

**MODELING FORCED OUTAGE IN HYDROPOWER GENERATING UNITS FOR  
OPERATIONS PLANNING MODEL**

by

Abhishek Agrawal

B.Tech., Indian Institute of Technology (BHU), Varanasi, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES  
(CIVIL ENGINEERING)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2018

© Abhishek Agrawal, 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Modeling Forced Outage in Hydropower Generating Units for Operations Planning Model

---

submitted by Abhishek Agrawal in partial fulfillment of the requirements for

the degree of Masters of Applied Science

in Civil Engineering

---

**Examining Committee:**

Professor Ziad Shawwash

---

Supervisor

Professor Gregory Lawrence

---

Supervisory Committee Member

---

Supervisory Committee Member

---

Additional Examiner

**Additional Supervisory Committee Members:**

---

Supervisory Committee Member

---

Supervisory Committee Member

## **Abstract**

Unplanned outages of generating units, also known as forced outages, act as a source of operational uncertainty for hydropower companies like BC Hydro. Forced outages reduce plant availability and causes loss of system flexibility and revenues. A combination of both likelihood of occurrence (frequency) and severity of outage event (duration) truly represents the risks posed by forced outages. Energy studies, using simulation and optimization models, are carried out by utility companies to incorporate different sources of uncertainties and maximize benefits in multi-purpose, multi-reservoir systems. The Department of System Optimization at BC Hydro is developing new quantitative approaches to model uncertainty of forced outages in their operations planning models and system energy studies. In this thesis, statistical properties of forced outage datasets are quantified, and different algorithms to generate scenarios of forced outages are developed. The statistical analysis methods and scenario generation algorithms are applied for a major hydroelectric facility in the BC Hydro system having 10 generating units and results are presented. Time to Failure and Time to Repair for outage events were obtained and checked for annual trends, seasonality and correlations. Outages of units were also evaluated for homogeneity. The impacts of planned outage on forced outages were quantified and suitable probabilistic distributions were developed to represent frequency and duration of outages. Three different scenario generation algorithms were developed using Markov/Semi-Markov based processes and Monte Carlo Simulation. It was found that Semi-Markov based scenario generation algorithm that comprehensively accounts for impacts of planned outages on forced outages is best suited to generate scenarios of forced outages for energy studies and operational planning models.

## **Lay Summary**

Forced (unplanned) outages of generating units in hydropower plants lead to loss of generation capacity. The uncertainty in occurrence and duration of outages affects the results of energy planning models and causes loss of flexibility and revenues during system operations. Scenarios of forced outages can model the uncertainty in generation capacity. This thesis presents the methods that can be used to comprehensively account for uncertainty of forced outages via scenarios. Recent data on forced outages were obtained and their statistical properties were investigated. Different scenario generation methods were developed to model those properties. Tests were carried out for a BC Hydro generating station. It was found that the impact of planned outages on forced outages should be incorporated to generate future outage scenarios. The most suitable method was identified and recommended for quantifying uncertainty due to forced outages in generating units.

## **Preface**

The author contributed in building the database, development of methods and results analysis for the research presented in Chapter 2, 3 and 4. A version of Chapter 4 was published in the proceedings of the Annual Canadian Dam Association Conference - 2017, under the title “Modeling Forced Outages of Hydropower Generator Units for Reliable Dam Operation” (Agrawal et al. 2017). The author’s research supervisor - Professor Ziad Shawwash and postdoctoral researcher Dr. Quentin Desreumaux guided the development of statistical analysis and scenario generation framework.

The data on historic outage records were collected from BC Hydro’s Unit Status Record with the help from its administrator Mr. Stan Mathews, who also provided useful insights about outage data. The research objectives were shaped by Prof Shawwash and the Department of System Optimization at BC Hydro. Dr. Andrew Keats provided valuable feedback for the methods developed in Chapter 3. Mr. Tim Blair and Dr. Dave Bonser helped in verifying the assumptions made and benchmarking criteria used in Chapter 4.

