidence as to what those expectations are. An exception is a well-documented empirical analysis of public perception of AI conducted recently in the UK. The results showed a high level of mistrust in AI. A key source of mistrust in AI, the authors claim, are dominant narratives the both over represent the powers of AI technologies or that only highlight their potential negative consequences. “Both excessively hopeful and excessively frightening narratives can have significant negative societal impacts. Exaggerated expectations for what AI can achieve, and when, risk undermining further research and investment” (Cave, Coughlan & Dihal 2018).

These “excessively hopeful and excessively frightening” narratives are exacerbated by a lack of AI governance. To enable appropriate public debates on AI capabilities and raise awareness of their risks and benefits we need an infrastructure of governance mechanisms that debate and regulate the design, deployment, and operation of AI technologies. After outlining the relations between trust and governance we unpack the concept of governance, argue for a specific role of hybrid and private governance mechanisms in enabling the trustworthiness of AI development and deployment, and elaborate on what type of private governance AI requires.

2.2 Trust and governance

Trust is intrinsically linked to governance. Trust is never a matter of blind deference, but rather a question of placing - or refusing - trust with good judgment (O’Neill 2002). Thus, the critical question about trust relates to the institutions and processes that enable us to actually make judgements about the trustworthiness of people, processes, services, products, or organisations. And as O’Neill argues, we always need social and political institutions that allow us to judge where to put trust (O’Neill 2002). These are the institutions of governance, institutions that execute accountability in the public and in the private arenas. This means there is an entanglement between the need for assessing trustworthiness and the power delegated to social and political institutions that enable trust. Delegating power to those institutions is a necessary condition for generating structured ways to enable good judgment. These are institutions that form the architecture of a social contract, enabling societies to function, enabling the deployment of technologies, and acting as sources of guidelines and accountability. It is important, however, not to confuse the term ‘governance’ with ‘government’. Governance refers to a certain degree of self-regulation by societal actors and to private-public cooperation in solving societal problems (Biemann and Pattberg 2008).

Based on a review of various broad definitions of governance, Andonova et al (2009) identify three common features of governance: first, governance is concerned with realising public goals; second, it steers a particular constituency of actors; and third, it is regarded as authoritative. The latter element – authoritativeness – is of critical importance as it distinguishes governance from other types of informal cooperation between actors. This cooperation factor is important for the establishment of credibility and authoritativeness. As Falkner (2003) argues in the context of international relations, governance emerges as the result of interaction and negotiation, a type of negotiation that is de facto institutionalized and of a more permanent nature. Within a governance system, participants do not decide to be bound by institutional norms because of mere self-interest, but rather adjust their behaviour in negotiation with other participants’ views, all in recognition of the legitimacy of the system of governance (Falkner 2003).

Negotiation and cooperation are important for the eventual construction of an AI governance system. Negotiation and cooperation will help legitimacy to emerge through the complex set of processes that build guidelines and standards (as well as ways to verify those guidelines and standards). Those processes will also build an understanding of what technologies are, what they can and can’t do. Building governance mechanisms for new technologies is often, in fact, a performative process, as it must be built on consultation and negotiation. Those consultations and negotiations help build societal awareness about specific technologies, and - in so doing – they contribute to determining the perception of such technologies.

Governance is also an iterative, ongoing process of negotiating perspectives and priorities. As new governance systems are built, diverse stakeholders in society must consult and strive to reach consensus. These negotiations and perception-building are important reminders that governing is an ongoing process; it is not necessary or desirable to wait until a perfect system is in place to start governing a field. This is most certainly the case for AI, which is in dire need of governance mechanisms. Lastly, governance is often multi-layered. It is the aggregate of a set of interacting mechanisms that enable the processes of cooperation and negotiations that are described in the preceding paragraphs. A governance system is never established by a single organisation, but rather is often the results of partnerships of many organisations, both public and private, and very often a hybrid of both.

3. Governing AI

Discussion of AI governance is exploding, driven by an increasing global debate on the ethical and societal dilemmas posed by advances in AI techniques and their deployment in society. Another driver of the AI governance discussion is the increased recognition of the strategic value of AI as a driver of national and regional power and as a driver of economic growth. Five distinct emerging perspectives on AI governance are explored in the following sub-sections:

1. Technical and political aspects
2. Sector-specific governance
3. Dynamic and iterative governance
4. Ethics and technical design
5. Machine-human interactions and regulatory forces

3.1 AI is both technical and political

An in-depth analysis of the research agenda needed for devising global norms, policies, and institutions to best ensure the beneficial development and use of advanced AI was produced in 2018 by the Future of Life Institute (Dafoe 2018). The document, although a work in progress, structures research on AI governance along the following dimensions.

Technical:

- Mapping technical possibilities
- Assessing progress
- AI safety

Political:

- Domestic and mass politics
- International political economy
- International security

Ideal governance:

- Values and principles
- Institutions and mechanisms
- Positive visions

The proposed 'Ideal Governance' is based on values and principles, institutions and mechanisms, and on positive visions. The rationale behind giving an equal weight to politics and technical issues alongside a positive vision of AI impacts lies in the transformational capabilities AI already presents, and the fact that a race for AI capabilities could easily become a race for power.

The document from the Future of Life Institute also gives special attention to the non-linear nature of AI's transformational capabilities. Although one can anticipate innovations in the field, we must be prepared for big jumps in capabilities that may render traditional governance mechanisms suddenly obsolete. The authors solve this problem of anticipating the actual technical capabilities of AI by providing an ideal vision of the 'end goal' of AI for humanity within the context of shared values and democratic principles.

But in the real world of fast development and deployment of AI solutions, driven by efficiency, cost savings and potential economic growth, it is crucially important to regulate AI immediately. This regulation should be part of an ongoing process of building suitable governance mechanisms and the shared values underpinning an ideal governance vision. The tripartite structure offered by Dafoe (2018) is both technical and political. It could be used to create an initial system of rules and regulations using existing private governance mechanisms. Before we address this point again we want to outline other AI governance perspectives.

3.2 AI governance must be sector-specific

A different take on AI governance is presented by the AI Now Institute latest report (AI Now Institute 2018). In a summary document offering ten key recommendations to ensure that AI serves humanity, governance is number one. According to this report, the key actors in AI governance are governments. Moreover, they argue for the "need to regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain (AI Now Institute 2018:4)." The logic behind this sectoral focus is that it is difficult to establish the risks or benefits of a particular AI application independently of the context of its deployment. Devising the risks requires a systemic perspective. Since we must pay attention to the specific context in which an AI is deployed, domain knowledge of this context is critical to the evaluation of AI performance. Over the years, individual sectors have created relevant regulatory frameworks and accountability mechanisms that are also relevant for AI governance. "Sectors like health, education, criminal justice, and welfare all have their own histories, regulatory frameworks, and hazards. However, a national AI safety body or general AI standards and certification models will struggle to meet the sectoral expertise requirements needed for nuanced regulation (AI Now Institute 2018: 4)".

The report argues that the application of AI in specific domains and its consequences must be prioritized, rather than the regulation of the technologies themselves. This focus on leveraging existing regulatory frameworks of specific domains is a key message. It implies that existing private governance mechanisms of different sectors can be leveraged for helping regulate the development and implementation of AI.

3.3 AI governance must be dynamic and iterative

Private sector consultancies are also jumping on the bandwagon of AI governance. Deloitte published a research paper focusing on the limitations of traditional governance systems in regulation of emerging technologies, with a special focus on "intelligent technologies" (Egger and Turley 2018). Intelligent technologies disrupt existing business models, posing serious challenges to traditional governance making. The paper puts forward the concept of seeing governance as a space, a regulatory spectrum, encompassing different stages: from a pre-regulatory stage to a stage of testing and evaluation, leading to a more mature regulatory approach. It also considers an iterative process enabling ongoing improvement and more mature regulatory approaches. This stepwise conception of regulation is useful as it reminds us of the need to create the rules, to create a system to enforce them, and to create an iterative process that enables ongoing revisions.

The authors also argue that we need to think about different types of regulation. These types are:

1. Adaptive regulation (enabling iterative regulatory processes)
2. Regulatory sandboxes (enabling prototyping and the creation of regulatory accelerators)

3. Outcome-based regulation (focusing on results and performance)
4. Risk-weighted regulation (moving away from one-size-fits-all to a data-driven approach)
5. Collaborative regulation (enabling international collaboration across countries and ecosystems of actors)

These five regulatory forms are complementary. Each contributes to the debate on how to create the necessary conditions for good judgments, and thus for trust in AI and its specific applications. All these characteristics also apply to hybrid and private governance.

3.4 AI governance must go beyond integrating ethics into technical design

The European Union's work to establish governance of AI provide a fourth example of how this task can be approached. In their efforts, the EU has mobilized the expertise of many scholars and societal actors. Of particular importance is the recent guidelines produced by the High-Level Expert Group on Artificial Intelligence, entitled "Ethics Guidelines for Trustworthy AI." These guidelines, although voluntary, put forward a structure that directs attention to ensuring the ethical purposes of AI development and deployment, founded on fundamental rights, societal values, and the ethical principles of beneficence (do good), non-maleficence (do not harm), autonomy of humans, justice, and explicability. They also call for exercising vigilance in areas of critical concern. The guidelines continue with guidance for realizing trustworthy AI. The Expert Group converges around the idea that trustworthy AI emerges from the integration of ethical considerations expressed as human rights, technical robustness and the leveraging of relevant governance mechanisms to ensure compliance and operationalisation.

The requirements for trustworthy AI according to the draft guideline are:

Accountability
- Data governance

Design for all
- Governance of AI autonomy (human oversight)

Non-discrimination
- Respect for human autonomy

Robustness
- Respect for privacy

Safety
- Transparency

These requirements are to be incorporated at the earliest design stage and it is important that teams building and deploying AI are interdisciplinary. They encourage the use of technical and non-technical methods to ensure the implementation of the requirements into all AI systems. Ensuring auditability and the integration of trustworthy AI into an organisation's culture are also highlighted as key elements. Although the draft Guideline also indicates a key role for private governance, this is not clearly developed.

3.5 AI governance must encompass machine-human interactions and all regulatory forces

Finally, we present the recent work on AI governance emerging from the Massachusetts Institute of Technology (MIT). Much of this work emphasizes a view of AI that includes interaction and role division with human beings. As such, AI governance cannot be about technical requirements only, but about the superminds that emerge in the combination of artificial and human intelligence (Malone 2018). Iyad Rahwan, professor of AI ethics at MIT, follows a similar thinking process. In his efforts to study AI from an interdisciplinary and business perspective, Rahwan (2018) argues for the need to think about the regulation of AI encompassing all the forces that regulate society, leading to what he calls the inclusion of society-in-the-loop or SITL. There are four key forces in society, Rahwan argues: Government regulations, norms, markets and architectures. AI governance needs to include elements of each.

To implement society-in-the-loop, Rahwan argues, it is crucial to understand what types of behaviours people expect from AI and to enable policy-makers to articulate these expectations (Rahwan 2018: 9). We also need metrics to evaluate AI behaviour against widely shared values for algorithms that are already used in many arenas in public and private life such as in banking education or finance. This means that AI requires not only governmental regulation, but also industry standards that represent the expectations of the public and the corresponding oversight, that is, assurance, verification or certification.

In short, all these perspectives on AI governance point to the need to look well beyond governmental regulation and policy instruments, and towards other forms of governance. In addition, they point to the integration of ethics with technical robustness, and close attention to the interactions between machines and humans. Although they all hint to a key role for industry standards, certification and verification, the types, roles and functions of hybrid and private governance are not yet appropriately developed. Although the five cases we have summarized are a small sample, they show a gap in the current literature on AI governance. That gap is a more in-depth analysis of the role of hybrid and private governance as a provider of trust and as an element of the governance architecture of AI. This is important because an in-depth exploration on how hybrid and private governance can contribute to regulating AI may promote open debate and consensus on expectations. Hybrid and private governance can provide tools, methods, existing accountability mechanisms, and institutions that have experience in setting and auditing standards. Hybrid and private governance can also enrich the debate and effectiveness of AI governance with lessons learned over long periods of time, collaboration and networking, sectoral approaches, and the fusion of technical and societal considerations in relation to AI technologies.

4. Hybrid and private governance

‘Hybrid and private governance’ refers to the governance functions taken up by partnerships between public institutions and industry or other private actors. These private actors are non-state actors that are active in the market. They both produce and abide by private governance, creating order in economic and social life. Selfgovernance or self-regulation is common in many industries as a complement to public governance mechanisms. From the first stock markets in Amsterdam dating to the 1600s (Stringham 2015), to the first classification societies issuing certificates to ship owners certifying that ships were technically fit for purpose (Paulsen et al 2014), private governance has enabled trust in technological advances through setting and enforcing rules and standards. Private governance has spread to all industries and all areas of socio-economic transactions. In fact, we live in what Michael Power (1999) calls the audit society. We check and double-check that things and technologies are fit for purpose and meet societal requirements. Contemporary societies have developed a myriad of methods and actors to verify and certify this is indeed the case.

In an in-depth study of the role of private regulation in society, Tim Büthe describes the processes of setting private regulation as follows:

These rules are set by a range of non-governmental bodies: industry associations, NGOs, networks of firms, technical experts, or groups of activists. Many of their rules are widely observed by producers of goods and services, and other economic actors, including many who did not participate in writing the rules. Private rules thus govern—that is, they enable and constrain—a broad range of activities in the world economy. Some of the private regulators, such as technical standards-developing organizations, operate largely out of the public view. Others, such as credit rating agencies, have recently become the object of intense scrutiny (Büthe 2010).

In a follow-up study, Büthe and Mattli explore the critical role that industry standards and those who create them play in regulating industries and their activities (Büthe and Mattli 2011). The resulting system of global rule-making by technical experts has a double-sided effect. On the one hand, it provides the structure for generating and verifying that risks (in particular technology risks) are unveiled and addressed. On the other hand, it creates ecosystems of very powerful actors that have global influence.

In light of the current lack of accountability in the development and implementation of AI, we argue that a transparently built system of private and hybrid governance may enable people and institutions to make appropriate judgments regarding the safe and responsible use of AI. Trust in AI will emerge when we have managed to knit together that 'fabric' of private and hybrid governance. We also argue that assurance mechanisms are a way to improve the existing accountability processes that are common to industries and the functioning of markets. Assurance mechanisms can help to integrate the new ethical dilemmas posed by the application of AI technologies in the real world into those accountability processes.

4.1 Standards: A key private governance mechanism

A key mechanism of private and hybrid governance is the development and application of industry standards. Industry standards support many of the complex aspects of the maturing and deployment of new technologies. They steer many issues, from internal company work processes, to the design and performance of the products and services they deliver, to the labour conditions of employees, to the transport systems by which they commute. Industry standards are more than just technical documents; they steer behaviour, embed norms into organizations, and provide "recipes for reality" (Busch 2011). They have the potential to transform both the way we think about different issues and the ways in which we act (Timmermans and Berg 2003).

The process of creating standards requires negotiation of the proper balance between technical capabilities and societal expectations. This process creates awareness of the technologies, and it gives those technologies a ‘social licence’ to operate along the way. Standardisation is the process of constructing uniformities across time and space, through the generation of agreed-upon rules. Standards promote coherence, and they facilitate interaction and the sharing of best practices (Timmermans and Epstein 2010). They also serve as one means for knowledge transfer, and act as translation mechanisms between technology providers and society. Furthermore, “standards are recipes not only for the best ways to ensure that a product or process is fit for purpose; they also shape the demand for certain products or services. Many of our everyday decisions are shaped by standards; we accept and adapt to them without even realizing that we do (St.Clair and Aalbu 2017).” Standards are an invisible infrastructure, both in technical and moral terms.

There are many different ways to classify standards. In general, we can distinguish between the following three types.

Design standards	Performance standards	Procedural standards
Set technical specifications of a physical system (For example, the ship classification rules used by classification societies)	Set outcome specifications (For example, energy efficiency standards)	Set specifications for processes (For example, management systems standards such as ISO 9001 for quality management or ISO 14001 for environmental management.)

Most standards are a combination of these three types. They harmonize processes, procedures and designs, they enhance transparency and traceability, they inform and accelerate regulatory processes, and they provide the basis for third-party verification. Standardisation is thus the result of long-term processes, often initiated by industry, and often in close collaboration with governments. Standards also set the terms for responsible deployment of technical solutions into the real world by promoting safety and accountability while often also enhancing innovation and levelling the playing field. Standardisation processes are by no means always fair or transparent. Although they are often regulated by standards organisations such as the International Standards Organisations (ISO) or by similar national bodies, standardisation processes can also represent the alliances of big players in a particular industry. Thus, standards may not take into account important societal issues that are relevant for smaller players or less influential regions of the world.

Standards are an essential ‘brick’ in the construction of a fair and representative governance system for AI technologies and their deployment into society. AI governance needs to start by examining existing standards, identifying which ones are relevant for regulating AI techniques and their deployment, and exploring ways to revise them to accommodate AI aspects. A next step will be to create new standards to fill gaps and shortcomings in existing (or modified) standards, and to prepare for future developments of AI. At the same time, it is fundamental to create the basis for the other key element of private and hybrid governance: third-party verification and certification.

4.2 Providing trust through assurance

We often refer to third-party verification and certification as ‘assurance’. Assurance refers to the validation and provision of confidence that a product or process is fit for purpose, and that it complies with existing safety, environmental, or other technical requirements. The provision of assurance is always based on credible technical information or knowledge. Assurance methods and tools set the boundaries of acceptable (and, often, insurable) risks. At the core of assurance are standards, but other types of governmental regulations and societal expectations can be embedded as well. The key tools of private governance are combinations of industry rules, standards, certification, and verification systems, including independent third-party verification of compliance with accepted rules and regulations.

The provision of trust through assurance is more than a century old. In the maritime industry, assurance has played a critical role since the mid-1880s. Ship-owners, some of the first true globalizers, were eager to take advantage of the liberalization of world trade, but their operations were very limited by providers of insurance (Paulsen et al 2014). The reluctance of insurance companies to accept liability for accidents was at the time very much related to a lack of control of the technical specifications and built quality of the ships. This reluctance by the insurance sector led ship owners to form alliances that jointly determined the conditions that would enable insurers to trust their ships and insure their cargo. This led to a long and iterative process of developing new standards, understanding uncertainties, and building a notion of acceptable and insurable risk that would satisfy insurers and governments while also meeting societal expectations. These processes of creating acceptable and insurable risks also led to the creation of a set of shared values across all societal actors that gave the ship owners a societal licence to operate. Soon, classification societies emerged to provide third-party assurance to the insurance companies and to governments and society. The primary source of legitimacy of these classification societies was the technical and scientific knowledge that enabled the actual assurance that technical designs were in fact fit for purpose. As technological advances led to increasingly sophisticated ships, these organisations added corresponding technical skills to their portfolios, advancing and often creating new technical and procedural requirements for safety in maritime operations. This concentration of expertise was also key in the increasing role of these classification societies. As governments found that they did not have staff sufficiently qualified to regulate technical requirements for ships in operation, they increasingly delegated government authority to these classification societies. This delegation reinforced the classification societies’ need to understand not only the technical requirements of a ship, but also the requirements for ships in operation, the job descriptions of the crew, and the interactions between ships and their flag states. This delegation of authority was underpinned by strict ethical rules guaranteeing neutrality and due process.

Providing third-party assurance was not possible without ensuring neutrality. Classification societies needed to demonstrate that they were neutral partners, representing neither the interests of a particular ship owner, nor of a particular government, nor of a particular vendor providing parts for the ship's construction. Both technical expertise and neutrality led to the forging of long-standing reputations, and many of the 19th century classification societies remain in operation, providing third-party assurance in a competitive market with one another.

The maturing of maritime standards and the provision of assurance against those standards was an iterative process, driven by the need to establish accountability and responsibility, often economic responsibility, for the resulting accidents and losses due to poor technical design. In the decades after these beginnings of the maritime assurance regime, we have seen the expansion of standardisation and third-party verification to all areas of life. Not only are we surrounded by all types of standards regulating food safety, education, health care, labour relations, financial reporting, or environmental impacts, we also have a myriad of organisations and systems that audit and verify that actors comply with these regulations. As mentioned earlier we live in an "audit society", but this is due to the need for accountability in an increasingly technically and scientifically complex world (Power 1999). The more powerful technologies become, the more societies seem to have a need for accountability and control.

With AI, science and technology are about to get smarter and more complex, and they will create more uncertainties. The lessons from private and hybrid governance regimes in other sectors show that we need similar processes to materialize the principles of AI governance that are currently emerging and to anchor them in the systems and in the language used by industry. It is perhaps even necessary to complement governmental regulation and leverage other forms of governance besides policy instruments. We must consider the integration of ethics with technical robustness, and we must take into account the relationships between machines and humans.

5. Concluding remarks

Governing AI is going to require the coordinated effort of multiple societal actors. Emerging perspectives on AI governance argue that AI governance must be both technical and political, sector specific, dynamic and iterative, and go beyond a mere integration of ethics into coding. Perspectives also point out to the need to look well beyond governmental regulation and policy instruments, and towards other forms of governance. In addition, there seems to be agreement that a special issue that requires close attention is the interactions between machines and humans through the whole life cycle of AI systems. Although they all hint to a key role for industry standards, certification and verification, the types, roles and functions of hybrid and private governance are not yet appropriately developed. Public governance mechanisms will need to be complemented with the hybrid and private governance in order to ensure that AI development and deployment is fit for purpose, abides by a socially negotiated definition of acceptable risk, and is bound to standards and recommended practices that are recognised by the industries and the sectors in which solutions are implemented. Without romanticising the role of private governance and voluntary mechanisms of accountability, we argue that lessons learned from the emergence of assurance, certification and verification have central importance in the debate on safe and responsible AI. For AI technologies to be scaled up in a way that does not conflict with widely accepted ethical values, and for those technologies to enjoy societal support, companies producing and deploying these technologies need to demonstrate compliance with safety and security requirements comparable to other technologies. We have seen how emerging work from many different actors points to the need for new forms of governance, and the integration of ethical considerations and AI explainability into existing systems of accountability. We have also seen how governments call for alignment with widely accepted societal values and existing standards, as well as established industry mechanisms for accountability and compliance, although some such standards and industry mechanisms would need to be adapted (and new ones created) to tackle the decision-making power of intelligent algorithms. Assurance of AI is not only about third-party professional opinions regarding how transparent an AI system is. Rather, assurance of AI would require a systems perspective, encompassing both the digital and physical components of an application as well as the roles and tasks of human agents and the consequences of these applications to society. In short, hybrid and private governance are elements of a meaningful governance system that could prevent harmful applications of AI, ensure robustness, prevent bias and violations of privacy, and provide trust without stalling the growth potential of these technologies and their potential to support the common good.

6. References

AI Now Institute. 2018. AI Now Report 2018.
https://ainowinstitute.org/AI_Now_2018_Report.pdf

Andonova, L. B., Betsill, M. M., & Bulkeley, H. (2009). Transnational Climate Governance. Global Environmental Politics, 9(2), 52–73.

Botsman, R. (2017). Who can you trust?: How technology brought us together and why it might drive us apart. Public Affairs.

Büthe, T. (2010). Private regulation in the global economy: A (P)review. Business and Politics, 12(3). https://doi.org/10.2202/1469-3569.1328

Büthe, T. & Matti, W. (2011). The New Global Rulers: The Privatization of Regulation in the World Economy. Princeton University Press.

Cave, S., & Coughlan, K. (2017). “ Scary Robots ”: Examining Public Responses to AI.

Dafoe, A. (2018). AI Governance : A Research Agenda.
https://www.fhi.ox.ac.uk/wp- content/uploads/GovAIAgenda.pdf

Eggers, Willian and Turley, M. (2018). The future of regulation: Principles for regulating emerging technologies. Deloitte Insights. Retrieved from
https://www2.deloitte.com/insights/us/en/industry/public-sector/future-of-

regulation/regulating- emerging-technology.html

Etzioni, A., & Etzioni, O. (2017). Incorporating Ethics into Artificial Intelligence. Journal of Ethics, 21(4). https://doi.org/10.1007/s10892-017-9252-2

European Union High Level Panel on AI (EU HLEG). 2019. Ethic Guidelines for Trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Falkner, R. 2003. Private environmental governance and international relations: exploring the links. Global environmental politics, 3 (2). pp. 72-87. https://core.ac.uk/download/pdf/207876.pdf

Firth-Butterfield, K. 2018. Interview with Thomson Reuters, 11 September 2018. https://blogs.thomsonreuters.com/answerson/wef-socially-responsible-ai-development/

Fukuyama, F. (1996). Trust: The social virtues and the creation of prosperity. The Free Press.

Ghahramani, Z. (2015). Probabilistic machine intelligence and artificial intelligence. Nature, 452–459. https://www.repository.cam.ac.uk/bitstream/handle/1810/248538/Ghahramani%202015%20Nature.pdf

Hao, K. Should a self-driving car kill the baby or the grandma? Depends on where you're from. MIT Technology Review https://www.technologyreview.com/s/612341/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/

Hind et al 2018 https://arxiv.org/abs/1808.07261

Kaplan, J. 2016. Artificial Intelligence: What Everyone Needs to Know. Oxford UP.
Luhmann, N. (1982). Trust and Power. John Wiley & Sons, Inc.

Microsoft. (2018). The Future Computed: Artificial intelligence and its role in society.

Malone, T. 2018. Superminds: The Surprising Power of People and Computers Thinking Together. Little, Brown and Company.

Miller, T. (2018). Explanation in Artificial Intelligence : Insights from the Social Sciences. https://arxiv.org/pdf/1706.07269.pdf

O'Neill, C. (2018). Auditing Algorithms. ORCAA. Retrieved from http://www.oneilrisk.com/

O'Neill, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books.

O'Neill, O. (2002). A question of Trust: The BBC Reith Lectures 2002. Cambridge

University Press.

Paulsen, G; Andersen, H.; Collett, J.P.; Stensrud, I. T. (2014). Building Trust: The History of DNV 1864- 2014. Dinamo Forlag.

Power, M. 1999. The Audit Society: Rituals of Verification. Oxford University Press.

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Society of Automobile Engineers (SAE). 2018. Taxonomy and Definitions For Terms Related To Driving Automation Systems For On-Road Motor Vehicles. https://webstore.ansi.org/Standards/SAE/SAE30162018

Spohrer, J. 2019. Surprise: AI In 2019. Forbes https://www.forbes.com/sites/cognitiveworld/2019/02/25/surprise-ai-in-2019/#1aabdb2b248a

St.Clair, A.L. and Aalbu, K. 2016. Business action for climate resilient pathways: the role of standards, DNV GL Position Paper, available at https://www.dnvgl.com/technology-innovation/climate-resilient- pathways.

Taddeo, M. (2010). Trust in Technology: A Distinctive and a Problematic Relation. Knowledge, Technology & Policy, 23(3–4), 283–286. https://doi.org/10.1007/s12130-010-9113-9

Taddeo, M. (2017). Trusting Digital Technologies Correctly. Minds and Machines, 27(4), 565–568. https://doi.org/10.1007/s11023-017-9450-5

Taddeo, M., & Floridi, L. (2011). The case for e-trust. Ethics and Information Technology, 13(1), 1–3. https://doi.org/10.1007/s10676-010-9263-1

Timmermans, S. and Esptein, S. 2010. A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. Annu. Rev. Sociol. 2010. 36:69–89.

Timmermans, S. and Berg, M. (2003) The Practice of Medical Technology. Sociology of Health & Illness, 25, 97-114.

Vallor, S. (2016). Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press.

Xiang, Y. 2002. Probabilistic reasoning in multiagent systems: A graphical models approach. Cambridge University Press.

Advanced analysis of reliability and risk of equipment subjected to degradation and obsolescence

Sebastian Martorell, Isabel Martón, Ana I. Sanchez
Universitat Politècnica de Valencia
MEDASEGI research group
Camino de Vera, S/N
46022, Valencia, Spain

Abstract

Equipment ageing and technical obsolescence is a major concern in the analysis of reliability and availability of systems, which affects reliability, maintainability and availability of maintained equipment. An increased level of detail of the reliability and risk (R&R) models is required in order to simulate the equipment behaviour in the long term under uncertain degradation and obsolescence conditions. The challenge is twofold. On one hand, there is a need of developing advanced and detailed R&R models that address such impact explicitly. On another hand, the development of such models must be compatible with the available data to estimate the model parameters. One can face two extreme situations, the available data is scarce, or the available data is huge but must be classified first. This paper introduces such advanced and detailed R&R models and provide some guidance on how both situations are and could be managed. The example of application provides interesting results not only of the usefulness of such advanced model but also on the importance of accurate estimation of parameters.

Keywords: Ageing, technical obsolescence, maintenance optimization, advanced RAMS modelling, parameter estimation

1. Introduction

Equipment ageing is a major concern in the analysis of reliability and availability of systems no matter their function is devoted to perform a safety related-mission or a production process.

At the beginning, ageing was mainly linked to chronological time elapsed sin the installation of the equipment and the accumulated degradation as a consequence of both environmental and working conditions, which is normally named physical ageing. Thus, ageing impact on both reparable and no reparable components.

On the other hand, obsolescence is inherent to the evolution of whatever technological system. It has been an issue since the very beginning of industrial revolution. However, technical obsolescence is becoming day after day a more important concern due to several facts, such as technological innovation and development, etc. Technical obsolescence affects reliability and availability of maintained equipment and is considered a type of non-physical ageing.

Maintenance activities are directed to manage ageing of repairable equipment with at aim at keeping equipment reliability and availability and system risk under control. One can find in the literature many approaches proposed to establish, and even optimize, maintenance plans based on reliability and risk, e.g. reliability cantered maintenance. They make use of diverse equipment reliability and system risk models with different level of detail depending on the particular application.

Nowadays, the impact of both physical and non-physical ageing on system equipment impose new challenge, because of an increased level of detail of the reliability and risk (R&R) models is required in order to simulate the equipment behaviour in the long term under uncertain degradation and obsolescence conditions.

The challenge is twofold. On one hand, there is a need of developing advanced R&R models that take into account the reality of the physical impact of both ageing and maintenance effectiveness. Such an impact must be introduced in the R&R formulation explicitly so that appropriate tuning of parameters would make possible to simulate and forecast the real performance of the system equipment in the long term.

On another hand, the development of such models must be compatible not only with the physical phenomena but also with the available data. For example, data on equipment degradation and failures, and maintenance scheduling and effectiveness must be used to estimate the relevant parameters within the advanced and detailed R&R models. Here, one can face two extreme situations, the available data is scarce, or the available data is huge. Equipment belonging to safety or production systems are two examples of extreme situations.

This paper introduces such advanced and detailed R&R models and provide some guidance on how both situations are and could be managed.

2. Ageing and technical obsolescence management

Technical obsolescence of an item can be defined as its becoming out of date in comparison with current knowledge, standards and technology. For example, ageing and technical obsolescence management of the nuclear equipment are an important factor to achieve the safe operation of the nuclear plants (NPP), maintaining and reducing the failure probability of the components and the downtime for testing and maintenance. Equipment inherent reliability and maintainability can be affected by the technical obsolescence and ageing. On one hand, it may increase ageing rate and reduce maintenance effectiveness, which, in turn, increases equipment ageing and reduces its reliability. On another hand, it may increase maintenance downtime mainly because of provisioning logistics of spare parts.

Nowadays, NPP are in the process of implementing ageing and technical obsolescence management programs, with the objective of keeping such adverse effects under control adopting in each case the most appropriate type of strategy, i.e. whether the obsolete item will be replaced in its entirety or repaired, etc.

Consequently, the establishment of an effective testing and maintenance policy at the NPP must consider all relevant issues in an integrated manner including equipment ageing, technical obsolescence, human resources, maintenance planning and scheduling, etc.

3. Advanced R&R models

Normally, it is assumed that R&R models for an equipment may consist of three major contributions depending on the equipment function. For example, a safety-related equipment has usually up to three main types of failure modes that contribute to the equipment unreliability: (1) by demand-caused, (2) standby-related failures and (3) running failures. The first is often associated with a demand failure probability (ρ), and the second and third with a standby failure rate (λ_s) and a mission failure rate (λ_m) respectively. They are generally associated with constant values in a standard Probabilistic Risk Assessment, i.e. ρ_0, λ_{0s} and λ_{0m} respectively, which do not take into account the component degradation nor the maintenance effectiveness, the latter connected with technological obsolescence.

Early studies reported in [1, 2] have provided a well-organized foundation for the effects of degradation, e.g. due to equipment testing, and ageing, due to chronological time under given environmental and working conditions. For example, Kim et al. (1994) [2] proposed a simplified but well organized unreliability model for a safety-related equipment, which can be formulated as follows:

$$u_R(n,t') = \rho(n) + \int_{nT}^{nT+t} \lambda(n,u)du \qquad \text{for} \quad t' \in [0,T] \qquad (1)$$

being the demand-caused unreliability contribution

$$\rho(n) = \rho_0 + \rho_0 p_1 n \qquad (2)$$

and the standby-related unreliability contribution:

$$\lambda(n,u) = \lambda_0 + \lambda_0 p_2 n + \alpha v \qquad \text{for} \quad v \in [0, nT+t'] \qquad (3)$$

where,

n = number of test performed on the equipment at chronological time t
T = test interval
t' = time elapsed since the last test
v= time elapsed since the last overhaul point
ρ_0 = residual demand failure probability
p1 = test degradation factor associated with demand failures
p2 = test degradation factor associated with standby failures
λ_0 = residual standby time-related failure rate
α = aging factor associated with ageing alone

However, this model does not take into account the important positive effects on the component unreliability of the maintenance activities as a function of their effectiveness in managing component degradation. So that, later on, several models were proposed to introduce this and other positive effects [2-4].

Results of recent research have concluded that it is possible to formulate the most relevant unreliability parameters depending explicitly on the most relevant physical parameters linked to equipment degradation and maintenance effectiveness, the latter accounts also for technical obsolescence, which can be represented in a simplified way as follows [5]:

$$\lambda = f(\lambda_0, \alpha, IMM, \varepsilon_S, M, T, TR) \qquad n = 1,2,\dots,N \tag{4}$$

$$\rho = g(\rho_0,\ p_1, IMM, \varepsilon_D, M, T, TR) \qquad n = 1,2,\dots,N \tag{5}$$

Eq. (4) provides a single result for λ. This shows that this result depends on the ageing rate, represented by α, the maintenance plan, represented by ε_S, M and the Imperfect Maintenance model (IMM) ,PAS or PAR, the test interval T and the time of reference TR in which λ is evaluated.

Eq. (5) provides a single result for ρ. This shows that this result depends on the test degradation factor, represented by p_1, the maintenance plan, represented by ε_D, M and the IMM, PAS or PAR, the test interval T and the time of reference TR in which ρ is evaluated.

Ageing PSA model has been proposed in Ref. [3], in which the previous reliability models can be integrated to provide a risk model that includes explicitly the effect of the ageing, obsolescence, testing and maintenance on the plant components. There, for example, the risk model is formulated in terms of the Core Damage Frequency (CDF) and the unreliability of a given component as follows:

$$R = CDF_0 + u_R \cdot B \tag{6}$$

Where, CDF_0 [$year^{-1}$] represents the reduced risk when the component is known not to be down and B corresponds to the traditional Birnbaum importance measure of the component. As established in Ref. [3], using a level 1 PSA, the required risk metrics for the evaluation of risk impact of ageing and obsolescence is the assessment of the annual increase of the baseline CDF (ΔR), which can be formulated as follows:

$$\Delta R = \Delta u_R \cdot B \tag{7}$$

Where Δu_R is the change on the component unreliability for a given period considering the degradation, obsolescence, testing and maintenance effects. Eq (7) refers to the risk impact of a single component, however extension of the risk impact for several components is straightforward.

4. Parameter estimation

In R&R modelling it is critical the estimation of the unreliability and maintainability parameters since an accurate estimation of them is essential in applying these models, for example, to optimize maintenance strategies.

In practice, for safety related equipment, it is very difficult to collect large amounts of reliability and maintainability data to make an estimation precisely. Usually, when the available data is scarce, a Bayesian approach can be used [6]. This approach evaluates the joint probability distributions of the parameters assuming their joint probability distribution. Also, there are studies in which Maximum Entropy is used to solve estimation problems with data uncertainty [7].

With enough amount of data, parameter estimation can be performed using the previous approach. In addition, in this scenario Maximum Likelihood Estimation (MLE) method can be used. For example, by means of this approach it has been possible to estimate the parameters associated to the standby-related unreliability (λ_0, α, *IMM* and ε_S) and demand-caused unreliability (ρ_0, p_1, *IMM*, ε_S) of the eq. (4) and eq. (5), as follows [5]:

$$L_\lambda(\xi|model, observed\ data) = \prod_{failures} \lambda(t) \prod_{maintenance}[-\Lambda(t)] \tag{8}$$

$$L_\rho(\xi|model, observed\ data) = \prod_{failures} \frac{\rho(t)}{1-\mathrm{P}(t)} \prod_{maintenance}[1 - \mathrm{P}(t)] \tag{9}$$

being Λ and P the accumulated standby-related unreliability and accumulated demand-caused unreliability, respectively. By maximizing the eq. (8) and eq. (9) the maximum likelihood estimators of the objective parameters are obtained. The maximum likelihood estimation method provides, in addition to the parameter estimates, information on its variability through the Fisher information matrix. To maximize the likelihood functions, for example, a genetic algorithm can be used. However, in the study reported in [5], equipment was not affected by technical obsolescence, so that, its impact on maintenance effectiveness was not considered.

On another hand, it is quite common to have a big amount of reliability and maintainability data, even for safety related equipment, for which failures are scares, for example, when the whole NPP fleet of components is considered. Here, the analyst faces other sort of problems.

Often, the plant has an equipment fleet of similar components but with different characteristics. For example, motor-operated valves with different sizes and functions, which undertake similar or different maintenance activities. In the specific case of having available the historical maintenance and test data of the plant, clustering algorithms can be used for grouping segmented equipment populations, which exhibits similar ageing and degradation patterns, based on their operational and physical characteristics [8].

However, technical obsolescence may affect such patterns. For example, imagine an equipment segment with components affect by obsolescence differently. In this specific case, maintainability parameters can be estimated considering only the group of components within the segment affected be same type of technical obsolescence. Thus, both reliability and maintainability parameters can be estimated together by reorganizing the segmented groups or, alternatively, once the reliability parameters have been found proceed to perform the estimation of the maintainability ones individually taking into account the different factors that influence maintenance effectiveness. Factors such as obsolescence, availability of stock or maintenance cost, type of repair and maintenance, etc. could be considered in estimating the effectiveness of maintenance. Classification models, such as Support Vector Machine (SVM), neural network or random forest, could be used to estimate the effectiveness of maintenance including as covariates the above factors.

5. Case of application

A sensitivity analysis of the R&R models parameters is presented, which are affected by ageing and obsolescence. A motor-operated valve (MOV) of the Auxiliary Feed Water System (AFWS) of a NPP is considered, which is one of the most important components for plant safety. Based on the R&R models introduced in section 3 and for sake of simplicity in the presentation and discussion of the results found, it is assumed that technical obsolescence degrades only maintenance effectiveness $\{\varepsilon_S, \varepsilon_D\}$ and no other compensatory measure are considered than re-adjusting testing and maintenance intervals. In this sensitivity study, optimal maintenance and testing intervals are sought for maintenance effectiveness $\{\varepsilon_S, \varepsilon_D\}$ ranging within [0.1; 0.9].

Figure 1 shows the optimal results found of test interval (T) and maintenance interval (M), and the corresponding unreliability (u_R) for each couple $\{\varepsilon_S, \varepsilon_D\}$. As shown in the optimization results, small variations in the maintenance effectiveness produce considerable changes in the optimal T and M planning.

Figure 1 shows also two type of results, acceptable (text in black) and unacceptable (text in red), according to the risk impact of ageing and obsolescence. Thus, no matter testing and maintenance intervals have been optimized in all cases, the annual risk increase of the baseline CDF, represented by ΔR in Eqn. (7), goes beyond the safety limit acceptable for a NPP for the solutions highlighted in red.

Therefore, other compensatory measures than just re-adjusting testing and maintenance intervals must be considered in order to manage technical obsolescence and ageing appropriately to keep the risk impact below the safety limit. Note, these situations correspond to low or very low maintenance effectiveness, which requires further research being conducted nowadays. This reinforce the fact that it is necessary to use statistical methods and tools that result in accurate estimations of the unreliability parameters to guarantee an optimal obsolescence and ageing management.

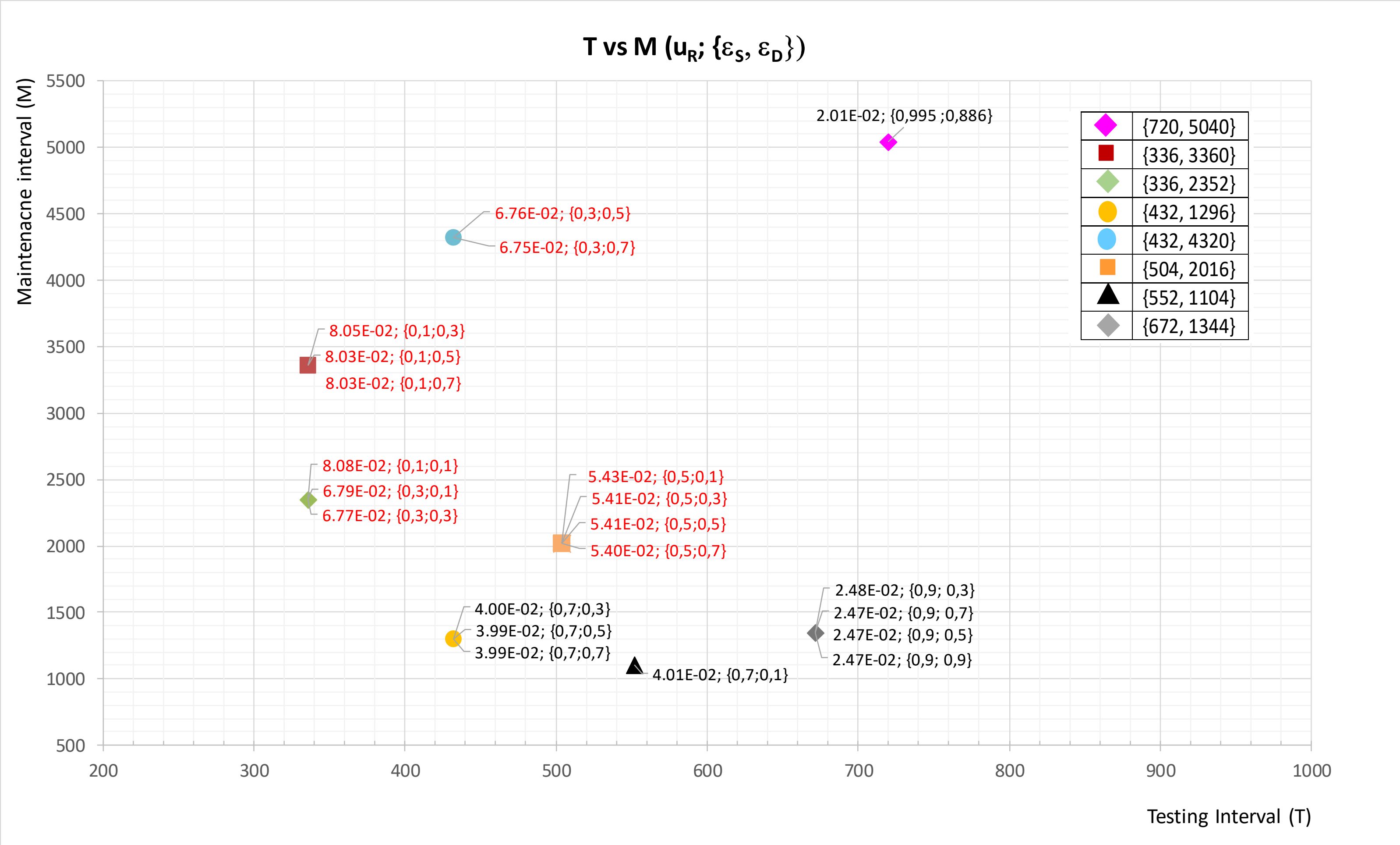


Figure 1. Optimal T versus M for different values of ε_S and ε_D.

Acknowledgements

The authors are grateful to the Spanish Ministry of Science and Innovation for the financial support received (Research Project ENE2016-80401-R) and the doctoral scholarship awarded (BES-2014-067602). The study also received financial support from the Spanish Research Agency and the European Regional Development Fund.

References

[1] Kim, I. S., Samanta, P. K., Martorell, S., & Vesely, W. E. (1992). Quantitative evaluation of surveillance test intervals including test-caused risks (NUREG/CR-5775). Washington, DC: US Nuclear Regulatory Commission.

[2] Martorell S., Sánchez A., Serradell V. (1999). Age-dependent reliability model considering effects of maintenance and working conditions. *Reliability Engineering and System Safety*; 64: 19–31.

[3] Martón, I., Sánchez, A. I. & Martorell, S. (2015). Ageing PSA incorporating effectiveness of maintenance and testing. *Reliability Engineering & System Safety*, 139, 131-140.

[4] Martorell, P., Martón, I., Sánchez, A. I., & Martorell, S. (2017). Unavailability model for demand-caused failures of safety components addressing degradation by demand-induced stress, maintenance effectiveness and test efficiency. *Reliability Engineering & System Safety*. Reliability Engineering & System Safety, 168, 18-27.

[5] Martorell, S., Martorell, P., Sanchez, A. I., Mullor, R., & Martón, I. (2017). Parameter estimation of a reliability model of demand-caused and stand-by related failures of safety components exposed to degradation by demand stress and ageing that undergo imperfect maintenance. *Mathematical Problems in Engineering*, 2017, 11 https://doi.org/10.1155/2017/7042453.

[6] Moneim A. F., Ghazy M. and Hassnien A. (2019). Bayesian Estimation of Parameters of Reliability and Maintainability of a Component under Imperfect Repair and Maintenance. *Arabian Journal for Science and Engineering* 44:2497–2502. https://doi.org/10.1007/s13369-018-3206-2.

[7] Muñoz-Cubo J.L., Mendizabal R., Miquel A., Berna C., Escrivá A. (2018) .Use of the Principles of Maximum Entropy and Maximum Relative Entropy for the Determination of Uncertain Parameter Distributions in Engineering Applications. Entropy 2017, 19, 486

[8] Cannarile, Francesco & Compare, Michele & Di Maio, Francesco & Zio, Enrico. (2018). A Clustering Approach for Mining Reliability Big Data for Asset Management. Proceedings of the Institution of Mechanical Engineers Part O Journal of Risk and Reliability. 232. 10.1177/1748006X17716344.

Robust Statistics for (big) data analytics

Riani M[1]., Atkinson A.C.[2], Corbellini A.[1], Perrotta D.[3] & Torti F.[3]
[1]University of Parma
[2]LSE
[3]European Commission Joint Researc Centre
mriani@unipr.it

Abstract

Data rarely follow the simple models of mathematical statistics. Often, there will be distinct subsets of observations so that more than one model may be appropriate. Further, parameters may gradually change over time. In addition, there are often dispersed or grouped outliers which, in the context of international trade data, may correspond to fraudulent behavior. All these issues are present in the datasets that are analyzed on a daily basis by the Joint Research Centre of the European Commission and can only tackled by using methods which are robust to deviations to model assumptions (see for example Perrotta et al., 2020).

This distance between mathematical theory and data reality has led, over the last sixty years, to the development of a large body of work on robust statistics. In the seventies of last century it was expected that in the near future "any author of an applied article who did not use the robust alterative would be asked by the referee for an explanation" (Stigler, 2010). Now, a further forty years on, there does not seem to have been the foreseen breakthrough into the wider scientific universe. In this talk, we initially sketch what we see as some of the reasons for this failure, suggest a system of interrogating robust analyses, which we call "monitoring" (Cerioli et al. 2018) and describe the robust and efficient methods which are currently used by the Joint Research Centre of the European Commission to detect model deviations, groups of homogeneous observations (Torti et al. 2018), multiple outliers and/or sudden level shifts in time series (Rousseeuw et al. 2019).

Particular attention will be given to robust and efficient methods (kwown as forwad search) which enables to use a flexible level of trimming and understand the effect that each unit (outlier or not) exerts on the model (see for example Atkinson and Riani, 2000, Riani Atkinson and Cerioli, 2009).

Finally we discuss the extension of the above methods to transformations (Atkinson et al. 2020) and to the big data context. With a large set of data, any model is liable to be approximate. Our first guesses at a model may even be seriously inadequate. But a suitable measure of lack of fit of themodel can provide low-dimensional plots that are informative about inadequacies. The inadequacy may be systematic, or it may be due to lack of homogeneity in the data as well as to the presence of, possibly large, numbers of outliers. For

these reasons, we need to fit the model in a robust manner. It is customary in robust analyses to assume that at least 50% of the observations come from the same uncontaminated distribution. However, this need not be the case when there are several clusters. For small samples, starting in different clusters can be informative about cluster structure. Random start forward searches (Atkinson et al. 2018) provide one way of starting in a variety of conditions. Theoretical results on such extreme trimming are given by Cerioli et al. (2019). All the methods described in the talk have been included in the FSDA Matlab toolbox freely donwloadable as a toolbox from Mathworks file exchange or from github at the web address https://uniprjrc.github.io/FSDA/

References:

Atkinson A.C., Riani M. (2000) Robust Diagnostic Regression Analysis, Springer–Verlag, NewYork.

Atkinson A.C., Riani M., Cerioli A. (2010) The forward search: theory and data analysis (with discussion). J Korean Stat Soc 39:117–134. https://doi.org/10.1016/j.jkss.2010.02.007

Atkinson A.C., Riani M., Cerioli A. (2018) Cluster detection and clustering with random start forward searches. J Appl Stat 45:777–798. https://doi.org/10.1080/02664763.2017.1310806

Atkinson A.C., Riani M., Corbellini A. (2020). The analysis of transformations for profit-and-loss data, Journal of the Royal Statistical Society, Series C, Applied Statistics, https://doi.org/10.1111/rssc.12389

Cerioli A., Farcomeni A., Riani M. (2019) Wild adaptive trimming for robust estimation and cluster analysis. Scand J Stat 46:235–256. https://doi.org/10.1111/sjos.12349

Cerioli A., Riani M., Atkinson A.C., Corbellini A. (2018) The power of monitoring: how to make the most of a contaminated multivariate sample (with discussion). Stat Methods Appl 27: 559–587. https://doi.org/10.1007/s10260-017-0409-8

Cerioli A., Riani M., Atkinson A.C., Corbellini A. (2018) The power of monitoring: how to make the most of a contaminated multivariate sample (with discussion). Stat Methods Appl 27:559–587. https://doi.org/10.1007/s10260-017-0409-8

Perrotta D., Cerasa A., Torti F., Riani M. (2020), The Robust Estimation of MonthlyPrices of Goods Traded by the Euro-pean Union A technical guide, JRC report, https://ec.europa.eu/jrc/en/publication/robust-estimation-monthly-prices-goods-traded-european-union

Riani,M., Atkinson, A. C. and Cerioli, A. (2009). Finding an unknown number of multivariate outliers. J. R. Stat. Soc. Ser. B Stat. Methodol. 71

447–466. https://doi.org/10.1111/j.1467-9868.2008.00692.x

Rousseeuw P.J., Perrotta D., Riani M., Hubert M. (2018). Robust Monitoring of Time Series with Application to Fraud Detection. Econometrics and Statistics. https://doi.org/10.1016/j.ecosta.2018.05.001

Stigler S.M. (2010). The changing history of robustness. The American Statistician, 64, 277–281.

Torti F.. Perrotta D., Riani M. Cerioli A. (2018). Assessing Trimming Methodologies for Clustering Linear Regression Data, Advances in Data Analysis and Classification: 227-257, https://doi.org/10.1007/s11634-018-0331-4

See slide presentation in Appendix 1.

A proposal of an algorithm to simulation right censored data type I in reliability field

D. Gaspar[1], L. Andrade Ferreira[2]
[1] Polytechnic Institute of Viseu, Viseu, Portugal
danigaspar@estgv.ipv.pt
[2] University of Porto, Faculty of Engineering, FEUP/DEMec, R. Roberto Frias, 4200-465 Porto, Portugal, lferreir@fe.up.pt

Abstract

In this presentation a method to develop an algorithm to simulate the behavior of right censored data, type I is proposed and explained. In simulation (and in particular in reliability), a lot of studies and works use censored data. There are a large number of studies that use censored data with different statistical distributions and perform sensitivity analyzes by changing the distribution and simulation parameters. These studies can be compared to verify their validity, but for this to be possible, they need to have some common indicators in order to validate the object of study, the model or the best optimization, etc.

In the case of censored data right type I simulation, the most common method that use different levels of percentage of censored data were studied. An analysis was made with the most common distributions in scientific research on reliability and maintenance: Weibull, normal, gamma, log-normal and exponential.

In this work, simulation algorithms for reliability models for complex equipment / systems were developed, when data collection is confronted with censored data.

The algorithms are innovative and their development was done in three different software: Python, Matlab and R.

A methodology of analysis (hypothesis tests) and validation with an evaluation matrix is proposed to test the i.i.d. data of RNG of censored data.

Keywords: Data censored, Reliability, Algorithm simulation, Statistical distribution

See slide presentation in Appendix 1.

IOT future in Energy Industries: an outlook

M. Raza
GE Power, Switzerland
Mohammad.Raza@ge.com

Abstract

As world evolves fast into a very different future and still 80% of the future jobs in 2030 are being defined, IOT is here to stay and influence everyday living. Businesses with a traditional way of doing things are seeing more and more incursion of IoT. It has already influenced the standard of living today from its impact on consumer-based products. The future will be the Industrial IoT (IIoT) infrastructure and platforms development along with influencing the way we operate and manage Industrial assets.

Energy sector along with Oil and Gas sector have been second in line in adopting digital transformation after space, aviation and health care industries. This brings in challenges specially on consolidation of data, advanced analytics, data sharing legislations and the most prominent being the cybersecurity both for device and network. Energy sector is most vulnerable to the attacks and so far, companies are using temporary security fixes and patches. Many are predicting that the blockchain with the merger of IoT and Artificial Intelligence (AI) would be applied to prevent attacks on large and valuable setups. The predication is implementing advanced technology-based security enabled hardware and software will be the major area of allocation of funds throughout the next decade.

This paper will provide an outlook on how IIoT is transforming the operational and financial benefits of Industrial set ups by combining machine to machine communication with industrial big data of predictive analytics and also the challenges it possesses from concept design till end of life. An independent view of the entire lifecycle perspective will be presented.

See slide presentation in Appendix 1.