# Table of Contents

<b>Abstract.....</b>	<b>iii</b>
<b>Lay Summary .....</b>	<b>iv</b>
<b>Preface.....</b>	<b>v</b>
<b>Table of Contents .....</b>	<b>vi</b>
<b>List of Tables .....</b>	<b>x</b>
<b>List of Figures.....</b>	<b>xi</b>
<b>List of Abbreviations .....</b>	<b>xiv</b>
<b>Acknowledgements .....</b>	<b>xvi</b>
<b>Dedication .....</b>	<b>xvii</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1    Problems due to uncertainty in unit availability .....	2
1.2    Probabilistic analysis of forced outages.....	4
1.3    Goals and Objectives .....	5
1.4    Organization of Thesis.....	6
<b>Chapter 2: Literature Review.....</b>	<b>7</b>
2.1    Database Management System for Generating Unit States .....	7
2.2    Unit States in the BC Hydro Unit Status Record.....	8
2.2.1    Available States .....	9
2.2.2    Forced Outage State.....	12
2.2.3    Maintenance Outage .....	13
2.2.4    Planned Outage.....	14

2.2.5	Forced Extensions .....	15
2.3	Statistical analysis of unit states .....	15
2.4	Reliability analysis in Forced outage modeling.....	22
2.4.1	Basic concepts of Reliability Analysis .....	24
2.4.2	Hazard Rate and Bath Tub Curves.....	25
2.4.3	Probability distributions for different hazard rates .....	28
2.4.4	Development of reliability indices for forced outages.....	30
2.4.5	Use of Markov process in Forced Outage modeling .....	32
2.5	Research literature on Modeling of forced outages .....	37
2.5.1	Studies on cost of forced outage and preventive maintenance .....	37
2.5.2	Forced Outage modeling for reserve computation.....	39
2.6	Suitable methods for hydropower units .....	43
<b>Chapter 3: Methodology.....</b>		<b>45</b>
3.1	Sources of Unit Unavailability.....	45
3.1.1	Classification and reclassification of outages, forced extensions and de-rates .....	46
3.1.2	Defining Time to Repair and Time to Fail.....	48
3.2	Statistical Analysis.....	50
3.2.1	Trend in Occurrence and duration of outages.....	51
3.2.2	Seasonality of Forced Outages.....	52
3.2.3	Independence of TTR and TTF values .....	53
3.2.4	Impact of planned outage on forced outages .....	54
3.2.5	Probabilistic model of outages.....	55
3.3	Methods of sampling from forced outage distributions .....	58

3.3.1	Monte Carlo Simulations .....	59
3.3.1.1	State Sampling Approach .....	60
3.3.1.2	State Duration Sampling Approach .....	61
3.3.1.3	System State Transition Sampling Approach .....	63
3.4	Methods for scenario generation of forced outages .....	64
3.4.1	Markov Process-State Sampling Method.....	64
3.4.2	Semi-Markov Process for FO-FO model .....	68
3.4.3	Semi Markov Process for PO-FO model .....	73
3.5	Summary of Methods.....	77
<b>Chapter 4: Case Study and Analysis of Results .....</b>		<b>78</b>
4.1	Gordon M. Shrum Generating Station .....	78
4.2	Database of forced outages .....	80
4.2.1	Preliminary Analysis.....	81
4.3	Statistical Analysis of data.....	85
4.3.1	Homogeneity of Units.....	86
4.3.2	Trends in Forced Outage Data .....	88
4.3.3	Seasonality of outages.....	89
4.3.4	Independence of TTR and TTF.....	93
4.3.5	Impact of Planned Outage.....	94
4.3.6	Probabilistic Model for TTR and TTF.....	97
4.4	Results of Scenario Generation.....	100
4.4.1	Markov Process Model .....	101
4.4.2	Semi-Markov Process (FO-FO) model.....	105

4.4.3	Semi-Markov process PO-FO model.....	113
4.5	Summary of scenario generation methods.....	119
<b>Chapter 5: Discussion and Conclusions.....</b>		<b>120</b>
5.1	Key takeaways for building an outage database.....	120
5.2	Conclusions from statistical analysis of data.....	121
5.3	Discussions on Scenario Generation Methods.....	123
5.4	Recommendations:.....	125
5.5	Future Work.....	126
<b>Bibliography .....</b>		<b>128</b>

## List of Tables

Table 1: Available On-Line States.....	10
Table 2 Available Off-line States .....	11
Table 3 Forced Out States.....	12
Table 4 Maintenance Outage State .....	13
Table 5 Planned and Upgrade Outages .....	14
Table 6 Forced Extension of Outages .....	15