List of Authors

Atkinson, Anthony C.	96
Corbellini, Aldo	96
Eisinger, Siegfried	4
Fackrell, Justin	4
Ferreira, Luís	66, 99
Fink, Olga	2
Gaspar, Daniel Mendes	99
Ge, Xiaocheng	20
Glomsrud, Jon Arne	4
Harrison, Chris	20
Lera St.Clair, Asun	68
Martón, Isabel	88
Martorell, Sebastian	88
Murgia, Alessandro	12
Pereira, Luís	66
Perrotta, Domenico	96
Raza, Mohammad	100
Riani, Marco	96
Sanchez, Ana I.	88
Stow, Julian	20
Torti, Francesca	96
Tourwé,Tom	12
Tsiporkova, Elena	12
Verbeke, Mathias	12
Wels, Henk	33

Mechanical seal failure prediction in an oil refinery: a first attempt to solve the problem using a data-driven approach

Luís F. Pereira[1 2] Luís A. Ferreira[1]

[1]Faculdade de Engenharia da Universidade do Porto

[2]lfpereira@fe.up.pt

October, 2019

Ideally

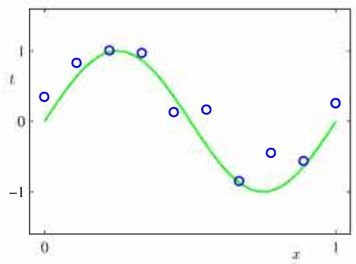
$$\boldsymbol{x} \longrightarrow \boxed{\text{model}} \longrightarrow y$$

- $\boldsymbol{x}$: input parameters
  - Variables with *predictive* power
  - e.g.:
    1. running time
    2. physical variables
- y: time to failure
  - Easily converted in operation/maintenance costs
- If the model is accurate enough: possible to simulate the cost impact of changing a given parameter

Reality

$\boldsymbol{x}^* \longrightarrow$ model $\longrightarrow y \in \{0, 1\}$

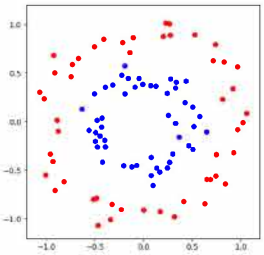
- $\boldsymbol{x}^*$: *incomplete* input parameters
  - Important variables are missing
  - Gathering and storing data is expensive
- y: binary/categorical output



Reality

$\boldsymbol{x}^* \longrightarrow$ model $\longrightarrow y \in \{0, 1\}$

- $\boldsymbol{x}^*$: *incomplete* input parameters
  - Important variables are missing
  - Gathering and storing data is expensive
- y: binary/categorical output

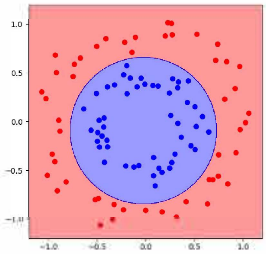


Motivation | Data processing | Machine Learning | Final remarks

Reality

$$x^* \longrightarrow \boxed{\text{model}} \longrightarrow y \in \{0, 1\}$$

- x^*: *incomplete* input parameters
  - Important variables are missing
  - Gathering and storing data is expensive
- y: binary/categorical output



L. F. Pereira, L. A. Ferreira, FEUP | Mechanical seal failure prediction in an oil refinery | 3 / 20

Motivation | Data processing | Machine Learning | Final remarks

Reality (cont'd)

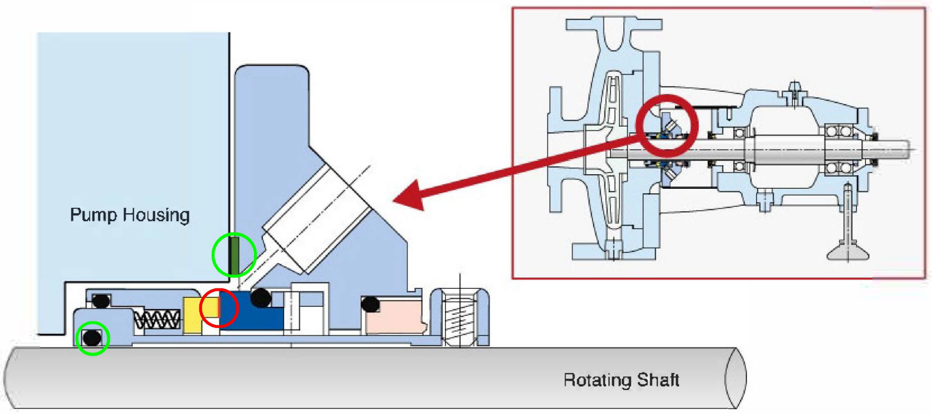
- Further simplification: independent datapoints, i.e. no information regarding chronological order
  - Special care must be taken during model selection and validation
  - Valuable information not considered
  - Alternative: recurrent neural networks, hidden Markov models, . . .
- Although reality is simplified, valuable information can be extracted from such a model
  - Potential application: planning of maintenance operations in a timely and thoughtful manner

L. F. Pereira, L. A. Ferreira, FEUP | Mechanical seal failure prediction in an oil refinery | 4 / 20

Data-driven approach in an oil refinery

- application of the conceptual ideas requires:
  - Very hard and meaningful problem to solve
    - otherwise, choose a simple and more interpretable strategy
  - Data availability
- Our case study:
  - Possibility of accessing maintenance data of an oil refinery
  - Plenty of mechanical equipments
  - Multicelular centrifugal pumps specially ”troubled":
    - Sealing system (mechanical seal) identified as main failure mechanism

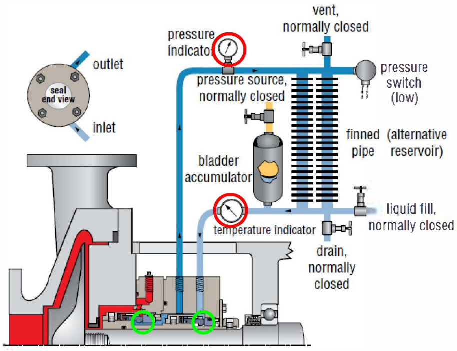
Mechanical seal

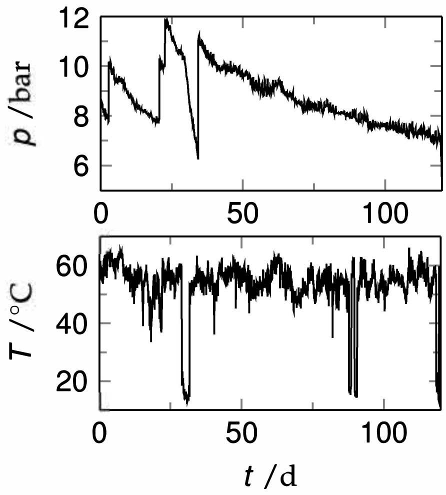


- Three sealing points: 2 static and 1 dynamic

Available data

- ANSI/API 682: Plan 53B





Available data (cont'd)

- Real-time data
  - Available variables:
    - Temperature
    - Pressure
    - Electric current
  - Highly-reliable data
  - But, missing points
    - Due malfunctioning of measure devices, extraction of protection devices or communication failure
    - Reconstructed using auto-regressive models (alternative: Machine Learning)
- Maintenance operations information (from SAP)
  - Available information:
    - Component identification and failure mechanisms
    - Maintenance dates
    - Number of used mechanical seals
  - Harder to analyze due its dependency on human intervention
  - Need for development of internal standards and formation of human resources

Creating information

- Plenty of data, but few variables
- Different sources of data: combination leads to more robust analyses and ensure complete exploitation of available information

$\Downarrow$

Create new information

Motivation | Data processing | Machine Learning | Final remarks

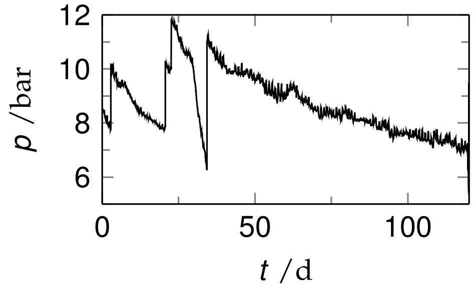
Creating information (cont'd)

- *motor*: binary variable with status of the motor in each instant (from temperature and electric current)
  1. Number of starts
  2. Time since last start/stop
  3. Pump operation time and calendar time
  4. Time of operation since mechanical seal's substitution and mechanical seal's life time
  5. Time since last pressurization
  6. Time since last substitution of opposite mechanical seal

Motivation | Data processing | Machine Learning | Final remarks

Creating information (cont'd)

- Pressurization: steep pressure rising
  1. $p \uparrow \geq 1$ bar
  2. $p_{final} \geq 5$ bar
- Slopes: change point detection (Matlab)
- Other variables:
  1. Number of pressurizations
  2. Pressure after pressurization
  3. Pressure increase during pressurization
  4. Number of pressure drops (4 bar)



L. F. Pereira, L. A. Ferreira, FEUP | Mechanical seal failure prediction in an oil refinery | 11 / 20

Motivation | Data processing | Machine Learning | Final remarks

Goal definition

$feature_1$	$feature_2$	...	$feature_{19}$	Classification
59.260	11.14	...	1	*Positive*
⋮	⋮	⋱	⋮	⋮
8.410	10.30	...	0	*Negative*

- Possible outputs:
  - *Positive*: mechanical seal will fail within 1 month
  - *Negative*: mechanical seal working properly
- But what is a failure?
  - A failure occurs when the pressure slope is higher than a threshold
  - Threshold defined taking into account maintenance team interventions
- From here, a Machine Learning framework must be followed (problem independent)

L. F. Pereira, L. A. Ferreira, FEUP | Mechanical seal failure prediction in an oil refinery | 12 / 20

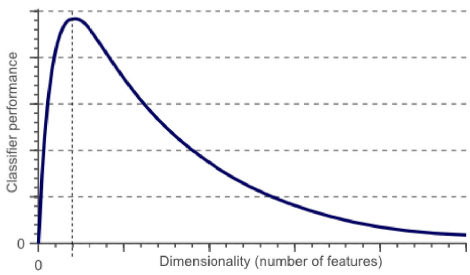
Model validation

- Data set split:
  - *Data set* = *Train set* + *Test set*
  - *Training set* = *training set* + *development set*
- Evaluation measures in the *training set* are highly optimistic
- Split cannot be random

$n = 6x10^5$	Quantity /%	Negative /%
training set (until 2016)	52.9	59.2
development set (2016)	23.1	74.0
Test set (2017)	24.1	64.8

Feature engineering

- Feature engineering is fundamental due the curse of dimensionality
  - What are the most important features?
  - How many features should be used?
  - Should we combine features (feature extraction)?
  - Take care of redundant information



Model selection and first results

- Plethora of Machine Learning algorithms available
  - Suggestion: if computational feasible, treat the algorithm as an hyper-parameter
- First results (only decision trees, naive Bayes and linear discriminant analysis models were trained):

		development set		*training set*	
		acc /%	*F_1 score /%*	*acc /%*	*F_1 score /%*
	10	66.8	56.4	89.3	87.1
DT	30	66.0	54.4	98.6	98.3
	50	66.7	55.1	100.0	100.0
LDA		71.8	54.1	76.7	72.5
NB		73.3	61.1	75.7	72.7

Motivation | Data processing | Machine Learning | Final remarks

Model selection and first results

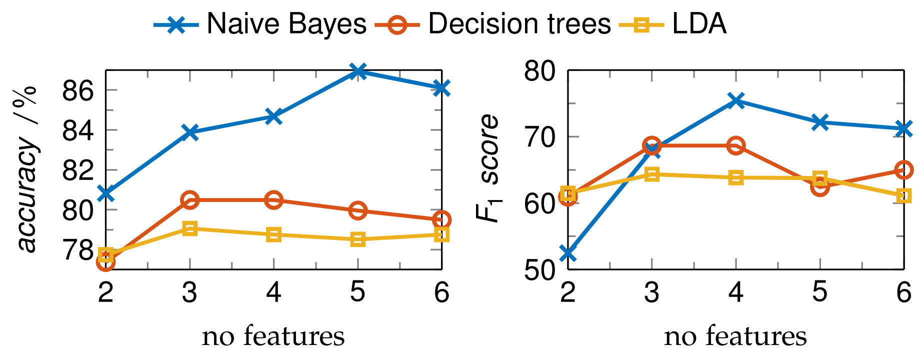
- Plethora of Machine Learning algorithms available
  - Suggestion: if computational feasible, treat the algorithm as an hyper-parameter
- First results (only decision trees, naive Bayes and linear discriminant analysis models were trained):

		development set		*training set*	
		acc /%	*F_1 score /%*	*acc /%*	*F_1 score /%*
	10	66.8	56.4	89.3	87.1
DT	30	66.0	54.4	98.6	98.3
	50	66.7	55.1	100.0	100.0
LDA		71.8	54.1	76.7	72.5
NB		73.3	61.1	75.7	72.7

L. F. Pereira, L. A. Ferreira, FEUP | Mechanical seal failure prediction in an oil refinery | 15 / 20

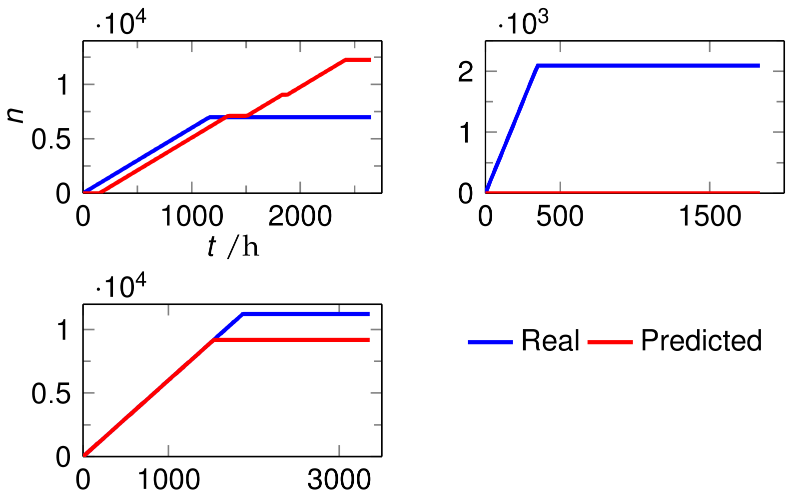
Further results

- Looking up an appropriate number of features:



- *Naive Bayes* with *accuracy*=87%

Further results (cont'd)



Further results (cont'd)

- Best models with a very poor performance in the test set

/%		no *features* 4	5
development set	*acc*	84.7	86.9
	F_1 *score*	75.4	72.2
Test set	*acc*	50.9	56.5
	F_1 *score*	59.3	56.1

Final remarks

1. Robust data-driven methodology to tackle problems in an industrial setting
2. Importance of gather and store proper data
3. Data under-exploitation and advantages of information integration
4. Importance of technical knowledge to guide data processing step
5. Importance of proper feature selection and extraction (could it give insights about what to measure?)
6. *Overfitting* and how it can be counter-intuitive: the importance of a correct model validation to avoid deceptions

57th ESReDA Seminar on
Advances in Reliability, Risk and Safety Analysis with Big Data
23rd and 24th October 2019
Universitat Politècnica de Valencia, Spain

Robust statistics for big data analytics

Marco Riani

mriani@unipr.it — http://www.riani.it

University of Parma — Robust Statistics Academy

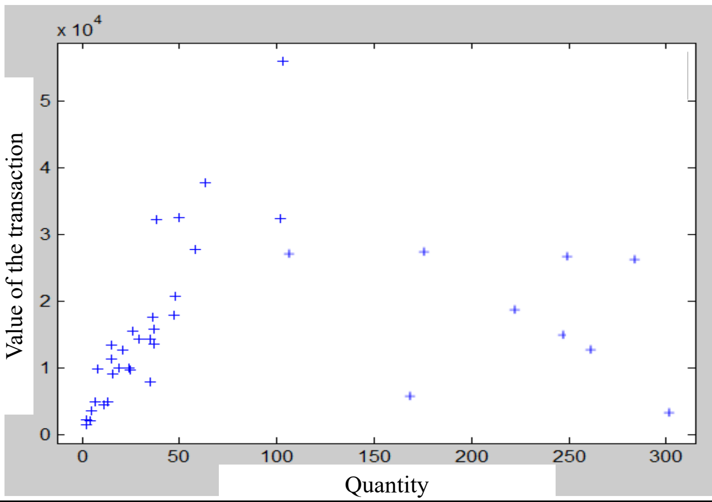
Joint work with Anthony Atkinson (LSE), Andrea Cerioli, Aldo Corbellini (Univ. of Parma), Domenico Perrotta, Francesca Torti, Andrea Cerasa, Emmanuele Sordini (JRC), Agustin Mayo Iscar, Luis Angel Garcia Escudero (Univ. of Valladolid)

1

Why robust statistics?

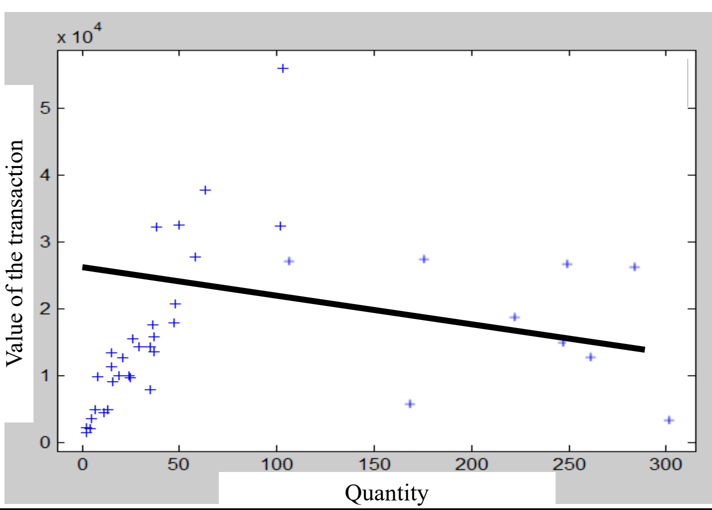
2

Example from international trade data (X-axis quantity Y-axis value of the transaction)



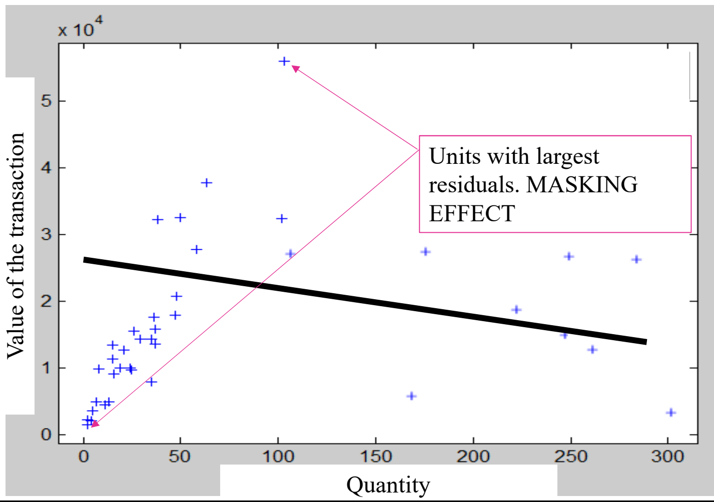
3

Example of non robust fit



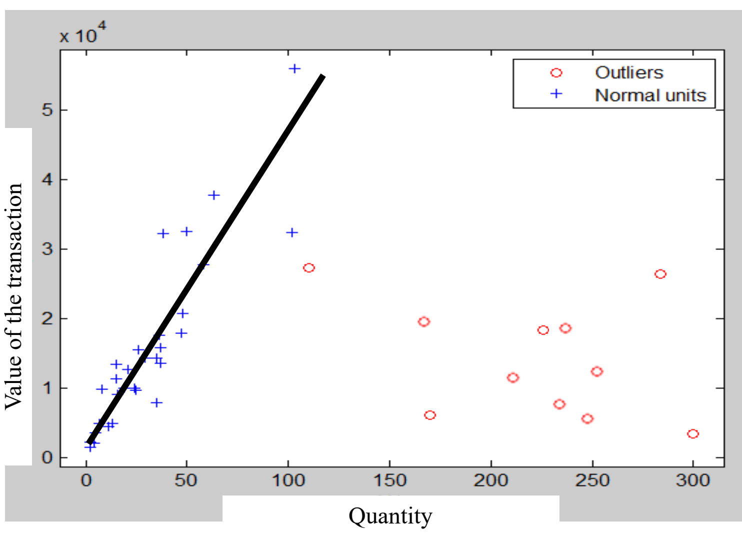
4

Example of non robust fit



5

Example of robust fit



6

History of Robust statistics

- Awareness of the importance of immunizing against outliers / gross errors is as old as the experimental approach to science
- **Thucydides** (History of The Peloponnesian War): "in 428 B.C. the Plataeans, besieged by the Spartans, **excluded extreme measurements** when estimating the height of the walls and managed to break the siege"

Reference due to Dr. Spyros Arsenis (JRC)

7

Classical and robust theory

Classical

- The observations are distributed according to F_θ
- Example
- $F_\theta = N(\mu, \sigma^2)$
- $\theta = (\mu, \sigma^2)$

Robust

- F_θ is considered as a mathematical abstraction which is only an ideal approximation to reality. The goal is to produce statistical procedures which still behave fairly well:
  - under deviations from the assumed model;
  - under the null model.

8

The grand plan

- Andrews *et al.* (1972) (the Princeton Robustness Study), at which time it was expected that all statistical analyses would, by default, be robust
- "**any author of an applied article who did *not* use the robust alternative would be asked by the referee for an explanation"**.
- "**from the 1970s to 2000 we would see ... extensions to linear models, time series, and multivariate models, and widespread adoption to the point where every statistical package would take the robust method as the default ...**"

Clearly, the Grand Plan is not what has occurred

9

Four possible explanations

- **Awkward asymptotics**: (**little interest on empirical performance**)
- **Awkward computation (**efficient algorithms, but with **black-box** style and the **need to select some tuning constants**)
- **Simplistic gaussian models** for the "good" part of the data: reality is usually more complex (clusters, skew distributions, heavy tails ...)
- **Empirical test size** is much bigger than nominal (outliers are found everywhere)

10

Which Method and How to Tune It?

Three classes of estimators:

- **Hard (0,1) trimming** (LTS, LMS, MCD, MVE) in which the amount of trimming is determined by the choice of the trimming parameter.
- **Adaptive Hard Trimming**. In the Forward Search (FS), the observations are again hard trimmed, but the amount of trimming is determined by the data, being found adaptively by the search.
- **Soft trimming** (downweighting). M estimation and derived methods (S, MM, tau). ρ function ensures that increasingly remote observations have a weight that decreases with distance from the centre.

11

Breakdown point and efficiency

- Breakdown point (bdp)= percentage of outliers the estimator can cope with
- Efficiency (eff) = $\frac{\mathrm{cov}(\beta_{ROBUST})}{cov(\beta_{LS})}$
- One needs to use tuning constants (consistency factors) which come from asymptotic theory

12

Regression setting

- Data (y_i, x_i) i=1, 2, …, n
- $y_i \in R$ Response
- $x_i \in R^p$ Predictors
- Regression model $y_i = x_i^T \beta + \sigma u_i$
- Predict y_i by $x_i^T \hat{\beta}$
- Residuals for given β: $r_i = r_i(\beta) = y_i - x_i^T \beta$

13

Least median of Squares (LMS)

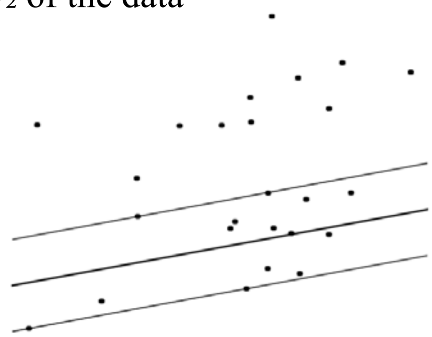
$$\min_{\beta} Me\{r_i(\beta)^2\}$$

- In the univariate case LMS becomes the Midpoint of the SHORTest Half = SHORTH
- **SHORTH = shortest** interval that covers **half** of the values

14

LMS with one explanatory variable

- It is the centre-line of the shortest (narrowest) strip containing ½ of the data



15

Properties of LMS

- It does not require a scale estimate
- Regression, scale and affine equivariant
- bdp=0.5, Fisher consistent and asymptotically normal

16

Drawbacks of LMS

- It displays marked sensitivity to central data values (not locally stable)
- It is very inefficient (converges at a rate $1/\sqrt[3]{n}$)

$$\sqrt{n}||\hat{\beta} - \beta|| \to_p \infty$$

$$\sqrt[3]{n}||\hat{\beta} - \beta|| = O_p(1)$$

17

LTS (least trimmed squares)
Rousseeuw (1984)

$$\hat{\beta}_{LTS} = \min_{\hat{\beta}_0,\, \hat{\beta}_1, \dots,\, \hat{\beta}_{p-1}} \sum_{i=1}^{h} r_{(i)}^2$$

$$r_{(1)}^2 \leq r_{(2)}^2 \leq \cdots \leq r_{(n)}^2 \quad h = [n/2]$$

- The extention to regression of the trimmed mean

18

Regression S estimators
Rousseeuw and Yohai (1984)

- LTS minimizes a robust residual scale estimate

$$\hat{\beta}_{LTS} = \arg\ \min_{\beta} \hat{\sigma}_{LTS}(\beta)$$

$$\hat{\sigma}_{LTS}(\beta) = \min \frac{1}{n-h} \sum_{i=1}^{h} [r_i(\beta)^2]_{(i)} \qquad \text{(biased)}$$

- Idea: minimize a more efficient robust scale estimator

- Regression S estimator minimizes an M estimate of scale

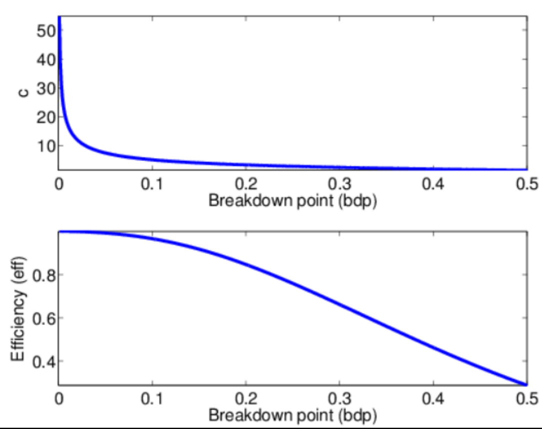
$$\hat{\beta}_{S} = \arg\ \min_{\beta} \hat{\sigma}_{M}(\beta)$$

Need to choose the consistency factor c of the associated ρ (ψ= ρ`) function

19

Efficiency and breakdown point

Consistency factor c (top panel) and efficiency (bottom panel) as a function of the breakdown point (bdp) for Tukey's Biweight.



20

More on the link between robustness and efficiency is in

TEST
DOI 10.1007/s11749-014-0357-7

ORIGINAL PAPER

On consistency factors and efficiency of robust *S*-estimators

Marco Riani · Andrea Cerioli · Francesca Torti

Received: 5 May 2013 / Accepted: 16 January 2014

21

Regression MM estimators

Yohai (1987)

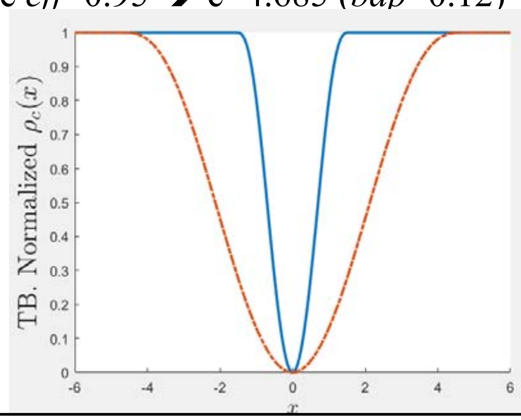
$$\min \sum_{i=1}^{n} \rho_{eff=0.95} \left(\frac{r_i(\beta)}{\hat{\sigma}_S(\hat{\beta}_S)} \right)$$

- Idea: fix *bdp*=0.5 and using $\rho_{bdp=0.5}$ find $\hat{\beta}_S$ and $\hat{\sigma}_S$ using S estimators
- Fix *eff*=0.95, and using $\rho_{eff=0.95}$ using $\hat{\beta}_S$ and $\hat{\sigma}_S$ as starting values in the weighted least squares loop
- The estimate of the scale is kept fixed in the iterative procedure

22

Claim of MM estimators

- HIGHLY ROBUST AND EFFICIENT
- Tukey's biweight rho (TB):
- blue line *bdp*=0.5 ➔ c=1.548 (*eff*=0.29)
- red line *eff*=0.95 ➔ c=4.685 (*bdp*=0.12)



23

More on this in

Statistical Science
2014, Vol. 29, No. 1, 128–143
DOI: 10.1214/13-STS437

A Parametric Framework for the Comparison of Methods of Very Robust Regression

Marco Riani, Anthony C. Atkinson and Domenico Perrotta

- Riani M., Atkinson A.C., Perrotta D. (2014). A Parametric Framework for the Comparison of Methods of Very Robust Regression Statistical Science, Vol. 29, No. 1, pp. 128–143.

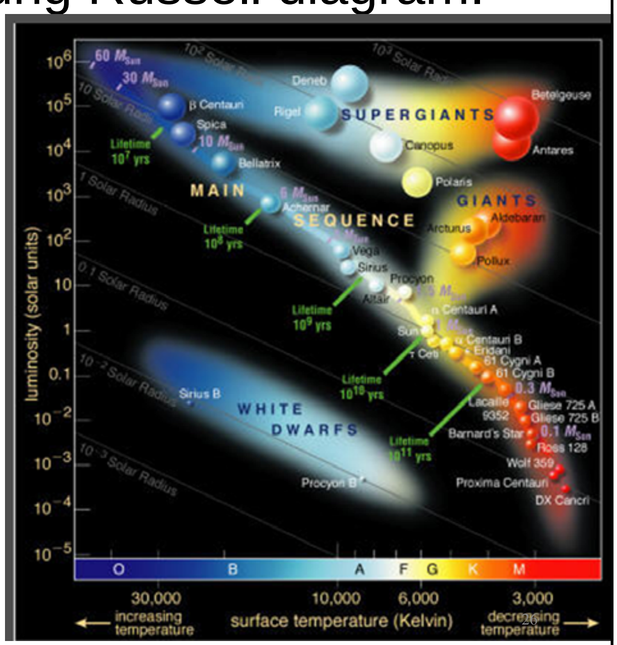
24

Robust regression in action

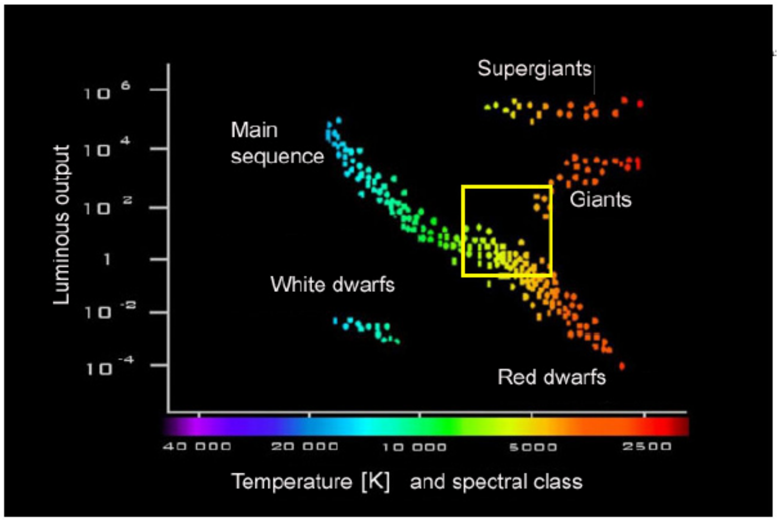
25

Hertzsprung Russell diagram.

- Graph showing the luminosity of a star as a function of its surface temperature

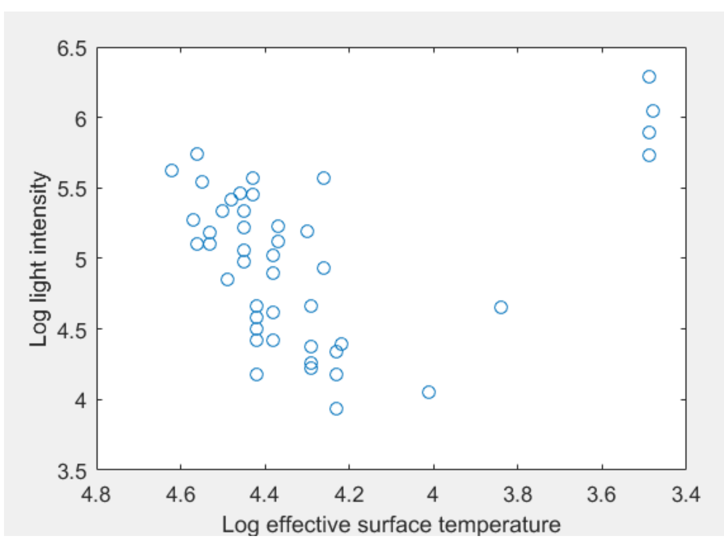


The extracted data come from the yellow square



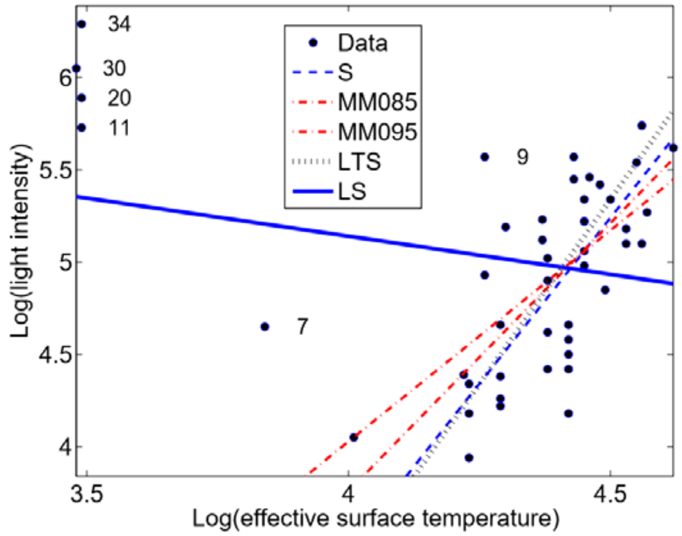
27

Log light intensity vs Log effective surface termperature (reverse order)



28

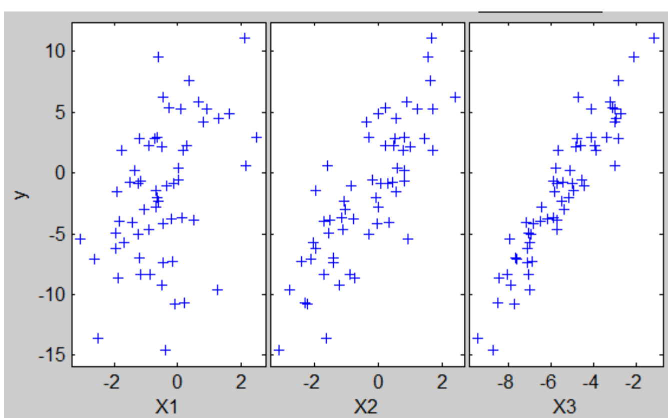
A comparison of different fits



29

Another example: a regression dataset with masked outliers (AR data)

- 60 observations, 3 explanatory variables



30

TRADITIONAL WAY OF DOING STATISTICS IN REGRESSION

```
>> mdlr = fitlm(X,y);
>> mdlr

Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    _______    ______    __________

    (Intercept)      11.174    0.67501    16.553    3.1288e-23
    x1             -0.21796    0.17244    -1.264       0.21146
    x2               1.4981    0.15534    9.6439    1.6733e-13
    x3               2.2596    0.13668    16.531    3.3265e-23


Number of observations: 60, Error degrees of freedom: 56
Root Mean Squared Error: 1.09
R-squared: 0.965,  Adjusted R-Squared 0.963
F-statistic vs. constant model: 510, p-value = 1.33e-40
```

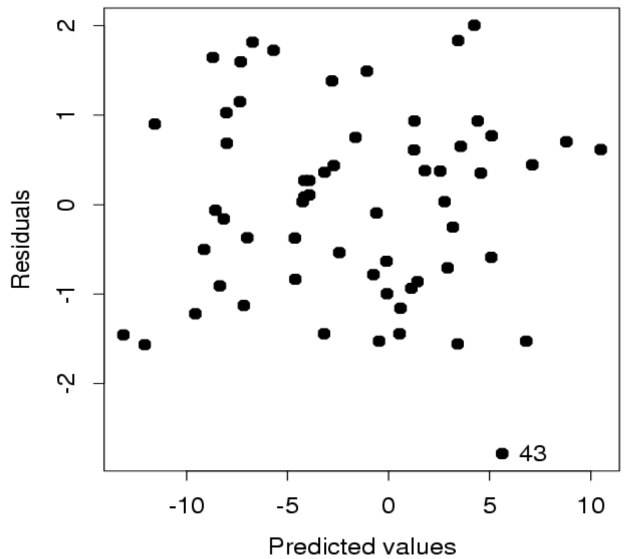
31

Statistics toolbox: RobustOpts on

```
>> mdlr = fitlm(X,y,'RobustOpts','on');
>> mdlr

mdlr =
Linear regression model (robust fit):
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate      SE       tStat       pValue
                   ________    _______    _______   __________

    (Intercept)      11.415    0.71721     15.915   1.9345e-22
    x1             -0.25422    0.18322    -1.3875      0.17078
    x2               1.4662    0.16506     8.8832   2.7871e-12
    x3               2.3066    0.14523     15.883   2.1262e-22


Number of observations: 60, Error degrees of freedom: 56
Root Mean Squared Error: 1.16
R-squared: 0.961,  Adjusted R-Squared 0.959
F-statistic vs. constant model: 456, p-value = 2.59e-39
```

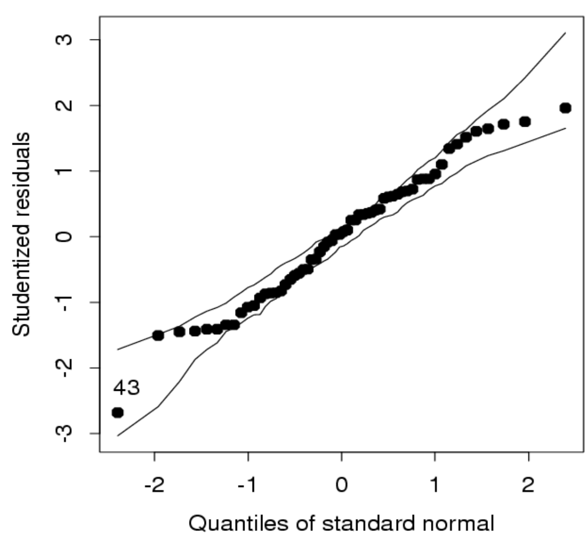
32

LS residuals against predicted values



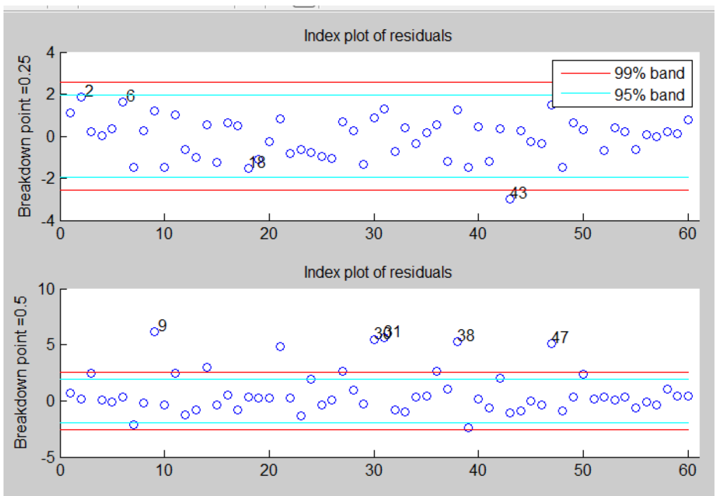
33

QQplot of studentized residuals



34

Analysis with robust S estimators:



bdp=0.25

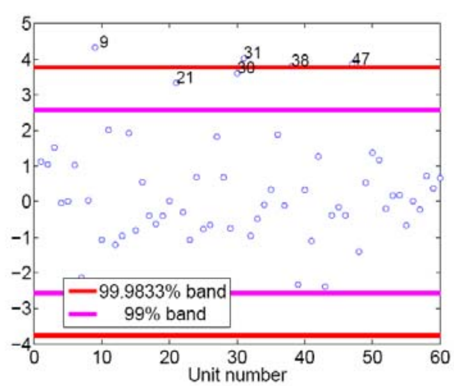
bdp=0.5

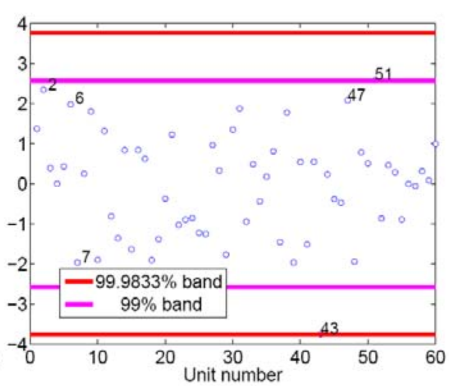
35

Analysis with robust estimators: MM

90% nominal efficiency

95% nominal efficiency





Individual and simultaneous confidence 99% bands

36

Traditional approach: compare robust and non robust fit

- Robust Inference as well as Classical Inference
- "... just which robust/resistant methods you use is not important – what is important is that you use some. It is perfectly proper to use both classical and robust/resistant methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard."
  – J. W. Tukey

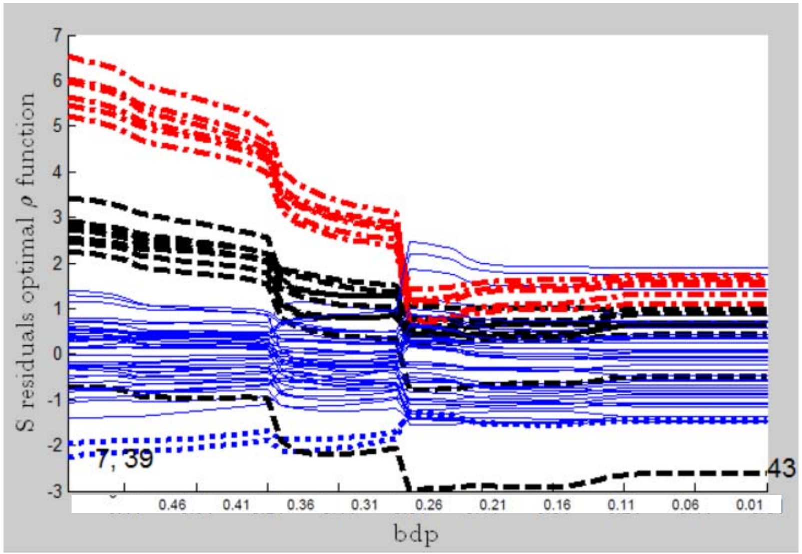
37

Consequences of the use of robust estimators

- Results obtained via a robust method are sometimes completely different
- Both in the use of traditional robust and non-robust statistical methods, researchers end up with a picture of the data.
- WHY NOT TO WATCH A FILM OF THE DATA ANALYSIS?

38

Monitoring of scaled S residuals



39

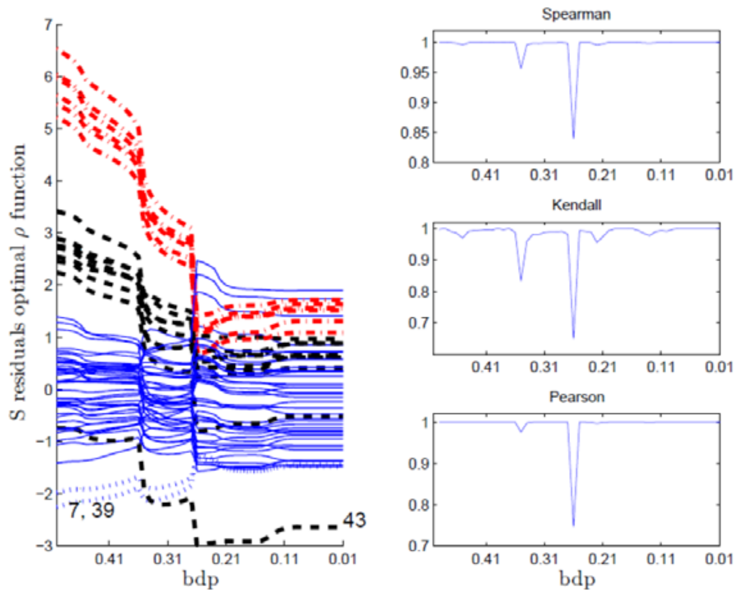
How to summarize changes in fit?

We consider three standard measures of correlation:

- Spearman. The correlations between the ranks of the two sets of observations.
- Kendall. Concordance of the pairs of ranks.
- Pearson. Product-moment correlation coefficient

40

Monitoring of scaled S residuals



41

EMPIRICAL BDP AND EFFICIENCY

Table: Empirical breakdown point and efficiency during monitoring for the transition between very robust and least squares regression: five estimators and four ρ functions. The values are for the step before the switch to a non-robust fit

Estimator		Bisquare	Optimal	Hyperbolic	Hampel
S	bdp	0.27	0.27	0.26	0.27
$\tau = 0.85$	bdp	0.38	0.40	0.41	0.41
$\tau = 0.90$	bdp	0.45	0.48	—[a]	0.50
$\tau = 0.95$	bdp	—[a]	—[a]	—[a]	—[a]
MM	effic.	0.91	0.97	0.90	0.87

[a] For these values of τ and ρ function, only non-robust solutions were obtained during monitoring.

42

More on this in

Electronic Journal of Statistics
Vol. 8 (2014) 646–677
ISSN: 1935-7524
DOI: 10.1214/14-EJS897

Monitoring robust regression

Marco Riani, Andrea Cerioli

Dipartimento di Economia, Università di Parma, Italy
e-mail: mriani@unipr.it; andrea.cerioli@unipr.it

Anthony C. Atkinson

The London School of Economics, London WC2A 2AE, UK
e-mail: a.c.atkinson@lse.ac.uk

and

Domenico Perrotta

43

Consequences of the use of robust estimators

- Even with monitoring the researcher loses the information that each unit, outlier or not, has on the final proposed estimate
- Purpose: to automatically order the data in terms of agreement to a fitted model
- Purpose: to find the effect that each unit has on your model
- Purpose: a tool that preserves the interpretative and computational simplicity of LS

44

Forward Search (FS) in Linear Regression

- STEP1: Start with very robust fit (LMS or LTS or S), then successively fit to larger subsets, found as those with the smallest residuals
- STEP2: Subset size increases until all the data are fitted. From LMS (LTS or S) to LS
- STEP3: MONITORING (scaled residuals, beta coefficients, …)

45

Step 2: Adding observations during the Forward Search

- Given *S*(m),* using *b*(m),* we compute the residuals for the *n* observations and select those which have the smallest squared *m*+1 residuals, *m=p*, *p*+1, …., *n*

Step 1

$$S^*(m) — b^*(m) \qquad e^2(m) = \begin{bmatrix} e_1^2 \\ e_2^2 \\ \cdots \\ e_n^2 \end{bmatrix}$$

m+1

Step 2

$$S^*(m+1) \longrightarrow b^*(m+1) \longrightarrow e^2(m+1) = \begin{bmatrix} e_1^2 \\ e_2^2 \\ \cdots \\ e_n^2 \end{bmatrix}$$

•This step is repeated up to when all units are included into the subset

……..

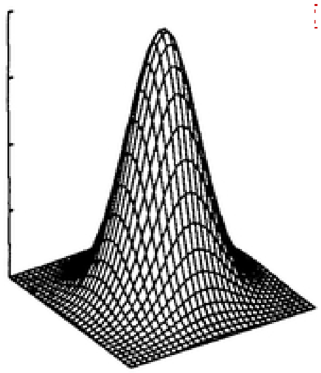
n

Step 3

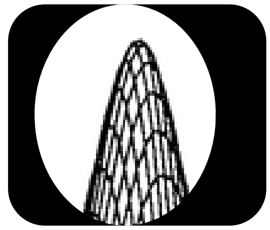
$$S^*(n) — b^*(n) — e^2(n) = \begin{bmatrix} e_1^2 \\ e_2^2 \\ \cdots \\ e_n^2 \end{bmatrix}$$

46

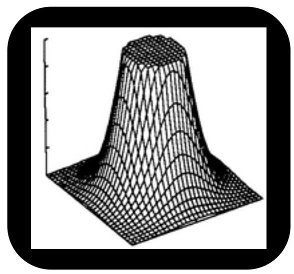
Characteristics of the FS



$m = p, p+1, p+2, \ldots, n$



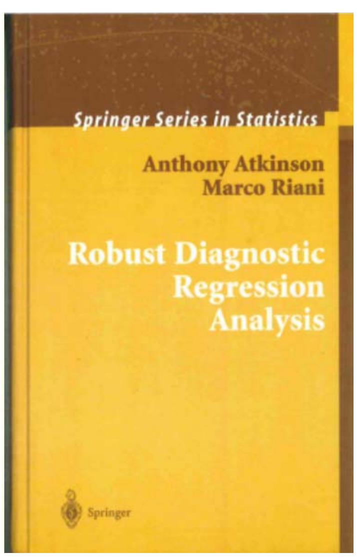
What is inside at step *m*

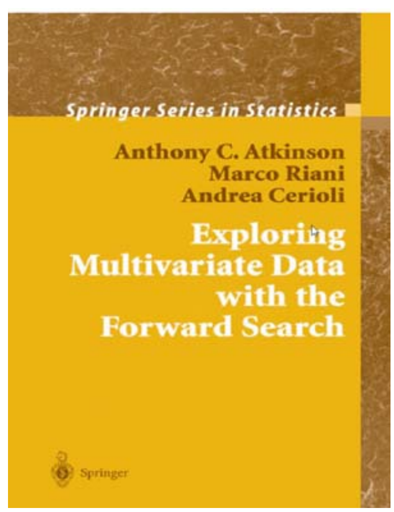


What is outside at step *m*

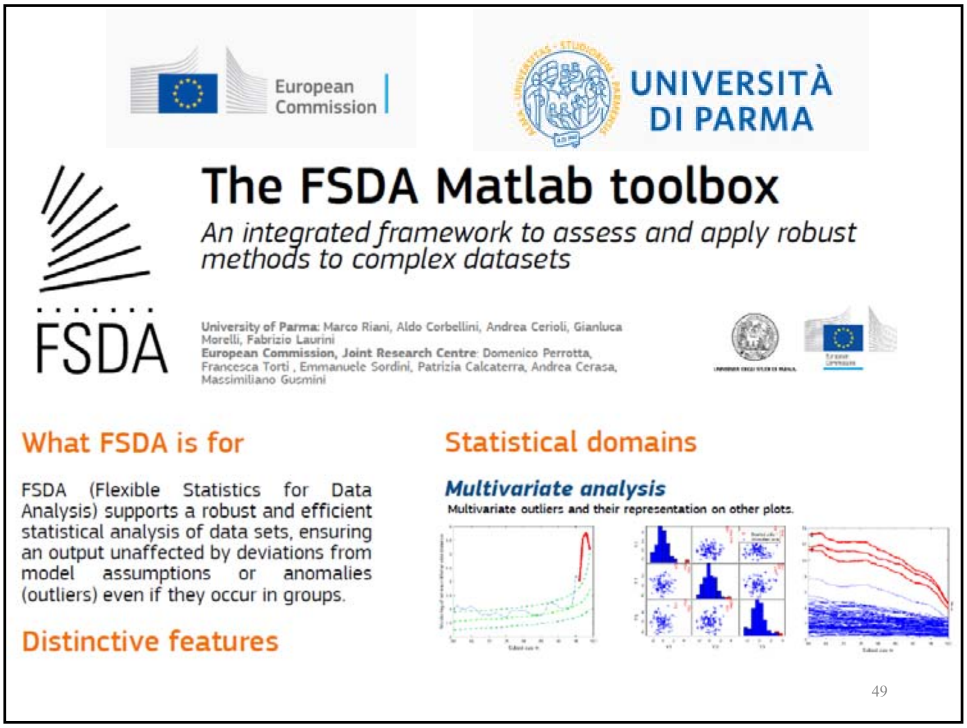
47

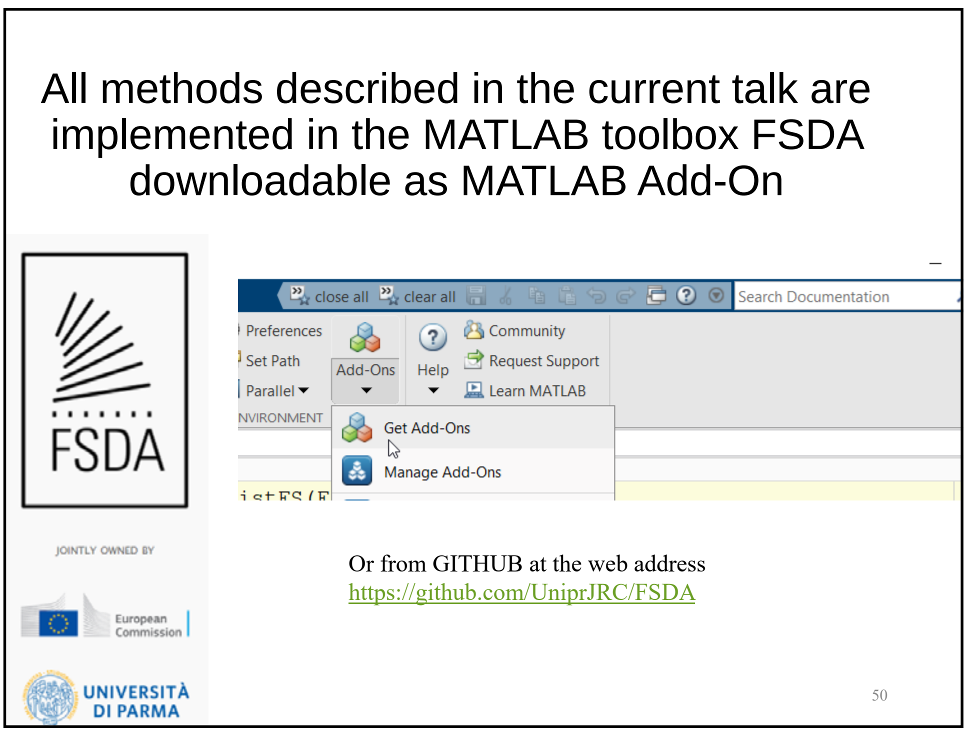
More on the forward search





48


European Commission
UNIVERSITÀ DI PARMA
The FSDA Matlab toolbox
An integrated framework to assess and apply robust methods to complex datasets
FSDA
University of Parma: Marco Riani, Aldo Corbellini, Andrea Cerioli, Gianluca Morelli, Fabrizio Laurini
European Commission, Joint Research Centre: Domenico Perrotta, Francesca Torti , Emmanuele Sordini, Patrizia Calcaterra, Andrea Cerasa, Massimiliano Gusmini
What FSDA is for
FSDA (Flexible Statistics for Data Analysis) supports a robust and efficient statistical analysis of data sets, ensuring an output unaffected by deviations from model assumptions or anomalies (outliers) even if they occur in groups.
Distinctive features
Statistical domains
Multivariate analysis
Multivariate outliers and their representation on other plots.
49


All methods described in the current talk are implemented in the MATLAB toolbox FSDA downloadable as MATLAB Add-On
FSDA
JOINTLY OWNED BY
European Commission
UNIVERSITÀ DI PARMA
close all
clear all
Search Documentation
Preferences
Set Path
Parallel
Add-Ons
Help
Community
Request Support
Learn MATLAB
NVIRONMENT
Get Add-Ons
Manage Add-Ons
Or from GITHUB at the web address
https://github.com/UniprJRC/FSDA
50

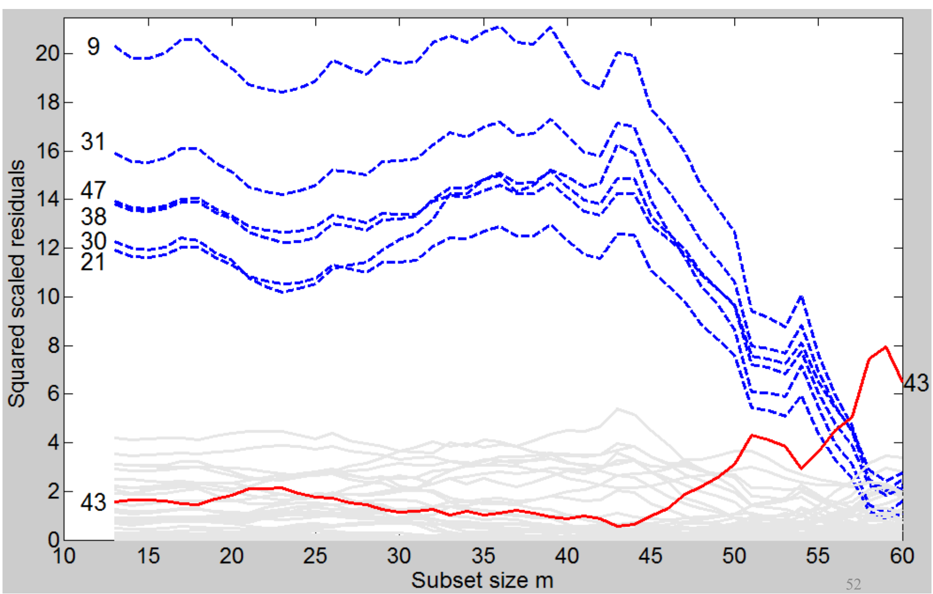
"New philosophy" of data analysis

- Our philosophy involves watching a film of data analysis rather than a snapshot.
- The crucial idea is to monitor how the fitted model changes as bdp decreases (S) or eff increases (MM) or, as in the "**forward search**", whenever a new statistical unit is added to the subset.