## List of Figures

Figure 1 Average Hydropower operational status .....	17
Figure 2 Mean FOD for units with different MCR ratings .....	19
Figure 3 Mean FOD for units at different age .....	19
Figure 4 Mean FOD for units operated differently .....	19
Figure 5 Major Component contribution to hydraulic unit ICbF due to unplanned outages .....	20
Figure 6: Typical bath tub curve for component failure. ....	27
Figure 7: Two-state representation of Unit Availability .....	33
Figure 8 Multi-state representation of Unit Availability .....	33
Figure 9 Trade off curve for Total cost of unit maintenance .....	38
Figure 10 Simplified chart for unit availability .....	49
Figure 11 : Description of TTR and TTF.....	50
Figure 12 TTR/TTF definition considering Planned Outages .....	55
Figure 13 Difference between ECDF and parametric CDF .....	56
Figure 14 System State Duration Method (Billinton and Li 1994) .....	62
Figure 15 Two-State model for generating units .....	65
Figure 16 Preventing overlap of sampled FO with existing PO .....	71
Figure 17 Case of TTF period crossing PO period .....	71
Figure 18 TTR- curtailment to prevent overlap with PO.....	72
Figure 19 TTR curtailment when FO exceeds PO period.....	72
Figure 20 Check for TTF value crossing a PO period .....	74
Figure 21 Accepted value of TTF and TTR.....	75
Figure 22 TTR curtailment and sampling from TTF_PF.....	76

Figure 23 Rejecting sampled TTF_FF and sampling from TTF_PF .....	76
Figure 24 Peace River System in British Columbia .....	79
Figure 25 Normalized Total Duration of Outages .....	83
Figure 26 Normalized Number of Outage Events .....	83
Figure 27 Statistical distribution of TTR of different Outages.....	84
Figure 28 ECDF curves of TTF for each unit.....	87
Figure 29 ECDF curves of TTR for each unit .....	87
Figure 30 Annual Trend in TTF.....	89
Figure 31 Annual Trend in TTR.....	89
Figure 32 Seasonality in TTR distribution.....	91
Figure 33 Seasonality in TTF distribution .....	91
Figure 34 Confidence band on TTF-winter months .....	93
Figure 35 Scatter Plot of TTR and TTF.....	94
Figure 36 Two State Unit representation .....	94
Figure 37 Impact of Planned Outage on Forced Outage.....	95
Figure 38 Impact of Planned Outages on TTR .....	96
Figure 39 Impact of Planned Outages on TTF.....	96
Figure 40 Best Fit distribution – TTR.....	98
Figure 41 Best Fit distribution - TTF.....	99
Figure 42 Best Fit distribution - TTF-FF.....	100
Figure 43 Best Fit distribution - TTF-PF.....	100
Figure 44 TTF- Markov Process.....	103
Figure 45 TTR- Markov Process .....	104

Figure 46 TTF FO-FO Model – (Parametric Input Distribution) .....	106
Figure 47 TTR FO-FO Model – (Parametric Input Distribution).....	106
Figure 48 TTF FO-FO Model (Non-Parametric Distribution) .....	108
Figure 49 TTR FO-FO Model (Non-Parametric Distribution).....	109
Figure 50 TTF FO-FO Model (PO Heuristics).....	111
Figure 51 TTR FO-FO Model (PO Heuristics).....	111
Figure 52 TTR from PO-FO model .....	115
Figure 53 Bias in higher TTR values .....	115
Figure 54 TTF-FF using Parametric distribution.....	116
Figure 55 TTF-PF using Parametric distribution.....	116
Figure 56 TTF-FF in PO-FO model with bias correction.....	118
Figure 57 TTF-PF in PO-FO model with bias corrections .....	118

## List of Abbreviations

ABNO	Available but Not Operating
AIC	Akaike Information Criterion
BC	British Columbia
BC Hydro	British Columbia Hydro and Power Utility
BCUC	British Columbia Utility Commission
CEA	Canadian Electrical Association
DAUFOP	Derating Adjusted Utilization Forced Outage Probability
ECDF	Empirical Cumulative Distribution Function
ERIS	Equipment Reliability Information System
ETA	Event Tree Analysis
FD	Forced De-rate
FO	Forced Outage
FOD	Forced Outage Duration
FOR	Forced Outage Rate
FTA	Fault Tree Analysis
GADS	Generating Availability Data System
GMS	Gordon Merritt Shrum
ICbF	Incapability Factor
LOLE	Loss of Load Expectation
LOLP	Loss of Load Probability
MC	Markov Chain

MCR	Maximum Continuous Rating
MCS	Monte Carlo Simulation
MLE	Maximum Likelihood Estimation
MMTR	Mean Time to Repair
MO	Maintenance Outage
MTTF	Mean Time to Fail
NERC	North American Electric Reliability Corporation
PDF	Probability Distribution Function
PO	Planned Outage
RBD	Reliability Block Diagram
SD	Scheduled De-rates
SNL	Speed No Load
TTF	Time to Fail
TTR	Time to Repair
US	United States
USACE	United States Army Corps of Engineers
USR	Unit Status Record
VaR	Value at Risk
WECC	Western Electricity Coordination Council