- The slides which follow show the analysis of the AR data using the forward search

51

Monitoring of scaled residuals



52

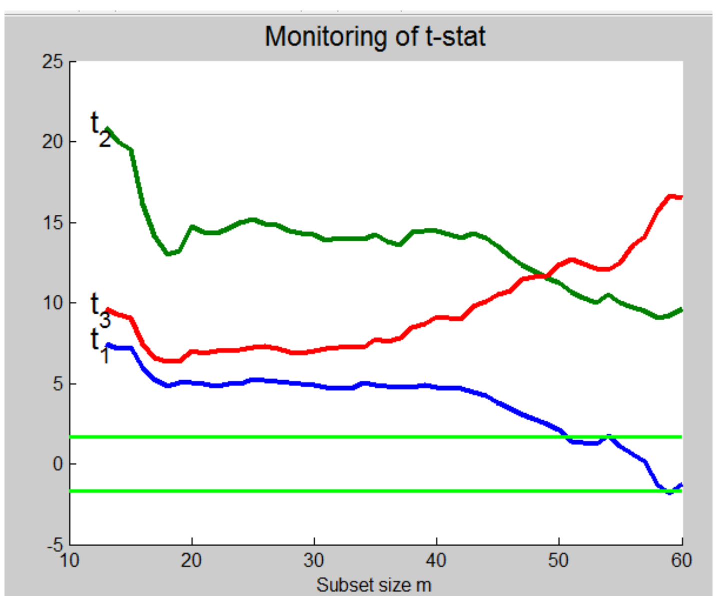
SIGNIFICANCE OF THE EXPLANATORY VARIABLES

Standard static approach

	All units	Without unit 43
t0	16.55	17.64
t1	-1.26	-1.93
t2	9.64	9.75
t3	16.53	17.66

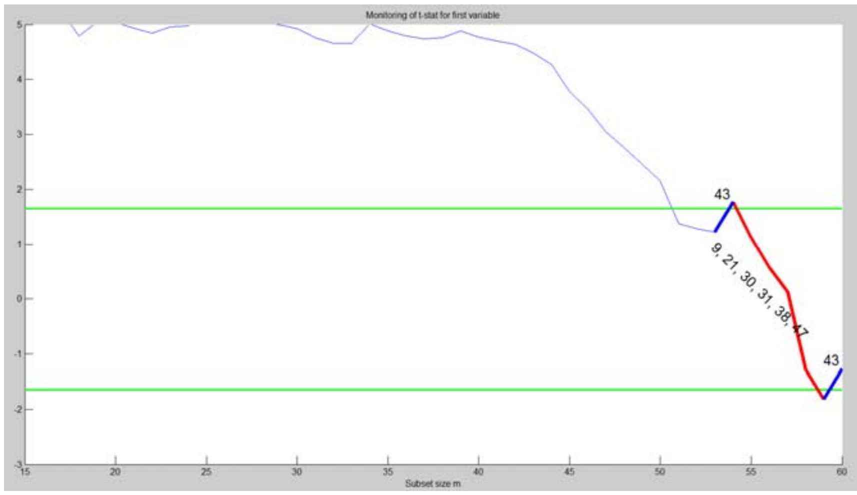
53

Monitoring of scaled *t*-statistics



54

Monitoring of scaled *t*-statistic for first variable

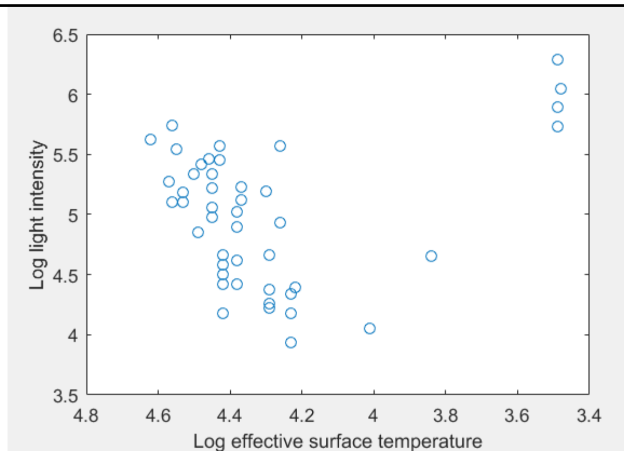


55

Robust regression estimators

- HARD TRIMMING
  - Least median of squares (LMS)
  - Least trimmed squares (LTS)
- SOFT TRIMMING
  - M
  - S
  - MM
  - Tau
- ADAPTIVE HARD TRIMMING
  - Forward search (FS)

56

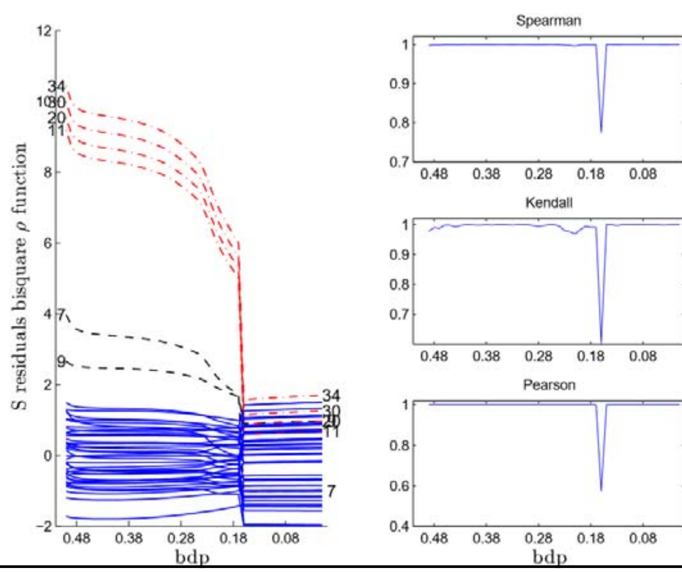


Stars data again

Analysis with the monitoring approach

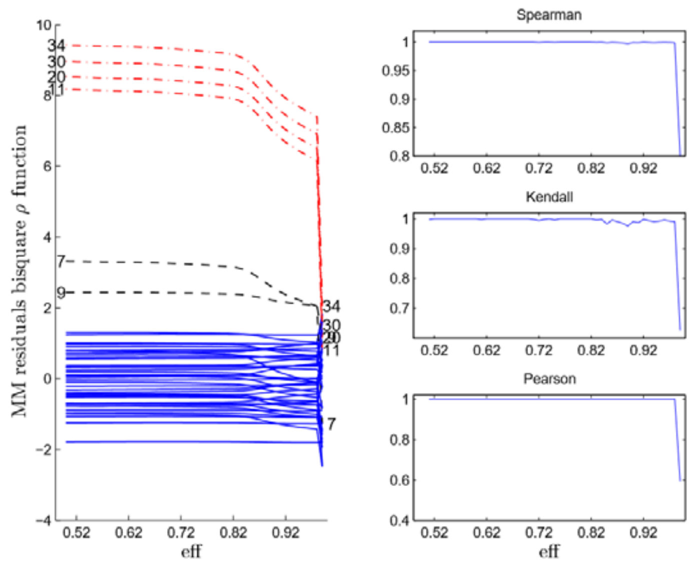
57

Monitoring of scaled S residuals



58

Monitoring of MM residuals



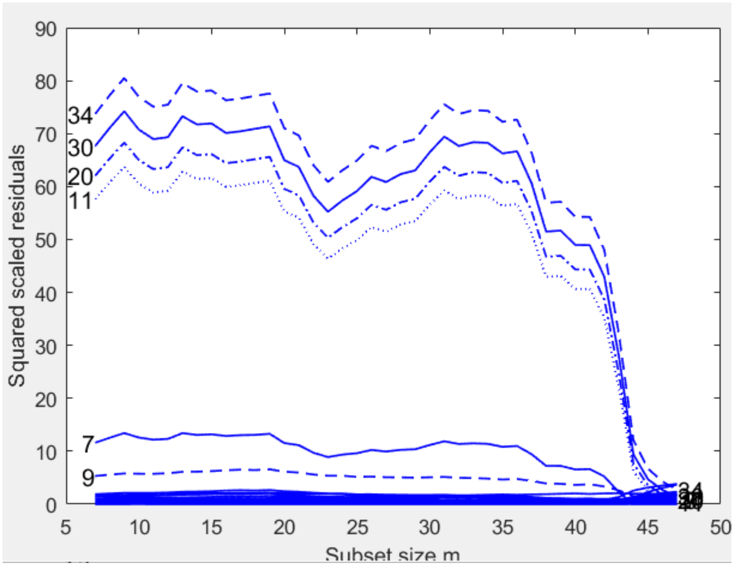
59

- Empirical breakdown point (bdp) or efficiency (eff) for S, MM and τ and four ρ functions. The values are for the step before the switch to a non robust fit

Estimator		Bisquare	Optimal	Hyperbolic	Hampel
S	bdp	0.17	0.17	0.17	0.17
$\tau = 0.85$	bdp	0.14	0.14	0.14	0.16
$\tau = 0.90$	bdp	0.17	0.16	0.16	0.20
$\tau = 0.95$	bdp	0.26	0.21	0.24	—[a]
MM	eff	0.98	0.99	0.97	0.96

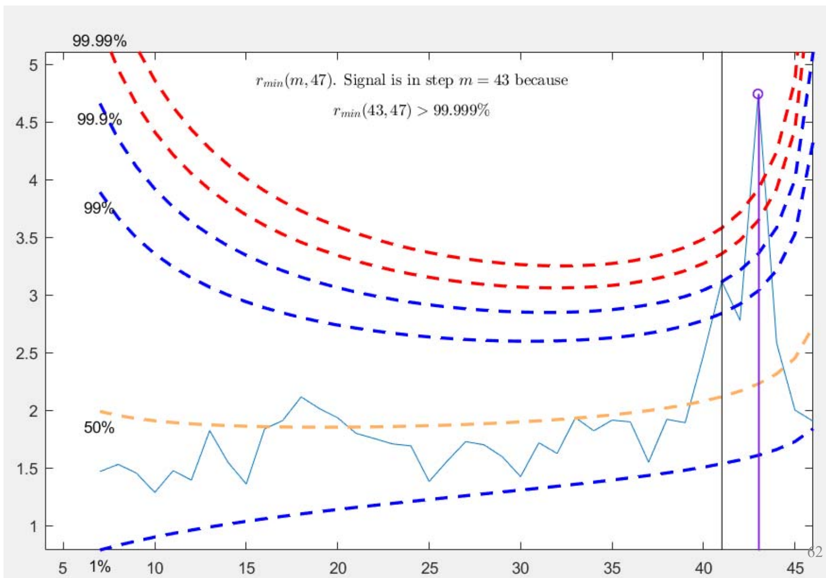
60

Monitoring of FS residuals



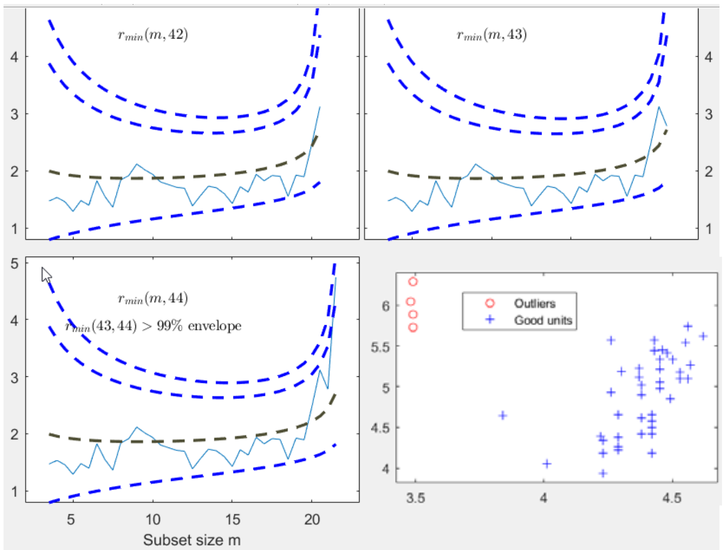
61

Stars data. automatic FS



62

Stars data: envelope resuperimposition



63

More about this in

Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2009)
71, *Part* 2, *pp.*

Finding an unknown number of multivariate outliers

Marco Riani,
Università di Parma, Italy

Anthony C. Atkinson
London School of Economics and Political Science, UK

and Andrea Cerioli
Università di Parma, Italy

64

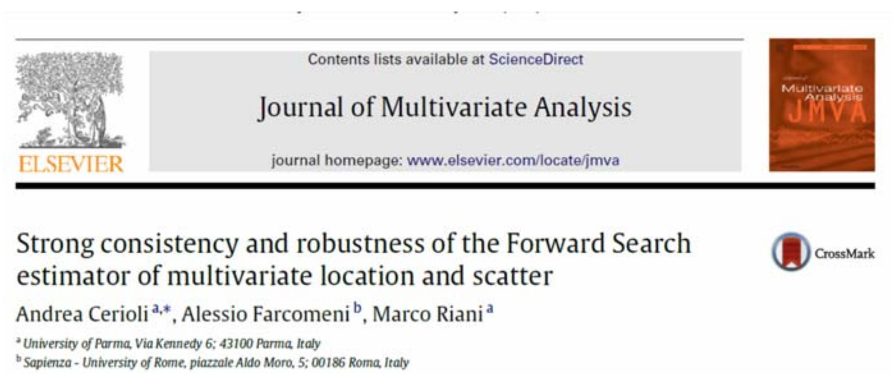
Theoretical properties of the FS

- Cerioli Farcomeni, and Riani (2014) show that the estimates obtained at step m and are **strongly consistent** under the null model and have **breakdown point** 1 − m/n under contamination: **the FS yields consistent high-breakdown estimators, but with adaptive breakdown point**

65

Theoretical properties of the FS

The FS yields consistent high-breakdown estimators, but with adaptive breakdown point


Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Strong consistency and robustness of the Forward Search estimator of multivariate location and scatter

Andrea Cerioli [a,*], Alessio Farcomeni [b], Marco Riani [a]

[a] University of Parma, Via Kennedy 6; 43100 Parma, Italy
[b] Sapienza - University of Rome, piazzale Aldo Moro, 5; 00186 Roma, Italy

66

Succesful applications of robust statistics

Ro.S.A. Robust Statistics Academy | UNIVERSITÀ DI PARMA

HOME ACTIVITIES RESEARCH FSDA TOOLBOX TEAM RESEARCH PARTNERS CONTACTS

- Dating of the Turin Shroud

Statistics and Computing
July 2013, Volume 23, Issue 4, pp 551–561 | Cite as

Regression analysis with partially labelled regressors: carbon dating of the Shroud of Turin

67

Fraud detection

Ro.S.A. Robust Statistics Academy | UNIVERSITÀ DI PARMA

HOME ACTIVITIES RESEARCH FSDA TOOLBOX TEAM RESEARCH PARTNERS CONTACTS

JOINTLY OWNED BY

European Commission

UNIVERSITÀ DI PARMA

THE WALL STREET JOURNAL.

WORLD

U.K. Faces $2.11 Billion EU Claim Over China Imports Fraud

Investigation comes as U.K. prepares for Brexit and the renewed debate over free trade

BBC Sign in News Sport Weather iPlayer Sounds

NEWS

Home UK World Business Politics Tech Science Health Family & Education

Politics Parliaments Brexit

EU demands 2.7bn euros of 'unpaid customs duty' from UK

68

Clustering


Journal
Journal of Computational and Graphical Statistics ›
Volume 27, 2018 - Issue 2

Enter keywords, authors, DOI, ORCID

184
Views
2
CrossRef citations to date

Data Mining
Finding the Number of Normal Groups in Model-Based Clustering via Constrained Likelihoods
Andrea Cerioli, Luis Angel García-Escudero, Agustín Mayo-Iscar & Marco Riani
Pages 404-416 | Received 01 Jun 2016, Accepted author version posted online: 16 Oct 2017, Published online: 20 Jun 2018

ELSEVIER
Pattern Recognition
Volume 88, April 2019, Pages 246-260

Efficient robust methods via monitoring for clustering and multivariate data analysis

Marco Riani, Anthony C. Atkinson, Andrea Cerioli, Aldo Corbellini
Show more
https://doi.org/10.1016/j.patcog.2018.11.016
Get rights and content

69

Summary and extensions

- Using the flexible approach provided by the forward search, together with monitoring, it is possible to combine robustness and efficiency
- The focus of this talk was on regression, but the approach described in these slides can be extended to
  - Generalized linear models
  - Transformations
  - Multivariate analysis and robust classification
  - Bayesian framework
  - Non linear models
  - Time Series

70

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

ESReDA | European Safety, Reliability & Data Association

A proposal of an algorithm to simulation of censored data rigth type I in reliability

Daniel Mendes Gaspar

Luis Andrade Ferreira

VALENCIA, 24th October 2019

1

Index

01. Introduction

02. Data analysis and type of data

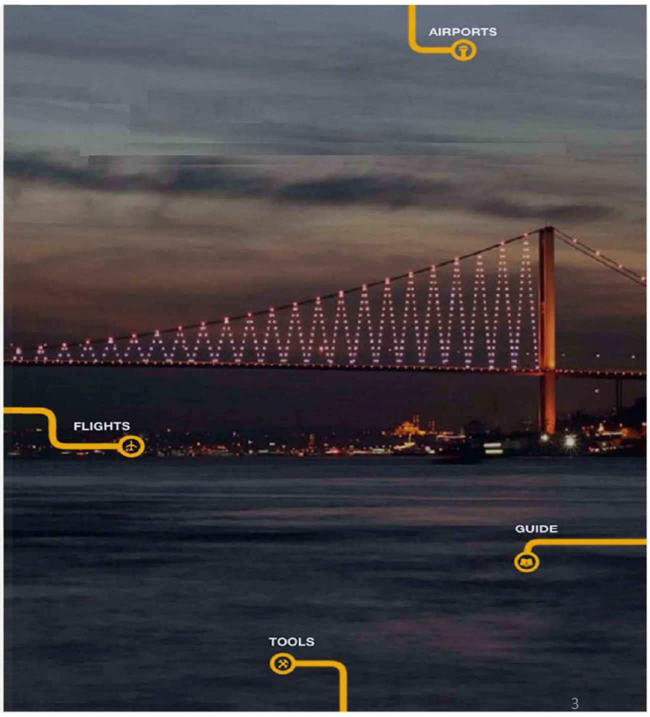
03. The algorithms

04 The Simulation results

05. Conclusions

2

01 Introduction



3

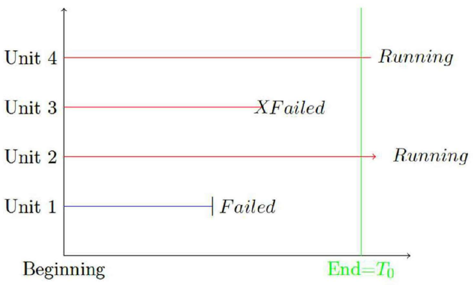
Introduction

Using Monte Carlo simulation on reliablity

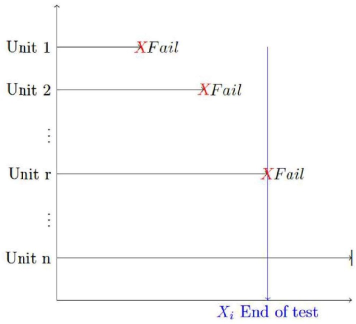
4

Introduction

The data censored (Appear on Reliability or medical studies, for example).



Censored Right type I



Censored Right type II

5

Introduction

- ❖ Estimation parameters to fit distribution

A number of models is developed need to be tested

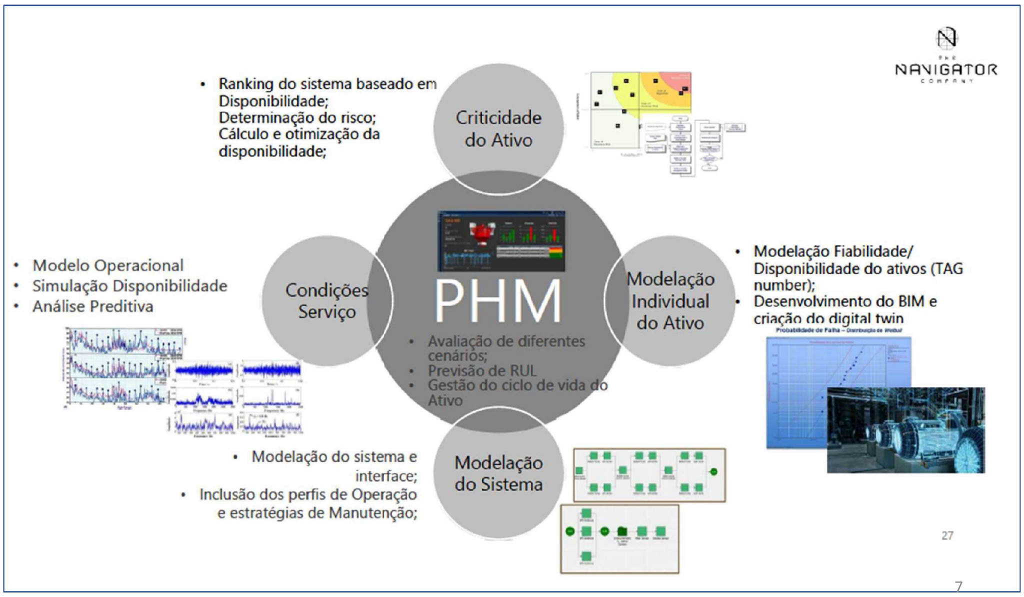
Right censored type I MLE Weibull $$L(\eta,\beta)=\prod_{i=1}^{n}\left\{\frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{\delta_i}\left\{\exp\left(\left(-\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{1-\delta_i}$$

- ❖ Optmizing simulation models (DES, ABM)

- ❖ Optmizing simulation models Predictive Analysis, ML, AI

6

The Navigator Company



Random number generating

The algorithms for simulating censored data are developed in a huge number of areas.

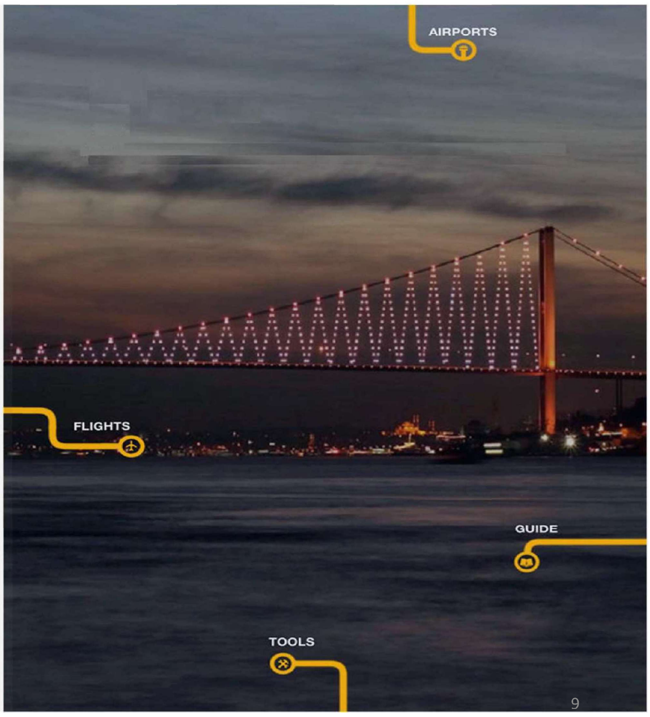
It's necessary to evaluate if the data that algorithms RNG produce can be considered randomly or not.

The distributions used are the most suitable in reliability and maintainability: Weibull, normal, gamma, log-normal and exponential.

8

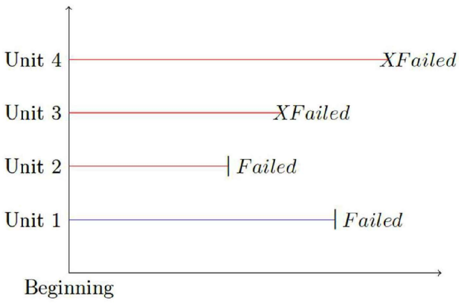
02

Type of data and data analysis



9

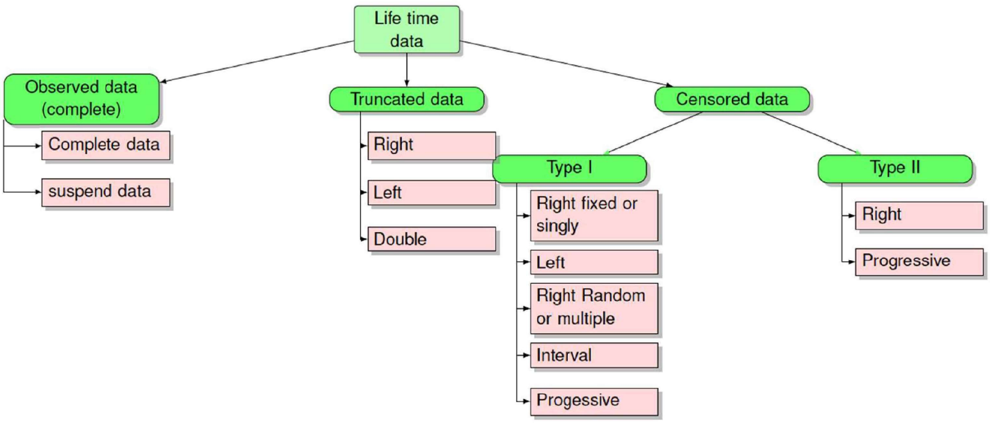
Type of data - reliability



Complete data

10

Type of data - reliability



11

Exponential and Gamma distribution

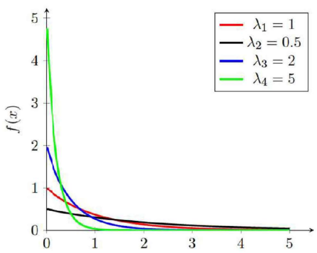


Figure 2.4: PDF for exponential distribution

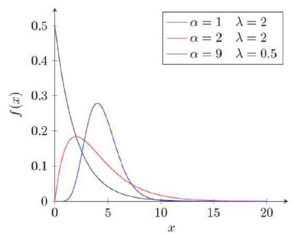


Figure 2.5: $f(x)$ - probability density function for gamma distribution

$$f(x|\lambda) = \begin{cases} 0 & (x \leq 0), \\ \lambda e^{-\lambda x} & (x > 0), \end{cases}$$

$$f(x, \alpha, \lambda) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\alpha x},$$

12

The statistical weibull distribution

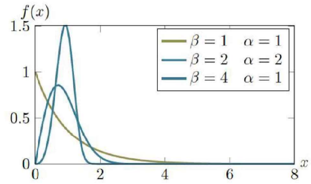


Figure 2.6: $f(x)$ - probability density function for Weibull distribution

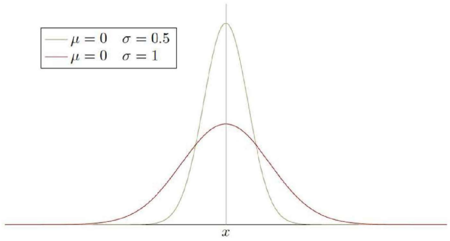


Figure 2.7: $f(x)$ - probability density function for normal distribution

$$f(x, \eta, \beta) = \frac{\beta}{\eta^{-\beta}} t^{\beta-1} e^{-(\frac{t}{\eta})^{\beta}}$$

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

General formulation of censored data

$$L(\theta; x, \delta) = \prod_{d\in D} f(x_d) \prod_{r\in R} S(x_r) \prod_{l\in L} S(x_r)[1 - S(x_l)] \prod_{i\in I} [S(U_i) - S(V_i)]$$

Where D is the set of death times, R is the set of right censored times, L is the set of left censored observations, I is the set of interval censored observations

Complete data MLE Weibull:

$$L(\eta, \beta) = \left(\frac{\beta}{\eta^{\beta}}\right)^n \prod_{i=1}^{n} \left[t_i^{\beta-1} \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right)\right]$$

The log-likelihood can be written as:

$$\ln L(\eta, \beta) = n \ln \beta - n\beta \ln \eta + (\beta - 1) \sum_{i=1}^{n} \ln t_i - \sum_{i=1}^{n} \left(\frac{t_i}{\eta}\right)^{\beta}$$

Right censored type I MLE Weibull

$$L(\eta,\beta)=\prod_{i=1}^{n}\left\{\frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{\delta_i}\left\{\exp\left(\left(-\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{1-\delta_i}$$

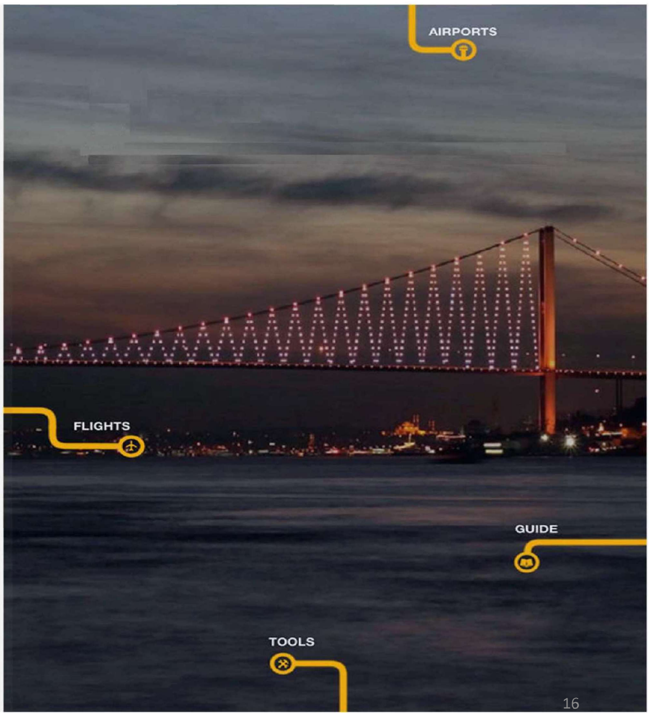
The log-likelihood can be written as:

$$\ln L(\eta,\beta)=r\ln\beta-r\beta\ln\eta+(\beta-1)\sum_{i=1}^{n}(\delta_i\ln t_i)-\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}$$

15

03

The algorithms



16

Proposed algorithm

Burton et al. (2006) proposed to generate a random non-informative right censoring with specific proportion of censored observations

Halabi and Singh (2004) in other way, provide formulas for determining parameters for standard survival and censoring distribution.

Gaspar and Ferreira (2019) proposed the calculated time of percentage of data censored.

17

The criteria for evaluating

The criteria for evaluating the change of parameters of each distribution results is obtained by three criteria:

The Mean $$m_u = \sum_{i=1}^{m} \frac{t_{ci}}{m}$$

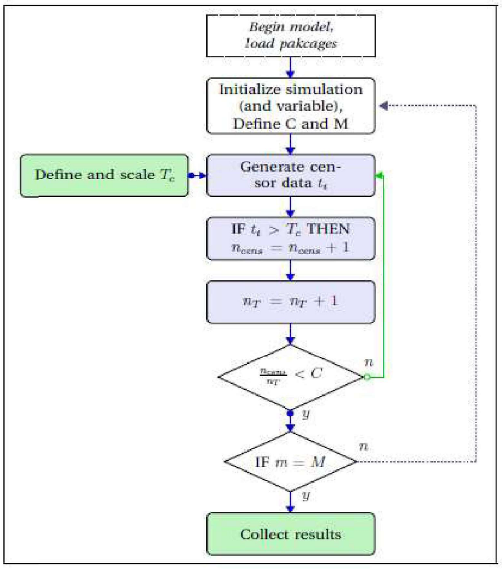
The Standard deviation $$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(t_{c_i} - \mu)^2}{m}}$$

The PE - percentage error $$PE_{T_u} = \xi = \frac{\left|t_c - T_{c(exact)}\right|}{T_{c(exact)}} x100$$

18

Algorithm 1

- The number of samples - n is not define and the time censoring is random (but scaled);
- The algorithm proceed to making a random scale of censorship time tc and then checking the number of censored data and the cycle/loop only follow to next step after the number of censored data is the same of value required;
- In this point the function used is the do while until get the target.
- This model can be used when don't know exactly the number of censor data required or the time of censoring it's not controlled.
- This algorithm have a great time-consuming and resources in computational point of view.



19

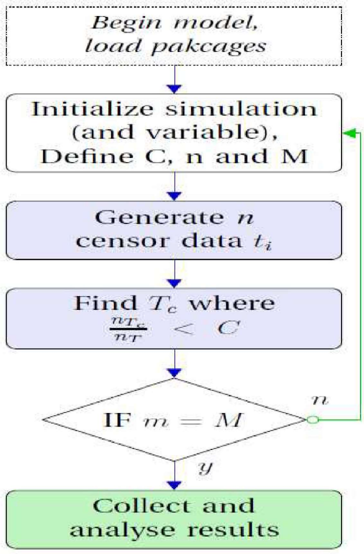
Algorithm 2

The second algorithm, define in the beginning all parameters except the censoring time t_c.

In this case, the censoring time is a result of the cycle and the random generation and their calculation depend of the %C censoring data.

The algorithm know exactly the value of %C, but can happens, sometimes not have enough random censor data or the inverse...

Is a simulation model that can be used in some application that the censoring times is not important and the time of computation must be optimize.



20

Algorithm 3

Begin to generate the vector data Y for ti of the distribution select and generate another vector *X* from binomial distribution with *(0,1)* a % C of zero's.

The model is a very practical solution, but sometimes not so accuracy.

It's need to define all parameters and in the initial step take out all the values ti that exceed tc... And some bias can be introduced in output of the model.

The solution needs to filter the values that exceed a certain time and for that reason it requires one more or two steps, and can take the algorithm no so efficiency.

Step 1: Chose parameters %C, n andM
Step 2: Define initial distribution parameters
Step 3: Chose random tc
Step 4: Generate the vector Y that represent ti and do generation until the all of n ti < tc
Step 5: Find tc that represent ntc/nt < C
Step 6: Generate the Binomial (0,1) vector X with random % C of zero's
Step 7: Multiply the two vectors and the O's represents the ti censored
Step 8: Collect and analyse results

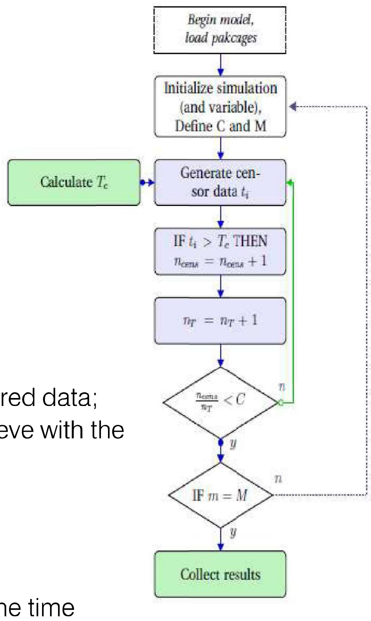
21

Algorithm 4 (the proposed algorithm)

This model optimize the simulation and give very good results.

It's necessary to calculate the time of censoring from each statistical distribution; generically is to do the inverse function of pdf function, calculate the time censoring tC and put this value in the algorithm of simulation with this value.

This procedure reduce the time consuming computation and with a large sample is very precise and comes closest to the percentage of censored data defined or theoretical.



All values random generates that fall in that region (x;+1) are the censored data; The relation between the reliability and the C% of censored data is achieve with the expression:

$$\mathbf{R(t)} = \int_{t_c}^{\infty} f(t)dt = C$$

In conclusion, define first the %C of censored data and then calculate the time tc to which the result have this probability.

22

An example (partial) of code to with weibull distribution, to the other distribution the code is reuse and adapt to calculate tC, the first part of the code in R software:

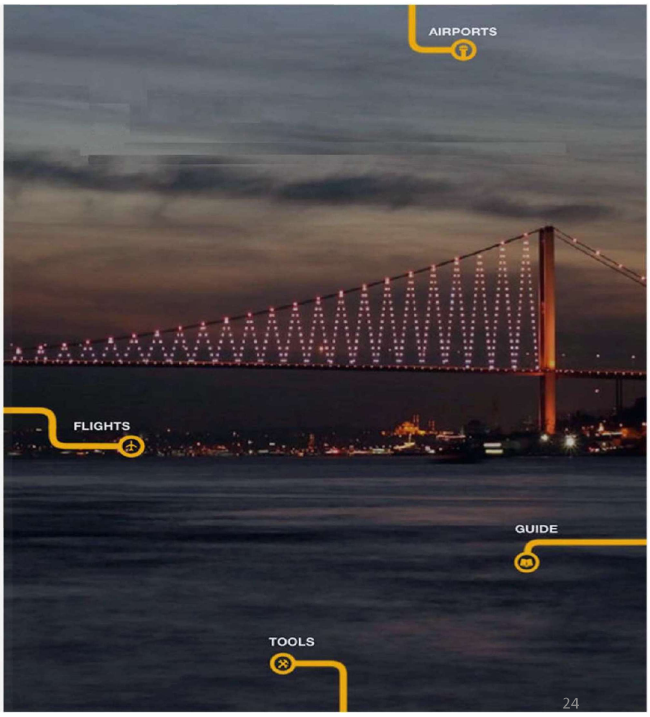
```
#Beginning of the cycle - repeat for different % data censored
for (i in 1:length(cen)){
   censoring_order <- n*(1-cen[i])# order in the vector censored
        for (j in 1:length(beta)){
                  #Execute m=100 cycles to obtain
                for (k in 1:m){
                        y <- rweibull(n,beta[j],wscale)
                        T_order <- sort(y) # sort by ascending
                        T_cens[k] <- T_order[censoring_order]
                }
        t <- wscale*(-log(cen[i]))^(1/beta[j]) # time censor
        #obtain the mean and the standar deviation of sample
        ysdt <- sd(T_cens) # standard deviation of sample
        ymeant <- mean(T_cens) # mean of the sample
        #obtain the mean and the standar deviation of sample
        wsdt[j] <- sd(T_cens) # standard deviation of sample
        wmean[j] <- mean(T_cens) # mean of the sample
        werf[j] <- abs(t-wmean[j])/t*100
        printgraf(n,wscale) #function to print the functions
        }
betadataframe <-cbind(wmean,wsdt,werf)
tablef <- cbind(tablef,betadataframe) # data frame to save
}
# final of the cycle!
```

Listing 1: Program in R from weibull distribution (partial)

23

04

Simulation Results



24

Weibull function

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	7.67	2.87	14.5	5.18	1.96	2.3	2.60	0.88	0.5	1.47	0.51	1.3
β_1	2.64	0.51	11.8	2.23	0.42	3.2	1.58	0.28	1.6	1.22	0.22	0.9
$\beta_{1.5}$	1.88	0.29	9.5	1.73	0.20	0.6	1.33	0.15	3.1	1.11	0.15	1.8
β_2	1.61	0.17	6.9	1.51	0.13	0.8	1.24	0.11	2.4	1.09	0.09	0.8
β_3	1.38	0.08	4.0	1.29	0.08	2.6	1.16	0.08	1.0	1.06	0.06	0.5
β_5	1.20	0.05	3.3	1.17	0.04	0.9	1.09	0.04	0.7	1.03	0.03	0.6

Table 1: The results from weibull distribution n=50

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	8.93	0.81	0.5	5.27	0.44	0.7	2.59	0.21	0.2	1.46	0.12	0.4
β_1	2.99	0.14	0.2	2.30	0.10	0.2	1.61	0.06	0.1	1.20	0.05	0.2
$\beta_{1.5}$	2.08	0.06	0.1	1.74	0.05	0.2	1.37	0.04	0.0	1.13	0.03	0.1
β_2	1.73	0.04	0.3	1.52	0.03	0.1	1.27	0.02	0.2	1.10	0.02	0.2
β_3	1.44	0.02	0.2	1.32	0.02	0.1	1.17	0.02	0.1	1.06	0.01	0.0
β_5	1.24	0.01	0.1	1.18	0.01	0.0	1.10	0.01	0.1	1.04	0.01	0.1

Table 2: The results from weibull distribution n=1000

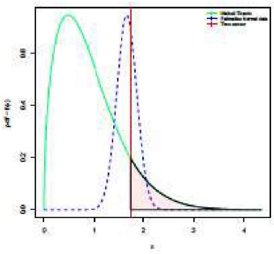


Figure 6: $\beta_{1.5}$,$C = 10\%$ and n=50

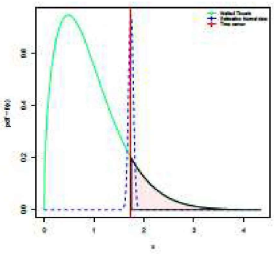


Figure 7: $\beta_{1.5}$, $C = 10\%$ and n=1000

25

Gamma function

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\alpha_{0.5}$	1.61	0.41	16.1	1.29	0.32	5.0	0.81	0.19	1.6	0.53	0.16	1.4
α_1	2.56	0.47	14.6	2.18	0.41	5.5	1.56	0.26	2.8	1.19	0.22	1.4
$\alpha_{1.5}$	3.55	0.58	9.1	3.01	0.44	3.7	2.23	0.34	4.0	1.82	0.26	0.5
α_2	4.27	0.61	10.0	3.79	0.54	2.5	2.88	0.37	3.8	2.43	0.29	0.5
α_3	5.83	0.74	7.4	5.25	0.64	1.3	4.24	0.41	1.0	3.64	0.40	0.8
α_5	8.64	0.88	5.6	7.88	0.62	1.5	6.67	0.51	0.7	5.82	0.50	1.1

Table 3: The results from gamma distribution - n=50

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\alpha_{0.5}$	1.91	0.11	0.4	1.35	0.08	0.3	0.82	0.05	0.2	0.54	0.03	0.3
α_1	3.00	0.14	0.0	2.30	0.09	0.0	1.60	0.06	0.3	1.20	0.05	0.2
$\alpha_{1.5}$	3.90	0.16	0.1	3.12	0.11	0.3	2.32	0.08	0.1	1.83	0.06	0.3
α_2	4.74	0.16	0.1	3.88	0.12	0.2	2.99	0.08	0.1	2.43	0.07	0.2
α_3	6.28	0.19	0.3	5.32	0.14	0.1	4.28	0.10	0.0	3.61	0.08	0.0
α_5	9.14	0.22	0.1	7.98	0.16	0.1	6.72	0.12	0.1	5.88	0.11	0.1

Table 4: The results from gamma distribution n=100

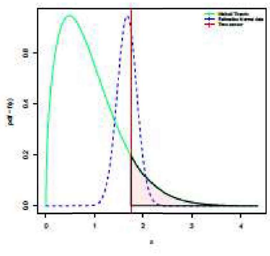


Figure 6: $\beta_{1.5}$,$C = 10\%$ and n=50

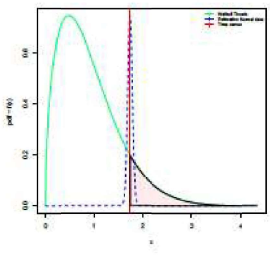


Figure 7: $\beta_{1.5}$, $C = 10\%$ and n=1000

27 | 45

26

Log-Normal function

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.3}$	4.23	0.34	5.0	3.98	0.29	0.4	3.47	0.22	0.9	3.17	0.17	0.4
$\sigma_{0.5}$	5.70	0.80	7.8	5.04	0.57	2.3	4.08	0.38	1.4	3.50	0.32	1.0
$\sigma_{0.7}$	8.01	1.41	6.8	6.54	0.97	1.9	4.78	0.77	2.4	3.97	0.49	1.2
σ_1	11.91	2.93	15.4	9.28	2.06	5.2	6.16	1.30	2.3	4.56	0.85	0.7
$\sigma_{1.3}$	18.86	7.16	18.2	13.69	4.11	4.8	7.72	2.09	4.9	5.29	1.28	1.6
$\sigma_{1.7}$	32.92	14.02	26.1	22.05	8.68	8.2	10.70	3.71	5.9	7.05	2.06	6.3

Table 5: The results from Log-Normal distribution n=50

	5			10			20			30		
	μ	σ	ξ	μ	σ	21 ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.3}$	4.45	0.09	0.1	3.99	0.06	0.0	3.50	0.05	0.0	3.18	0.04	0.0
$\sigma_{0.5}$	6.17	0.21	0.3	5.15	0.15	0.1	4.14	0.09	0.0	3.53	0.07	0.0
$\sigma_{0.7}$	8.58	0.41	0.2	6.67	0.24	0.0	4.90	0.15	0.1	3.92	0.11	0.1
σ_1	14.06	0.98	0.1	9.77	0.52	0.3	6.30	0.28	0.2	4.58	0.20	0.2
$\sigma_{1.3}$	22.98	2.02	0.4	14.39	1.00	0.0	8.10	0.47	0.2	5.37	0.29	0.0
$\sigma_{1.7}$	44.16	4.94	0.8	23.93	2.16	0.4	11.43	0.89	0.5	6.61	0.47	0.3

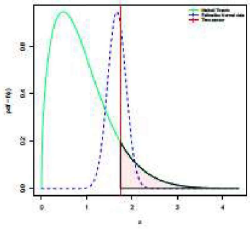


Figure 6: $\beta_{1.5}$,$C = 10\%$ and n=50

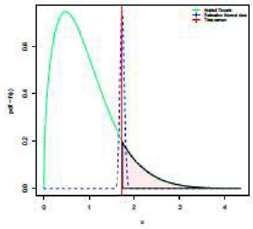


Figure 7: $\beta_{1.5}$, $C = 10\%$ and n=1000

Exponential function

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\lambda_{0.5}$	5.44	0.90	9.2	4.40	0.87	4.5	3.17	0.53	1.4	2.39	0.46	0.6
λ_1	2.74	0.55	8.7	2.25	0.43	2.5	1.59	0.30	1.2	1.16	0.20	3.4
$\lambda_{1.5}$	1.81	0.35	9.1	1.48	0.27	3.8	1.03	0.17	3.6	0.81	0.13	0.5
λ_2	1.36	0.25	9.0	1.13	0.20	1.9	0.76	0.13	5.2	0.60	0.11	0.4
λ_3	0.89	0.14	10.9	0.74	0.14	3.3	0.52	0.09	2.9	0.40	0.08	1.2
λ_5	0.53	0.09	11.8	0.44	0.10	4.2	0.32	0.05	1.7	0.24	0.04	1.3

Table 7: The results from Exponential distribution n=50

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\lambda_{0.5}$	5.96	0.27	0.6	4.59	0.20	0.3	3.21	0.13	0.3	2.40	0.10	0.2
λ_1	2.99	0.13	0.2	2.30	0.09	0.0	1.61	0.06	0.1	1.20	0.05	0.2
$\lambda_{1.5}$	1.99	0.09	0.2	1.53	0.06	0.0	1.07	0.04	0.2	0.80	0.03	0.1
λ_2	1.49	0.07	0.3	1.15	0.05	0.4	0.81	0.03	0.1	0.60	0.02	0.1
λ_3	1.00	0.05	0.1	0.77	0.03	0.3	0.54	0.02	0.1	0.40	0.02	0.1
λ_5	0.60	0.03	0.1	0.46	0.02	0.1	0.32	0.01	0.2	0.24	0.01	0.0

Table 8: The results from Exponential distribution n=1000

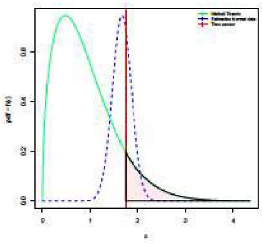


Figure 6: $\beta_{1.5}$,$C = 10\%$ and n=50

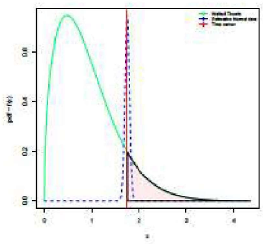


Figure 7: $\beta_{1.5}$, $C = 10\%$ and n=1000

R(t)= $\int_{t_c}^{\infty} f(t)dt = C$

Normal function

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.5}$	1.74	0.14	4.7	1.61	0.12	2.2	1.40	0.11	1.7	1.24	0.10	1.6
σ_1	2.46	0.26	7.0	2.19	0.24	3.8	1.82	0.21	1.3	1.49	0.20	2.4
$\sigma_{1.5}$	3.14	0.41	9.5	2.77	0.35	5.1	2.21	0.29	2.5	1.69	0.27	5.5
σ_2	3.85	0.50	10.3	3.48	0.49	2.3	2.64	0.40	1.8	1.90	0.31	7.2
σ_3	5.34	0.72	10.0	4.60	0.69	5.0	3.42	0.59	3.1	2.52	0.53	2.0
σ_5	8.55	1.40	7.3	7.22	1.03	2.5	5.09	1.04	2.2	3.56	0.95	1.8

Table 9: The results from normal distribution n=50

	5			10			20			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.5}$	1.82	0.03	0.1	1.64	0.03	0.1	1.42	0.02	0.1	1.26	0.02	0.1
σ_1	2.64	0.07	0.3	2.28	0.05	0.2	1.84	0.05	0.1	1.52	0.04	0.1
$\sigma_{1.5}$	3.46	0.10	0.2	2.92	0.08	0.1	2.26	0.07	0.1	1.79	0.06	0.1
σ_2	4.28	0.13	0.1	3.56	0.11	0.1	2.68	0.09	0.1	2.04	0.09	0.3
σ_3	5.92	0.20	0.3	4.84	0.17	0.0	3.52	0.14	0.1	2.57	0.12	0.2
σ_5	9.20	0.34	0.2	7.39	0.27	0.3	5.20	0.22	0.2	3.61	0.21	0.3

Table 10: The results from normal distribution - n= 1000

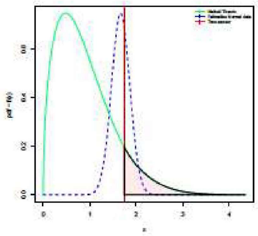


Figure 6: $\beta_{1.5}$,$C = 10\%$ and n=50

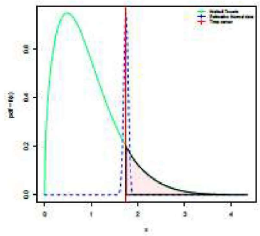


Figure 7: $\beta_{1.5}$, $C = 10\%$ and n=1000

29

Results

The results of the weibull distribution are very interesting. With a sample of n = 50 and a cycle of simulations M = 1000, can see that the standard deviation is higher when = 0:5 and then slowly decrease until = 5. The percentage error it's much higher when the C=5% and then goes down when the shape factor increase.

There is a different behavior in C=20%, in this case the standard deviation and the PE - percentage error is higher when = 1:5, this could have some explanation because the transition of shape from exponential to normal shape. The simulation to a sampling number of n = 1000 we don't have the same behavior, probably could be some phenomena with the random generation number in this particular parameters of distribution.

The table 2 show an simulation to a sample of 1000 and in this case the standard deviation and PE is smaller than in the case of that sample are 50 from table 1..

30

Results

To all simulation it can be noted that the error is less than 1% , which is very small and even the dispersion itself is very small,

as can be seen by the tables and in figures (in annex). There is a small decrease in dispersion and error as the shape factor increases. Finally,

These algorithm can be used but it's need to have caution and choose a higher number of sampling in order to give more accuracy to the simulation study!

Graphically can be see the dispersion, the bias and the mean of simulation and compare theoretical curve with the curve from a normal estimation data and the value of tc calculated

There is an exhaust comparison figure for all statistical distribution we study and for different values of sampling, shape factor and %C percentage of data censored.

31

Results

The results of the gamma distribution are quite similar to the weibull function.

The Gamma distribution variation of the shape factor don't have so influenced like in weibull distribution. In this case, the values is very similar to the standard deviation from sampling 50 and sampling 1000. The standard deviation is sightly increase when the shape factor increase and the values is small.

The percentage error are very different from sampling 50 and sampling 1000. In sampling 50 the PE is very high and in sampling 100 is small and quite stable.

The algorithm that we develop to gamma distribution are very suitable, but it's need to have caution and choose a higher number of sampling in order to give more accuracy to the simulation study!.

32

Results

The Log-normal distribution have an behavior very interesting. When we increase the standard deviation the Percentage error and the standard deviation increase. It's logic that if we increase and growth the dispersion the result in M simulations will growth too.

There is some difference in the values from a simulation that have an sample of n=50 and a sample of 1000. The value is much higher and the bias and Percentage error is much higher in the sample of 50.

In the case of the log-normal distribution we conclude that we need to have care in use these algorithm, but probably it's in all kind of algorithms using log-normal distribution. It's depend a lot from the shape and probably the scale parameter.

A simulation to a sample of 1000 and in this case the standard deviation and PE is smaller than in the case of that sample are 50..

33

Results

To exponential distribution performed a simulation with changing the hazard rate from 0.5 to 5. The standard deviation and the PE is much higher in the simulations with sample 50.

The exponential distribution is stable and more robust when the sampling is higher. The bias is reduced when the %C and the hazard rate increase.

The simulations with normal distribution use the mean = 1 and the standard deviation range from 0.5 to 5. Comparing the two table is quite easy to conclude that the bias and the error is much higher in the simulations with sample 50.

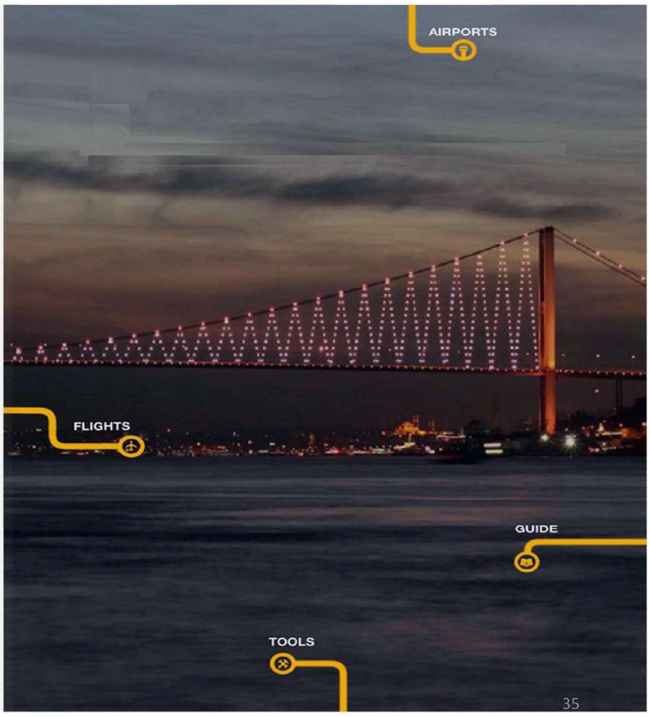
The normal distribution, when the sampling is higher have their values consistent, low bias and reliable results.