## **Acknowledgements**

The research presented in this thesis was supported and funded by grants provided to Dr. Ziad K. Shawwash by BC Hydro and the Natural Sciences and Engineering Research Council of Canada (NSERC; CRDPJ 476296 - 14). I would like to express my sincere gratitude to my supervisor Prof. Shawwash for the continuous support, patience, mentorship and inspiration throughout my graduate studies and research. He offered me the opportunity to get involved in the hydropower industry and gain valuable experience. I am also grateful to Dr. Gregory A. Lawrence for his support and guidance during my program at UBC and for reviewing this thesis.

I would like to express my immense gratitude to Dr. Quentin Desreumaux for providing invaluable technical guidance during my research. His critique and inputs were instrumental in achieving many of my research objectives. I would also like to thank the BC Hydro operations planning engineers: Mr. Tim Blair, Dr. Dave Bonser, Dr. Andrew Keats and Mr. Amr Ayad. Without their passionate participation and regular feedback, the application of methods would not have been successful.

I offer my heartfelt gratitude to Jonathan van Groll, Mehretab Tadesse, Luis Galindo and Erica Kennedy for making the Masters program so enjoyable. I would also like to thank my friend Ms. Neha Kothari for her irreplaceable assistance and encouragement during my stay in Vancouver. Special thanks are owed to my parents and other family members, who have supported me throughout my years of education, both morally and financially. Last but not the least, I am grateful to my landlords D.J. and Hari Singh for their love and affection.

*To my parents, Rajkumar and Laxmi*

## **Chapter 1: Introduction**

Utility companies around the world are using robust models for risk management and optimization of generating resources to participate in competitive electricity markets. System optimization studies are essential for finding a set of operating conditions for multi-purpose, multi-reservoir systems that maximizes the net benefits while satisfying all system wide constraints. BC Hydro is one of the largest utility companies in Canada that manages multi-purpose, multi-reservoir system for generation and supply of power. It operates 30 hydroelectric facilities and 2 thermal plants that provide power to its 4 million customers (BC Hydro 2016). Energy studies are executed every month over a 5-year planning horizon to forecast an optimal set of reservoir storage targets and generating station operations under forecasts of market, inflow and weather conditions. These studies also provide forecasts of energy imports and exports, corporate financials and price signals that are used as a decision support tool for hydro and thermal plant operations on a system scale relative to markets.

There is a significant level of data uncertainty in the decisions made during operations planning. In order to make informed decisions within such a context, it is fundamental to properly model the nature and consequences of the uncertainty involved (Barroso and Conejo 2006). Most of the studies related to planning and optimization of reservoir operations involve managing many types of uncertainty such as reservoir inflows, electricity demand and market prices. In addition to these natural or market driven uncertainties, there are operational uncertainties pertaining to the system as a whole such as the available system capacity. The system capacity is a function of individual unit availabilities. Availability of generating units is impacted by the occurrence and duration of planned and unplanned outage events.

The unavailability of a generating unit can be caused by outages that may occur suddenly or are scheduled ahead in time. Scheduled outages or ‘planned outages’ (PO), are generally deterministic as there is negligible uncertainty in their duration and timing. These are known to system modelers in advance and hence accounted for in energy planning studies. However, the complete shutdown of units or sudden decrease of available generating capacity also occurs due to various unforeseen conditions. These unscheduled shut downs are called ‘forced outages’ (FO) and temporary reduction of maximum generating capacity are called ‘forced de-rates’ (FD). A generating unit may be forced out, or de-rated, due to various reasons related to generator, turbine, water conduit, exciter, transformer, circuit breakers, auxiliary parts among others (NERC 2015).

### **1.1 Problems due to uncertainty in unit availability**

Regulators in North America and Europe require utility companies to keep contingency reserves to deal with uncertainty in the system capacity caused by unplanned outages. These reserves are defined as the capability above the firm system demand required to compensate for outages, errors in load forecasting and also to balance system voltage and frequency. Reserves can be computed as some percentage of the installed capacity, or can be the sum of the largest generating unit capacity plus some constant value of the installed capacity (Cepin 2011). The North American Electric Reliability Corporation (NERC) divides the geographical areas under its jurisdiction into regions and sub-regions. These sub-regions must comply with the reserves requirements mandated by NERC. For entities within the western interconnection, such as British Columbia, the reserves are based on a proposal by the Western Electricity Coordination Council (WECC) which states that the minimum amount of contingency reserve should be the greater of the most severe single

contingency, or the sum of 3% of the balancing area load and 3% of the balancing area generation (Milligan et al. 2010).