27 | 45

34

05

Conclusions



35

The reliability function

The survival testing and reliability studies are normally focus in estimate an unknown cumulative distribution function (cdf).

In simulation Studies we use the computational power to test a particular hypotheses and asses the validity and accuracy of a variety of statistical methods or procedures in relation to a known truth.

These procedures and algorithms provide empirical estimation of sampling distribution of the parameters of interest

In this study we are trying to help the development of the best procedures to generate a sample of data with a particular characteristic (data right censored and type I) and need to be random and i.i.d to used to simulation
in reliability field.

36

The reliability function

For this study, under the physically motivated assumption that the distribution of the generalized deviations does not depend on changes in certain parameters (e.g the scale parameter in the distribution in study).

Based on the experiences and intuitions we observed that a value of σ in the neighbourhood and below to 1 tends to make the distribution of the deviation close to *iid* $N(\mu,\sigma^2)$ over a wide range of testing conditions.:

Under the general model the probability of a sample have the number of censored data (or above) is:.

$$\text{POD} = \Pr(T(t_c) \geq t_{c_{exact}})$$
$$= \Pr[f(X(t), \tilde{x}(t_c)) \geq f(t_c, \tilde{x}(t))]$$
$$= \Phi_{nor}\left[\frac{f(X(t), \tilde{x}(t_c)) - \tilde{\mu}_t}{\tilde{\sigma}}\right]$$

Where Φ_{nor} is the standard normal cdf and $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ are estimates from the generalized simulation and deviation data..

37

The reliability function

The uncertainty of the estimation of the POD curve for a range of different shapes for different %C percentage of data censored are calculated

Compare four distribution: - exponential, weibull, gamma and log-normal, in which we can see the different behaviour when we change the scale parameters and have the same %C percentage of censored data.

The exponential distribution all the %C percentage of censored data have the same trend except the 10% line that
in the λ_1 and $\lambda_{1.5}$ have an erratic movement and 0% percent no difference between the exact value and the estimate.

The weibull distribution have the some smooth decrease trend to all different %C.

In gamma distribution the trend is more significant than weibull, but is increase, e.g. if the shape is more higher the error is lesser.

The log-normal have a very good results except the 5% and σ_1 and $\lambda_{1.5}$.

38

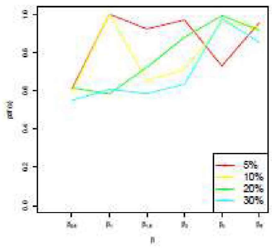


Figure 10: Gamma desviation changing β and n=1000

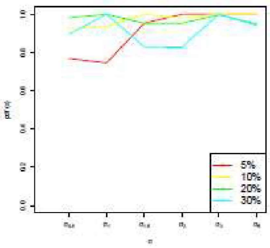


Figure 11: log-normal desviation changing σ and n=1000

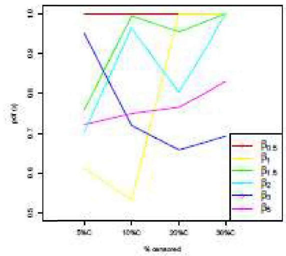


Figure 12: Gamma desviation changing %C and n=1000

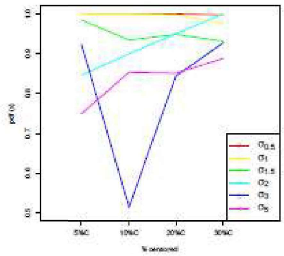


Figure 13: log-normal desviation changing %C and n=1000

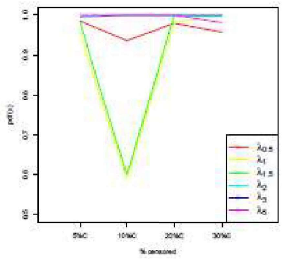


Figure 14: Exponential desviation changing %C and n=1000

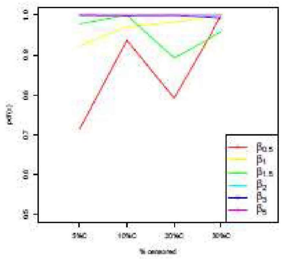


Figure 15: Weibull desviation changing %C and n=1000

Conclusions

_This research work was used to develop reliability model simulation algorithms for complex equipment/systems, when the data collection is faced with censored data.

_The algorithms are innovative and their development was done in three different software: Python, Matlab and R.

_A methodology of analysis (hypothesis tests) and validation with an evaluation matrix is proposed to test the i.i.d. data of RNG of censored data.

_The majority of simulation studies reported in the literature are not providing sufficient details of simulation random generator data.

To final conclusion, we hope our work permit to generate some more discussion and attention to the algorithms that simulate data censored, and give some tools and results to made the simulations and the studies more accuracy and optimized..

41

_There are influence of parameters of distribution or the parameters of model simulations in the randomness of data generation.

_Modifications of the simulation process, such as altering the number of simulations or other parameters are possible, but this modifications can be a time-consuming processing.

_The hypothesis test to apply to the specific generator censored data must be selected very carefully in order to have good results and to optimize the simulation time.

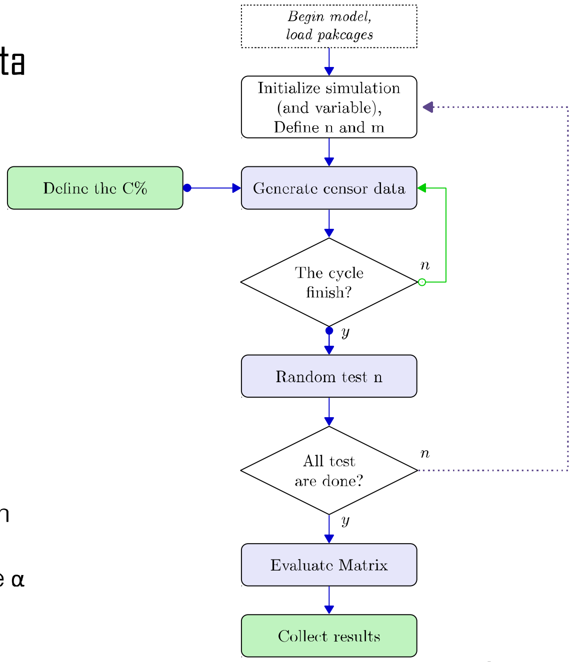
42

Flow chart for simulation of censored data

To illustrate and compare the methods as described above, a random sample of size, n=10, 100, 500, 1000, 10000 to take care of small, medium and large data sets.

The scale parameter was chosen to be 1 and the shape parameter have five values (0.5, 1, 1.5, 3, 5) and M=1000 (cycles of simulation).

The simulation cycle complete for each distribution doing to each percentages of censored data: 5%, 10%, 20% and 30% and for the level of confidence α = 0.05 and α = 0.1.



Test for randomness

Kolmogorov-Smirnov Test is the two-level test is to compare the empirical distribution of these U_j's to the uniform distribution, via a goodness-of-fit (GOF)

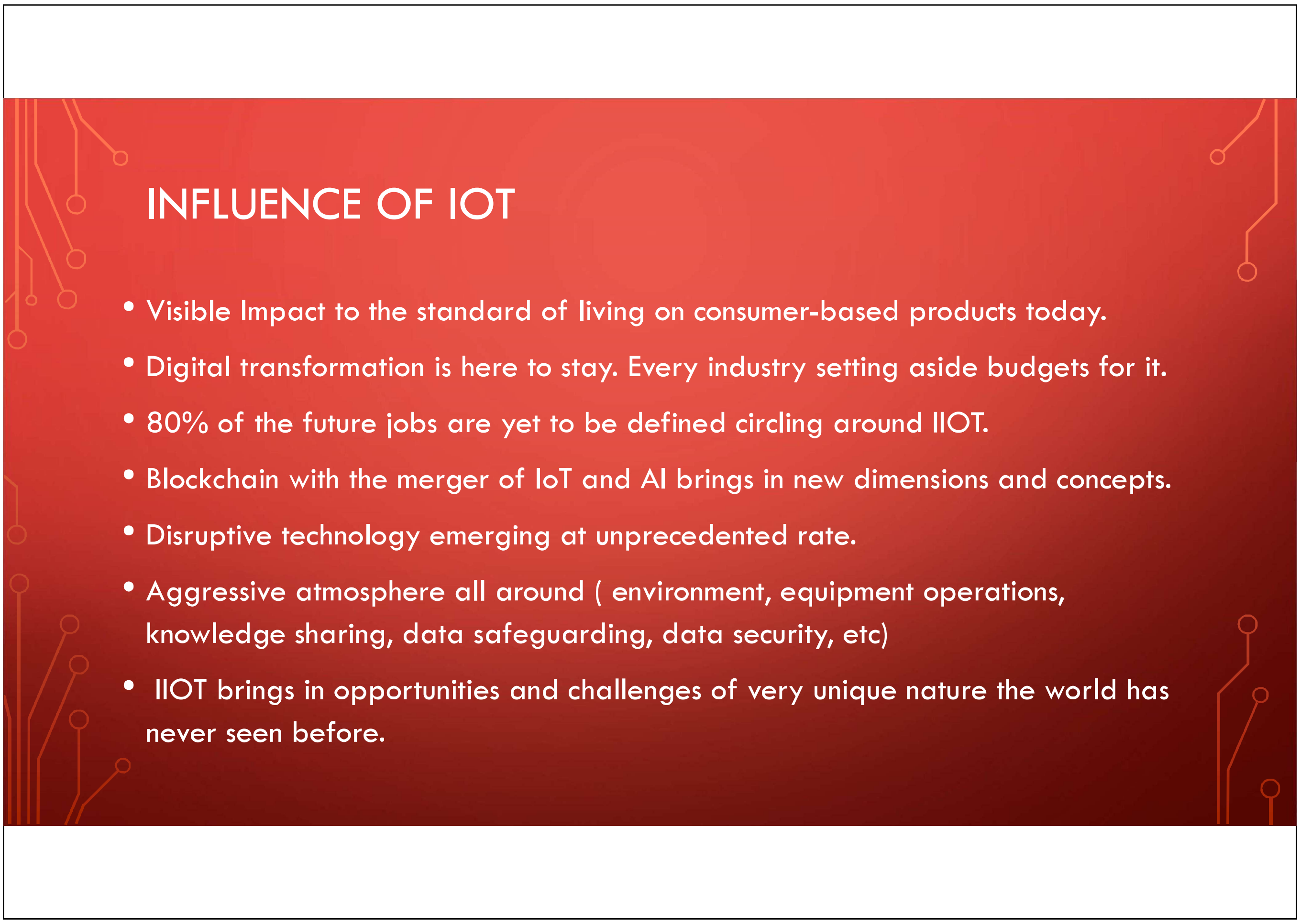
The **Wald–Wolfowitz** runs test (or simply runs test), is a non-parametric statistical test that can be used to test the hypothesis that the elements of the sequence are mutually independent

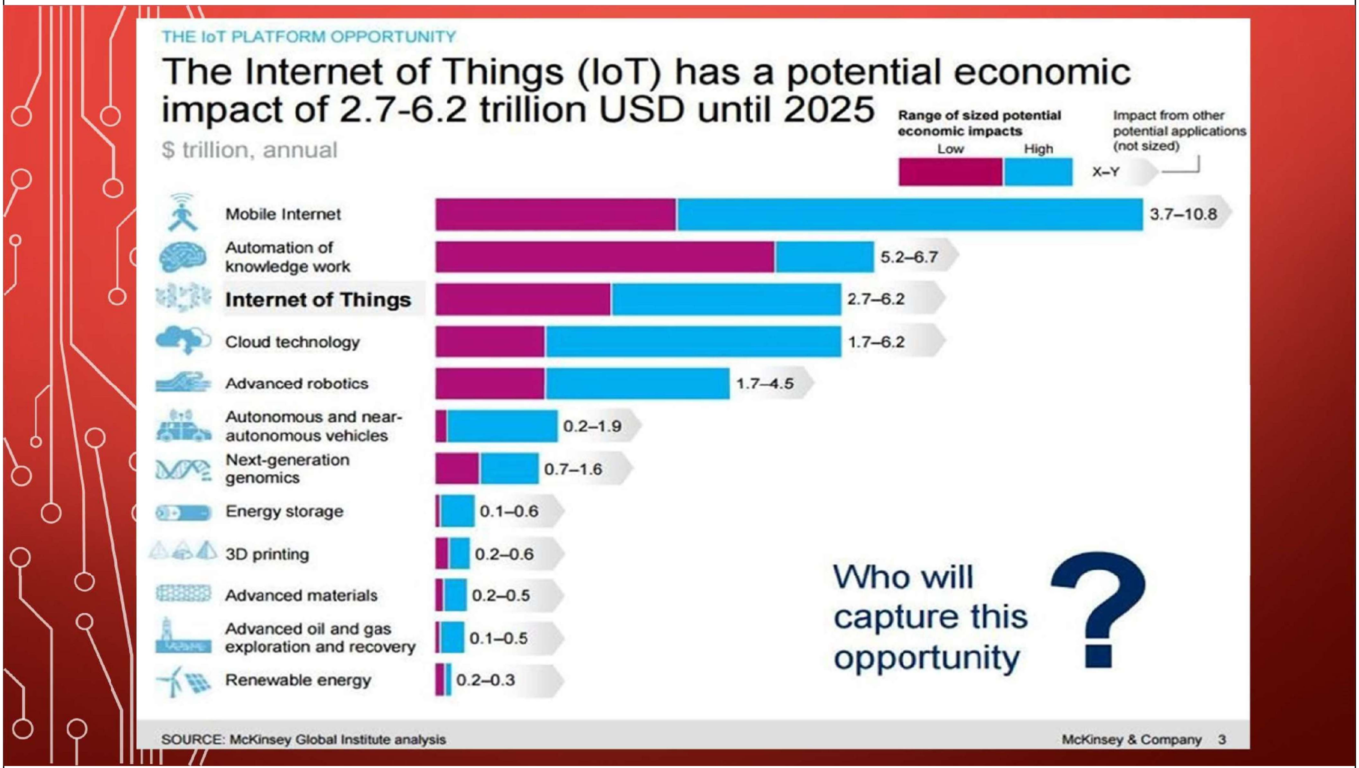
The non-parametric **Mann-Kendall test** is commonly employed to detect monotonic trends in series of random or reliability data.

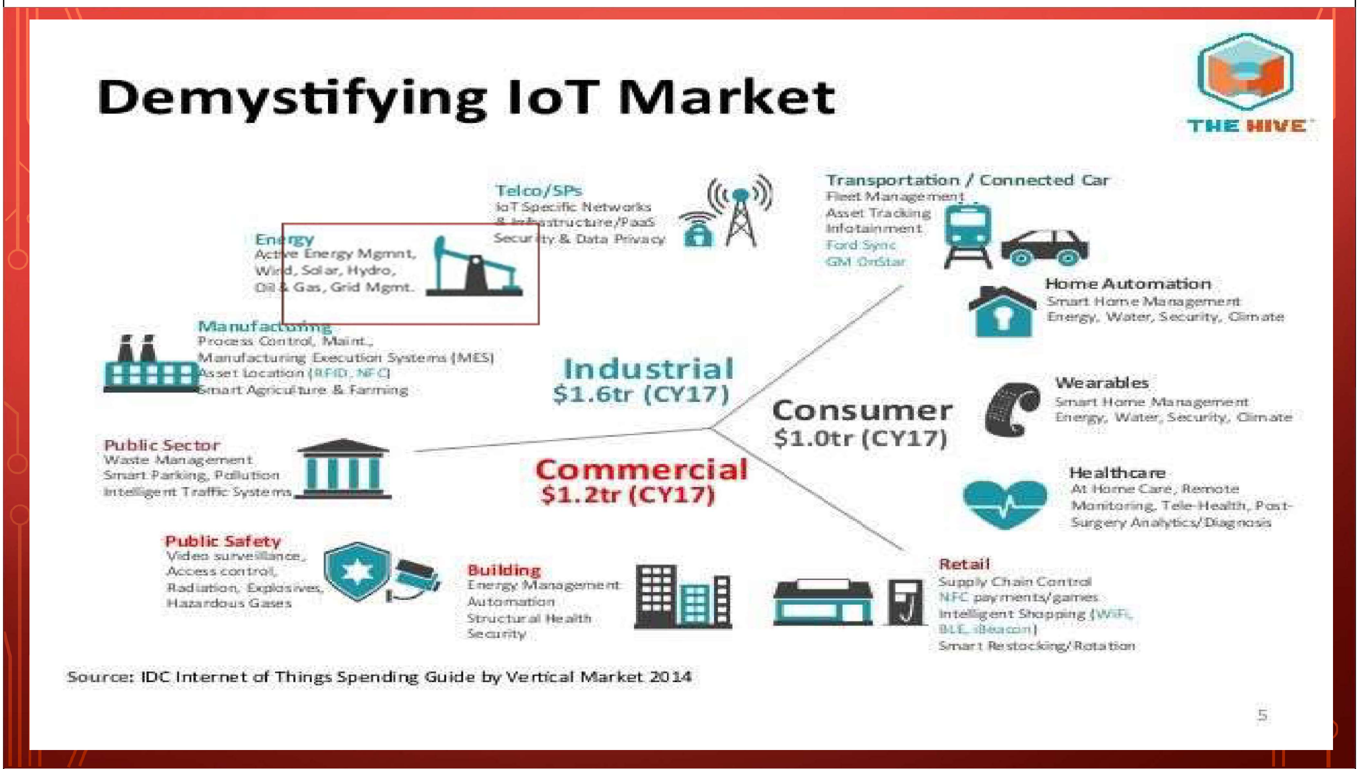
The **turning point test** for randomness is used to determine if the peaks and troughs (or turning points) of a serial data set (time-series) is independent of the order of the observations.

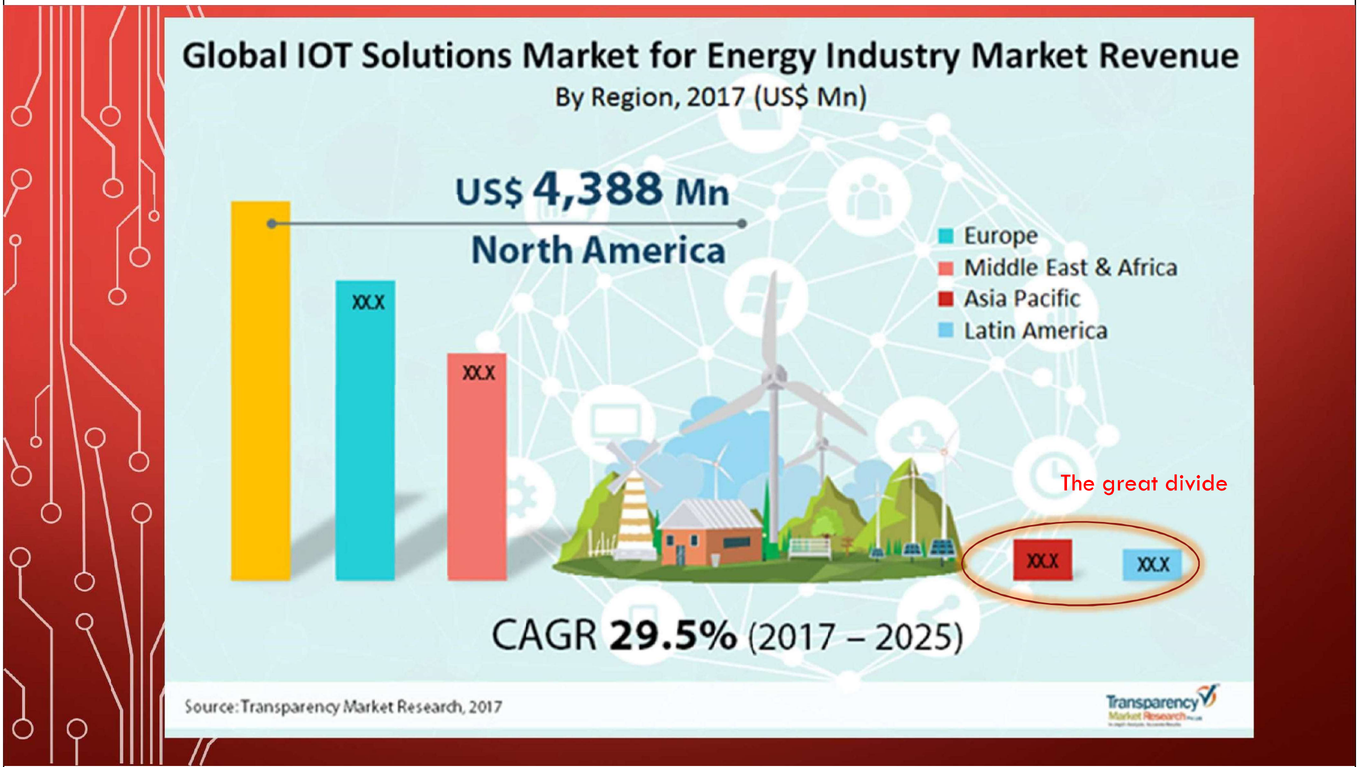
The **runs test – up and down** examines the arrangement of the numbers in a sequence to test the hypothesis of independence.


IOT FUTURE IN ENERGY INDUSTRIES
BY MOHAMMAD RAZA, ENGINEERING MANAGER, LIFE CYCLE PRODUCTIVITY, GE & BOARD MEMBER OF ESREDA


INFLUENCE OF IOT
• Visible Impact to the standard of living on consumer-based products today.
• Digital transformation is here to stay. Every industry setting aside budgets for it.
• 80% of the future jobs are yet to be defined circling around IIOT.
• Blockchain with the merger of IoT and AI brings in new dimensions and concepts.
• Disruptive technology emerging at unprecedented rate.
• Aggressive atmosphere all around (environment, equipment operations, knowledge sharing, data safeguarding, data security, etc)
• IIOT brings in opportunities and challenges of very unique nature the world has never seen before.

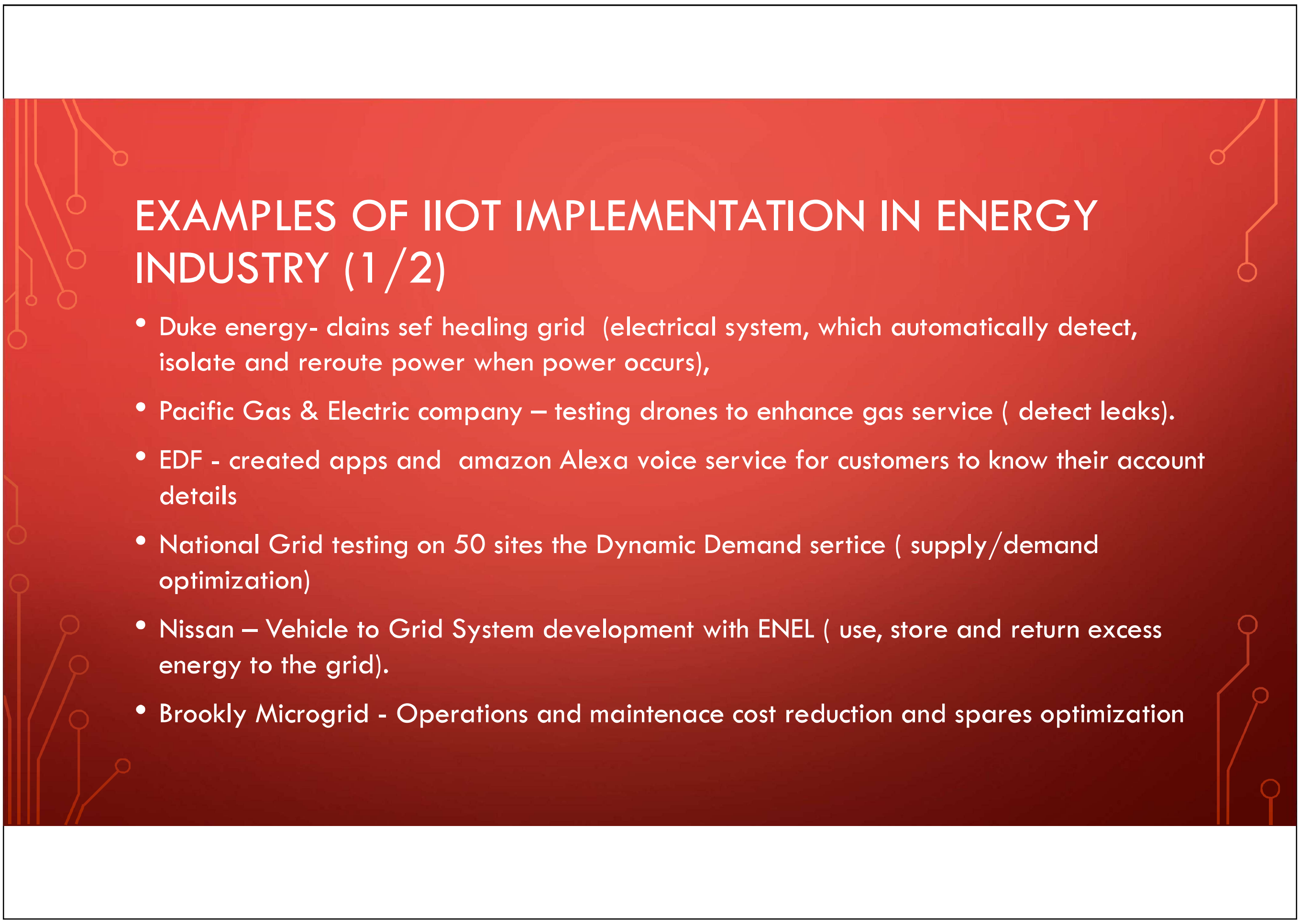

THE IoT PLATFORM OPPORTUNITY
The Internet of Things (IoT) has a potential economic impact of 2.7-6.2 trillion USD until 2025
$ trillion, annual
Range of sized potential economic impacts
Low
High
Impact from other potential applications (not sized)
X–Y
Mobile Internet 3.7–10.8
Automation of knowledge work 5.2–6.7
Internet of Things 2.7–6.2
Cloud technology 1.7–6.2
Advanced robotics 1.7–4.5
Autonomous and near-autonomous vehicles 0.2–1.9
Next-generation genomics 0.7–1.6
Energy storage 0.1–0.6
3D printing 0.2–0.6
Advanced materials 0.2–0.5
Advanced oil and gas exploration and recovery 0.1–0.5
Renewable energy 0.2–0.3
Who will capture this opportunity
?
SOURCE: McKinsey Global Institute analysis
McKinsey & Company 3

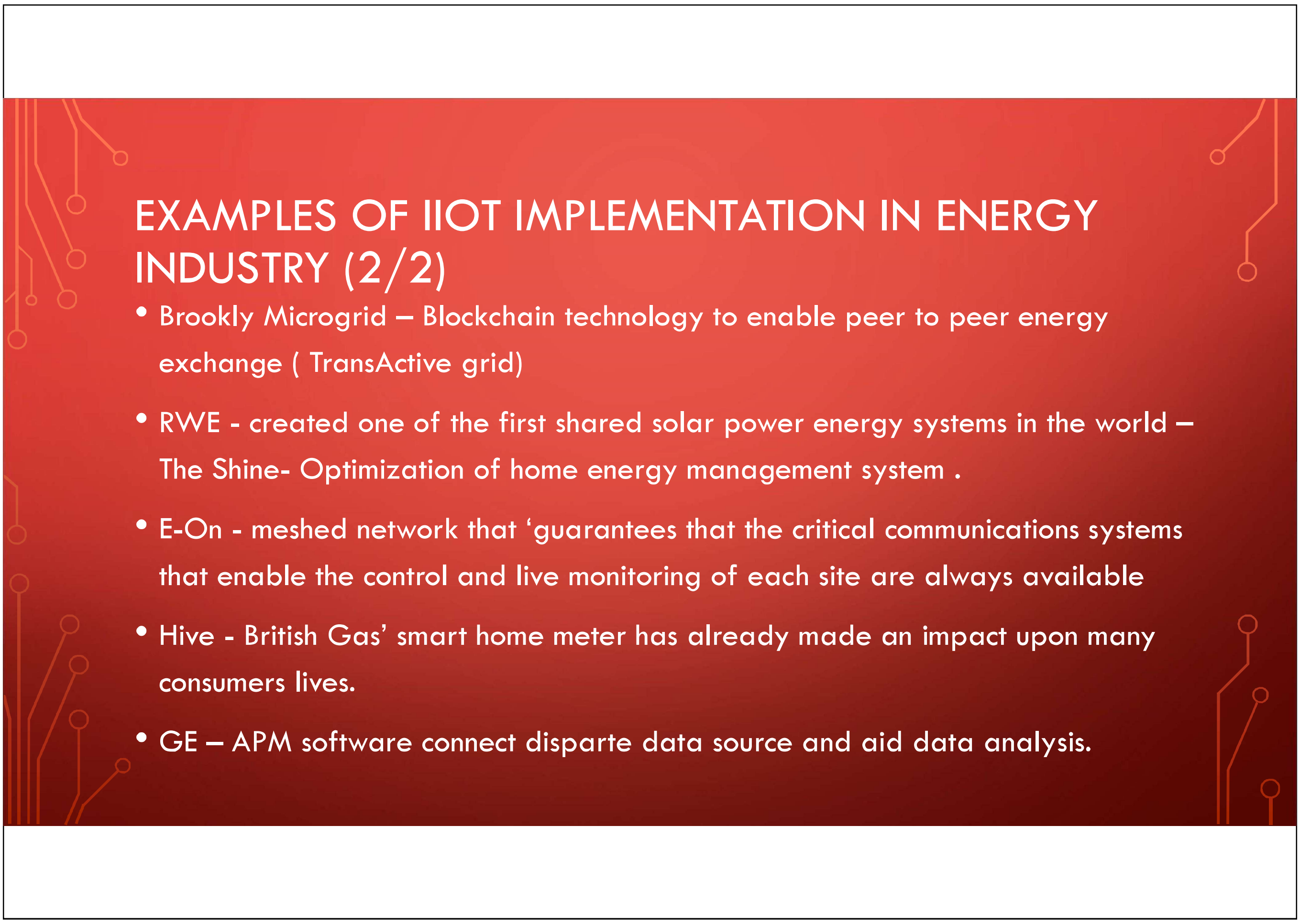

Demystifying IoT Market
THE HIVE
Telco/SPs
IoT Specific Networks
& Infrastructure/PaaS
Security & Data Privacy
Transportation / Connected Car
Fleet Management
Asset Tracking
Infotainment
Ford Sync
GM OnStar
Energy
Active Energy Mgmnt,
Wind, Solar, Hydro,
Oil & Gas, Grid Mgmt.
Home Automation
Smart Home Management
Energy, Water, Security, Climate
Manufacturing
Process Control, Maint.,
Manufacturing Execution Systems (MES)
Asset Location (RFID, NFC)
Smart Agriculture & Farming
Industrial
$1.6tr (CY17)
Consumer
$1.0tr (CY17)
Wearables
Smart Home Management
Energy, Water, Security, Climate
Public Sector
Waste Management
Smart Parking, Pollution
Intelligent Traffic Systems
Commercial
$1.2tr (CY17)
Healthcare
At Home Care, Remote
Monitoring, Tele-Health, Post-
Surgery Analytics/Diagnosis
Public Safety
Video surveillance,
Access control,
Radiation, Explosives,
Hazardous Gases
Building
Energy Management
Automation
Structural Health
Security
Retail
Supply Chain Control
NFC payments/games
Intelligent Shopping (WiFi,
BLE, iBeacon)
Smart Restocking/Rotation
Source: IDC Internet of Things Spending Guide by Vertical Market 2014
5

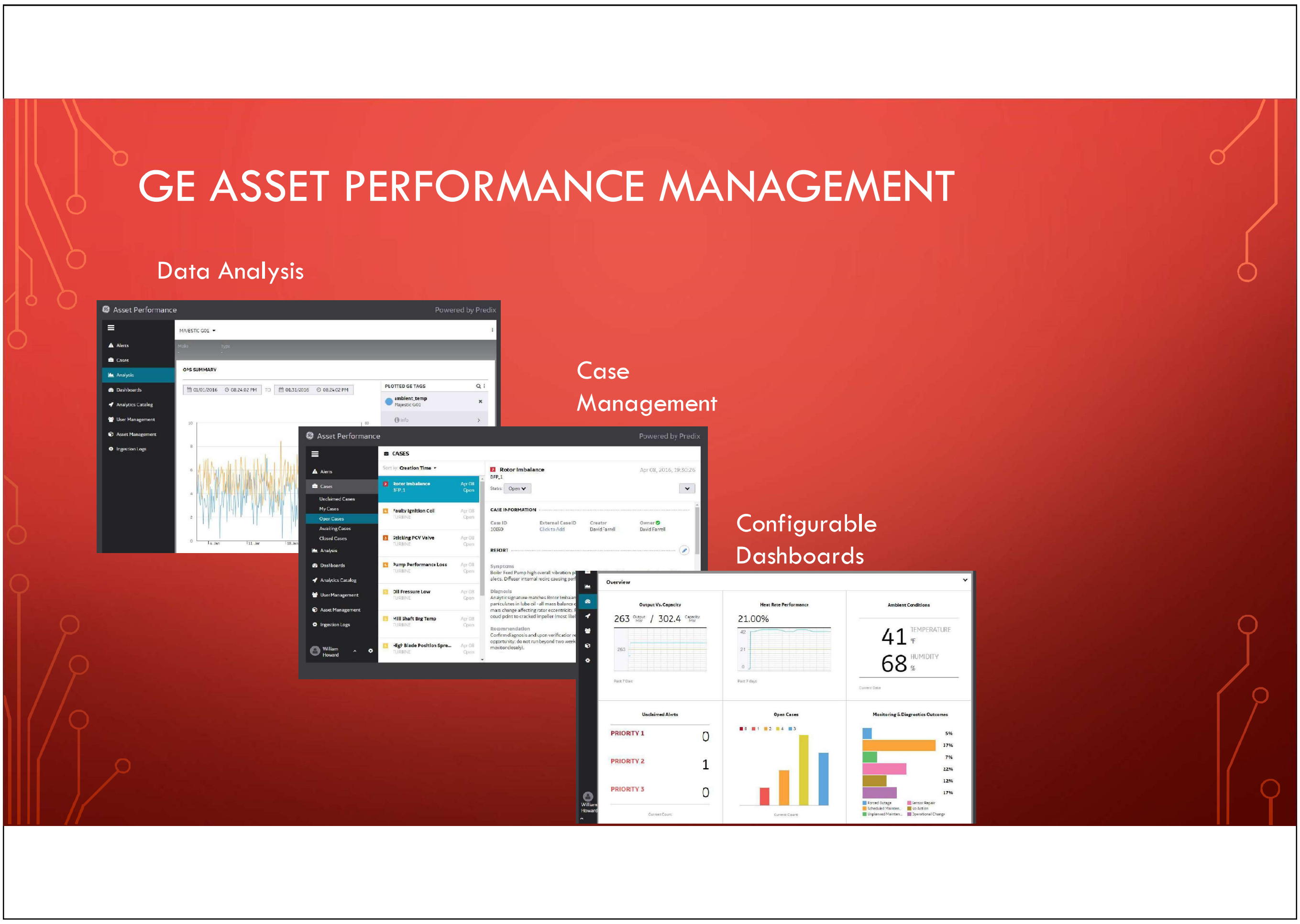


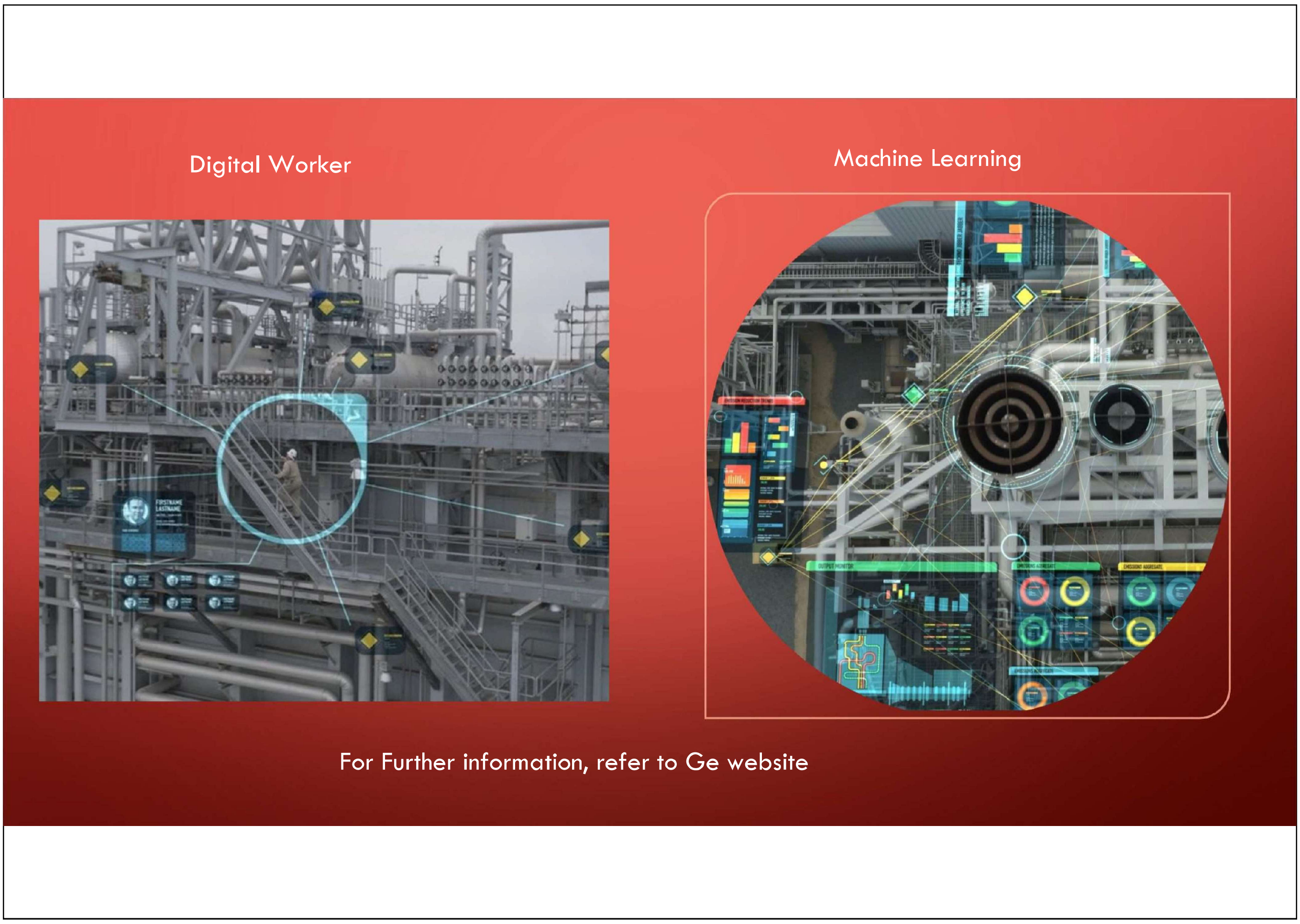
IIOT IN ENERGY INDUSTRY- WHERE IS THE FOCUS

- Sensor technology,
- Analytics development, cloud computing, Software algorithms
- Supply and demand optimization (smart Grid, micro grids),
- Energy storage and control processes
- Manufacturing facilities.
- Operations and maintenace cost reduction and spares optimization
- Risk based inspections methodolgies (predict and proactive know precisely what to do)
- Overall plant health monitoring


EXAMPLES OF IIOT IMPLEMENTATION IN ENERGY INDUSTRY (1/2)
• Duke energy- clains sef healing grid (electrical system, which automatically detect, isolate and reroute power when power occurs),
• Pacific Gas & Electric company – testing drones to enhance gas service (detect leaks).
• EDF - created apps and amazon Alexa voice service for customers to know their account details
• National Grid testing on 50 sites the Dynamic Demand sertice (supply/demand optimization)
• Nissan – Vehicle to Grid System development with ENEL (use, store and return excess energy to the grid).
• Brookly Microgrid - Operations and maintenace cost reduction and spares optimization


EXAMPLES OF IIOT IMPLEMENTATION IN ENERGY INDUSTRY (2/2)
• Brookly Microgrid – Blockchain technology to enable peer to peer energy exchange (TransActive grid)
• RWE - created one of the first shared solar power energy systems in the world – The Shine- Optimization of home energy management system .
• E-On - meshed network that 'guarantees that the critical communications systems that enable the control and live monitoring of each site are always available
• Hive - British Gas' smart home meter has already made an impact upon many consumers lives.
• GE – APM software connect disparte data source and aid data analysis.


GE ASSET PERFORMANCE MANAGEMENT
Data Analysis
Case Management
Configurable Dashboards


Digital Worker
Machine Learning
For Further information, refer to Ge website

CHALLENGES OF IIOT

- Consolidation of data, making few very powerful against many deprived of it
- Advanced analytics, with machine to machine interfaces – wish vs real value
- Data sharing legislations – how will international laws being framed ?
- Cybersecurity (device and network)- vulnerability and impact.
- World order with socio-economic implications using IIOT – will the gap of equality and justice among nations (rich and poor) increase ?

SUMMARY

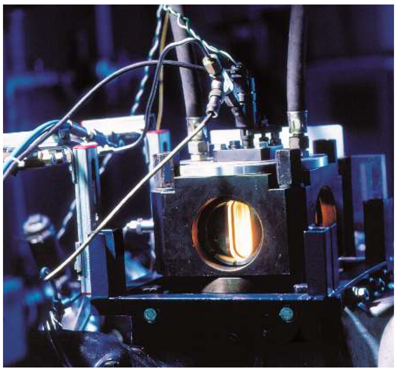
- Huge budgets and investments into IIOT.
- Rapid progress with disruptive methodology.
- IIOT is here to stay and influence all sections of life.
- Several use cases in IOT and relatively few in IIOT. Long way to go
- Challenges at every front.
- Time will tell if it is helping humanity vs disorienting and damaging humanity

57th ESReDA Seminar on

Advances in Reliability, Risk and Safety Analysis with Big Data

23rd and 24th October 2019

Universitat Politècnca de Valencia, Spain


Program and Venue

Scope of the Seminar

Industrie 4.0 is an industrial action that corresponds to the increasing integration of industrial production and information and communication technologies. It includes different aspects, among them cyber-physical systems, big data, internet of things, augmented reality, cloud computing and cognitive computing.

With recent improvements in sensor technologies, including miniaturization, performance, cost and energy consumption and in information systems resulting in increased functionality at lower costs, obtaining very important quantities of data from running industrial equipment in a cost-effective manner is now a standard practice.

To treat all the data gathered by the sensors and to transform it in useful information, industries seek to make a greater use of Artificial Intelligence (AI). There are several AI techniques, as are Machine Learning, Predictive Modelling and Deep Learning.

Among the most promising applications of these concepts can be found in Reliability, Risk and Safety Analysis. In seeking, opportunistically, the benefits from these new technological capabilities, it is important to remain critical and to address potential side or adverse effects as well especially for high-risk industries where errors can become dramatic. It is the role of the ESReDA association to organise an expert debate and further collaborative work on this topic.

For this 57th ESReDA Seminar we are concerned and invite to focus on Big Data challenges and applications. So the main topics will be the discussion of the following subjects:

- Retention and quality of data
- Data analytics
- Feature selection and extraction
- Identifying potential biases
- Data ownership and security
- Databases

The main point is: what can be done to improve the management of reliability, risk and safety making good use of these new capabilities?

This Seminar will be a forum to explore and discuss these topics. The Seminar is aimed at addressing issues met by different industries.

The programme proposes technical papers which cover different topics concerned with the application of Artificial Intelligence to Reliability, Risk and Safety Analysis. Besides, a specific round table discussing the different topics is organised. The technical programme includes plenary presentations by leading academics and scientists.

Seminar Organisation and Venue

Location

The School of Engineering Design, building 7B
http://www.upv.es/plano/plano-2d-en.html
Technical University of Valencia / Universitat Politècnica de València (UPV)
Camino de Vera, s/n
46022 Valencia
SPAIN

Organisation

The Seminar is jointly organised by ESReDA and CMT Motores Térmicos (UPV).

Chairman of the Seminar

L. Ferreira (ESReDA President, Prof. Universiy of Porto, Portugal)

B. Tormos (CMT senior researcher, Prof. Universitat Politècnica de València, Spain)

Technical Programme Committee

- Antonio Sola - Consultant, Spain
- André Lannoy - IMdR/ESReDA, France
- Bernardo Tormos - Universitat Politècnica de València, Spain
- Henk Wells – Consultant, The Netherlands
- Kaisa Simola - EC JRC Petten, The Netherlands
- Luís Ferreira – Universidade do Porto, Portugal
- Marco Riani – University of Parma, Italy
- Maria Grazia Gnoni – Università di Salento, Italy
- Micaela Demichela – Politecnico de Torino, Italy
- Mohamed Eid - CEA, France
- Mohammad Raza – GE Power, Switzerland
- Nicolas Dechy - Institut de radioprotection et de sûreté nucléaire, France
- Rasa Remenyte-Prescott - University of Nottingham, United Kingdom
- Siegfried Eisinger - DNV GL, Norway
- Tuuli Tulonen – Tukes, Finland,
- Vytis Kopustinskas - EC JRC Ispra, Italy
- Victor Borges - Thales, UK

Opening of the Seminar

To be announced

Closing of the Seminar

The President of the Board of Directors of ESReDA

Relevant dates:

- 22nd October 2019: Project Groups meetings, Board of Directors meeting.
- Seminar: 23rd and 24th October 2019
- ESReDA Gala Dinner: 23rd October 2019

Draft Program

Note*: All the meetings and Seminar Sessions will take place at* The School of Engineering Design, building 7B

Tuesday, October 22nd 2019

10:15 Coffee Break
10:30 Project Group meetings
12:30 Lunch

14:30 – 17:30 Board of Directors meeting

Wednesday, October 23rd 2019

Seminar Day 1

09:00 - 09:10 Welcome address by President and Seminar chair, Luis Ferreira
09:10 - 09:30 Welcome address by the Sub-Director for International Relations, University of Valencia.

09:30 - 10:30 Key Note speech 1:
Dr. Prof. Olga Fink – "System health monitoring with deep learning: Is big data all we need?"

10:30 - 11:00 Coffee Break

11:00- 12:30 Session I - Safety and big data

11:00 – 11:30	Siegfried Eisinger, Jon Arne Glomsrud, Justin Fackrell	Recommended Practice for Assurance of Data-driven Algorithms and Models
11:30 – 12:00	Mathias Verbeke, Alessandro Murgia, Tom Tourwé, Elena Tsiporkova	Fleet-based Remaining Useful Life Prediction of Safety-critical Electronic Devices
12:00 – 12:30	C. Harrison, X. Ge, J. Stow	Assessing GB Train Accident Risk Using Red Aspect Approaches to Signal Data

12:30 – 13:30 Lunch

13:30 – 15:00 Session II - Data and prediction

13:30 – 14:00	Henk Wels	Quality in data for unavailability of power plants

14:00 – 14:30	Luís Pereira, Luís Ferreira	Mechanical seal failure prediction in an oil refinery: a first attempt to solve the problem using a data-driven approach
14:30 – 15:00	Asun Lera St.Clair	Trust in AI

15:00- 15:30 Coffee Break

15:30- 17:00 Panel Discussion: Risk Based Inspections using big data analytics- do we have a clear path way?

Panelists: Henk Wels, Mohammad Raza
Moderator: Luís Ferreira

17:00- 17:15 Presentation of next Seminar and other ESReDA and ESRA activities

17:15 – 18:30 - Visit to the University of Valencia Research Centre: CMT-Motores Térmicos

19:30 Dinner at Restaurant (*Restaurant in downtown Valencia, to be announced)*

Thursday, October 24th 2019

Seminar Day 2

09:00- 10:00 Key Note Speech 2:
Prof. Sebastián Martorell - Advanced analysis of reliability and risk of equipment subjected to degradation and obsolescence

10:00- 11:00 Key Note Speech 3:
Prof. Marco Riani - Robust statics for big Data Analytics

11:00- 11:30 Coffee Break

11:30 – 12:30 Session III - Data Analytics and the IOT Future

11:30-12:00	Daniel Gaspar, Luís Ferreira	A proposal of an algorithm to simulation censored data right type I in reliability field
12:00- 12:30	Mohammad Raza	IOT future in Energy Industries

12:30- 12:45 - Thanks and End of the Seminar

12:45- 14:00 Lunch

Bibliographic notes about Key Note Speakers

Prof. Dr. Olga Fink, ETH Zurich

She is at the SNSF (Swiss National Science Foundation) and it is Professor for intelligent maintenance systems at ETH Zürich. Before, she was heading the research group "Smart Maintenance" at the Zurich University of Applied Sciences (ZHAW). Holds Ph.D. in civil engineering from ETH Zurich, and Diploma degree in industrial engineering from Hamburg University of Technology. Gained valuable industrial experience as reliability engineer for railway rolling stock and as reliability and maintenance expert for railway systems. Research focuses on Data-Driven Condition-Based and Predictive Maintenance, amongst others.

Prof. Sebastián Martorell, Universitat Politècnica de València

He is Full Professor of Nuclear Engineering, Director of the Radiation Service and Ex-Director of the Chemical and Nuclear Department at the Universitat Politècnica de València, Spain. Prof. Martorell received his Ph.D. in Nuclear Engineering from Universitat Politècnica de València in 1991. Head of the MEDASEGI research group, his research areas are probabilistic and deterministic safety analysis, uncertainties, risk-informed decision making, and RAMS modelling and optimization. In the past 27 years, he has served as consultant to governmental national and international agencies, nuclear facilities and private organizations in areas related to risk and safety analysis, especially applications to safety system design and testing and maintenance optimization of nuclear power plants. Prof. Martorell has taken part as Main Researcher in 64 national and international research projects and contracts. Prof. Martorell`s publications include 142 SCI and JCR papers in journals and proceedings of conferences in various areas of reliability, maintainability, availability, safety and risk engineering (h-index 21, about 1600 citations in WOS). He serves as a member of the Editorial Board of Reliability Engineering and System Safety International Journal. He is also an editorial board member of the Journal of Risk and Reliability, Proceedings of Institution of Mechanical Engineers, Part O. He has been Vice-Chairman of European Safety and Reliability Association (ESRA).

Prof. Marco Riani – University of Parma

He is a Full Professor of Statistics at the University of Parma, where he is teaching and research in Statistics and Informatics. Teaching has concerned courses at graduate, post graduate and PhD level. Supervisor of PhD students (two of them, namely Tiziano Bellini now at HSBC bank and Francesca Torti now at the Joint Research Centre of the European Commission won the prize for the best Italian PhD thesis in statistics). He is currently Director of the Interdepartmental center Ro.Sta.Bi.Da.C -– ROBUST STATISTICS FOR BIG DATA CENTRE of the University of Parma, Member of the Steering Committee of the SIS CLADAG (Classification and Data Analysis Group of the Italian Statistical Society), member of the board of the PhD programme in Statistics and Financial Mathematics, University of Milan Bicocca, Italy, Member of the Steering Committee of ICORS (International Conference of Robust Statistics), Scientific coordinator of the module "Advanced personal computing" of the "Marketing Management Master" organized by the University of Parma jointly with financial Italian newspaper “Il Sole 24 Ore”, and scientific coordinator of the module “Informative systems and statistics tools for market management” of the master in Agribusiness and Food Management organized by the University of Parma.

Registration and Seminar Fee

Registration will be accepted until the 11th October 2019.

A registration form and information package for the venue will be made available on the ESReDA website:

https://www.esreda.org/event/57th-esreda-seminar

The fees according to ESReDA’s rules are:

- Speakers: one speaker per accepted paper may participate without paying seminar fees.
- ESReDA members: up to three participants of ESReDA members are taken in charge by organization.
- Participant: 300€ per participant.
- Accompanied persons for Gala Dinner: 50€ per participant

Fees are to be paid by bank transfer to ESReDA account:

Holder: ESReDA

Bank: BNP Paribas Fortis Bank, Boulevard Jamar 1 D, 1060 Bruxelles, Belgium

IBAN: BE69 0012 3728 1678

BIC: GEBABEBB

Subject: Registration to the 57th ESReDA Seminar

Organizers

Universitat Politècnica de València

The Universitat Politècnica de València is a public, dynamic and innovative institution, dedicated to research and teaching that, while maintaining strong links with the social environment in which it carries out its activities, opts for a strong presence abroad. It is a young university, which celebrates its 50th anniversary during the academic year 2018-2019.

Its community is made up of about 34,000 students, 3,600 professors and researchers and 1,500 administrative and service professionals distributed among its three campuses located in Alcoy, Gandia and València.

At present, the UPV is constituted by 13 university centers, of which 9 are higher technical schools, 2 are faculties and 2 are higher polytechnic schools. In addition, it has a Doctoral School and 3 affiliated centers (Florida University, Berklee College of Music and EDEM Business School).

The Seminar will be supported by C-Motores Termicos at the Universitat Politècnica de València:

CMT-Motores Térmicos is a research and educational center fully involved in the development of the future combustion engine, and incorporating more than 100 people. For more than 35 years have conducted basic research for better understanding the relevant physical processes involved, and applied studies for optimizing the engine behavior and assisting in its development.

From the deep scientific knowledge to the real-life problems of the automotive industry, we combine experimental tests conducted in our state-of-the-art facilities, and theoretical studies providing relevant technical and scientific results. Our interdisciplinary approach covers different research areas, and aims at Excellence and Innovation.

More information available in: https://www.cmt.upv.es/

ESReDA

ESReDA is a European Association which provides a forum for the exchange of information, data and current research in Safety and Reliability and a focus for specialist expertise.

The Safety and Reliability of processes and products are topics which are the focus of increasing interest Europe wide. Safety and Reliability Engineering is viewed as being an important component in the design of a system. However the discipline and its tools and methods are still evolving and expertise and knowledge dispersed throughout Europe. There is a need to pool the resources and knowledge within Europe and ESReDA provides the means to achieve this.

ESReDA was established in 1992 to promote research, application and training in Reliability, Availability, Maintainability and Safety (RAMS). The Association provides a forum for the exchange of information, data and current research in Safety and Reliability and a focus for specialist expertise.

More information at:

https://www.esreda.org/

Contact Point:
Inga Šarūnienė
Lithuanian Energy Institute
Breslaujos str. 3, 44403, Kaunas
Lithuania
E-mail: inga.saruniene@lei.lt

ESReDA Project Group on Big Data, Reliability, Risk and Safety Analysis

The project group (PG) “Big Data, Reliability, Risk and Safety Analysis” aims to write a working technical document, if possible a book, in which it will try to identify the evolutions, paradigm shift and challenges caused by the emergence of Big Data in the Reliability, Risk and Safety Analysis of industrial equipment.

In doing so, the PG will attempt to identify the advantages and disadvantages of its use for equipment users by identifying the techniques to be applied, the standardization needs (if any) and the existing challenges to an application of new scientific knowledge in these areas.

We expect that this technical document will be published with a EUR Tech-Doc reference number.

Venue

Valencia

The port city of Valencia lies on Spain's southeastern coast, where the Turia River meets the Mediterranean Sea. It's known for its City of Arts and Sciences, with futuristic structures including a planetarium, an oceanarium and an interactive museum. Valencia also has several beaches, including some within nearby Albufera Park, a wetlands reserve with a lake and walking trails.

Valencia is a very touristic city with a great offer of hotels.

KJ-NA-30415-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/836400

ISBN 978-92-76-23807-2