Since forced outage are uncertain events, statistical representation of outages become necessary. There is a cost of keeping conservative operating reserves, which could have been otherwise allocated for power generation (Bornak 2013). Beyond the issue of reserve capacity, correct representation of FO is necessary because it has been established that the net worth of energy produced by the system is impacted by the frequency of outages (failure rates) and by the duration of outages (repair rates) (Ryan et al. 1990).

(Parrish 2015) discussed methods to quantify costs of outages of the US Army Corps of Engineers (USACE) hydropower facilities. The report argued that the economic impact of unit outages is a function of the plant's *operating strategy, water availability, regional generation mix, and regional electricity demand*. Changes in *operating strategy* due to outages lead to loss of revenues. For example, take the case of forward contracts, which are often used by producers to reduce financial risk in energy markets. If the producer has insufficient capacity at the time of delivery due to unit unavailability, then they might be forced to buy energy from spot markets to honour their contract. This can lead producers to incur financial losses (Barroso et.al., 2006, Das et.al., 2012). *Water availability* impacts cost of outages because the water level in the reservoir is seasonal and thus financial cost of outages is also seasonal. Outages lead to wasteful spill which becomes a cost of outage. *Regional energy demand* is important because when electricity demand increases in a region, more expensive generating units are used. This increases the value of generation lost due to a unit outage during certain periods. And lastly, if fossil fuels are used to

replace a hydropower unit outage, it may lead to emission costs due to *regional generation mix*. It is, therefore, safe to conclude that non-inclusion of unplanned outages affects all operating objectives by varying degrees. These adverse impacts demonstrate why it is important to account for uncertainty in unit availability in operations planning studies.

## **1.2 Probabilistic analysis of forced outages**

Understanding the underlying failure processes and predicting when equipment might fail is challenging since outages can depend on several parameters like age of unit, periodic maintenance, load on the unit, start-stops, component wear and tear, etc. The relevance of each of the factors is almost impossible to quantify (Braglia, et.al, 2012). Two methods of analysis of forced outages can be used: deterministic and probabilistic. Use of deterministic analysis would imply perfect knowledge of future forced outages. The weakness of deterministic analysis for failure events is that it will invariably over-predict or under-predict the consequences of events. Probabilistic analysis provides an appropriate mathematical framework for characterising uncertainty and making informed decisions to overcome this challenge. Probability based analysis of risks were originally developed for application in the military and aerospace industry. It was then applied to the nuclear industry by using a reliability analysis framework to prevent catastrophic failures. Reliability analysis is applied to study the probability of the system to remain in the operating state without failure. However, this concept can be modified and applied to repairable systems such as the generating units of hydropower plants. If a modeler considers a serious failure event that is highly unlikely to occur, it leads to wastage of resources during regular operations. On the other hand, if the modeler ignores an event considering it to be less severe but that event occurs rather frequently, then the system would have insufficient reliability. A probabilistic assessment of

generating unit outages can consider the operational impact of outage events on system operation and the probability of its occurrence (Billinton and Li 1994). A combination of both severity of event and likelihood of occurrence truly represents the risks posed by forced outages.

Access to long term historic unit outage data, representable probabilistic description of forced outages and improved computational power enables the use of advanced computational algorithms such as Markov Chain processes and Monte Carlo Simulations to generate scenarios of forced outages that are statistically similar to the historic forced outage values. These scenarios of forced outages combined with the planned outage data can significantly improve operations planning of generating facilities.

### **1.3 Goals and Objectives**

The aim of this research is to identify suitable methods for modeling forced outage of hydropower generating units and to develop scenario generation methods for system optimization studies of the BC Hydro system. To achieve these research objectives the following tasks were undertaken:

- 1) Study of literature on modeling of forced outages to find methods relevant for hydropower generating units.
- 2) Build a database from historical outage records and perform statistical analysis of outage data to understand key statistical properties of forced outages that need to be modeled.
- 3) Develop and test a probabilistic model to represent outages.
- 4) Develop and test different scenario generation methods to create potential outage scenarios and benchmark the results against historic values and to select the most suitable model.

## **1.4 Organization of Thesis**

This thesis is divided into five chapters. This first chapter gives a brief background of the problem and rationale for the research objectives. In the second chapter, a literature survey is carried out to analyse different modeling approaches of forced outages and identify the most appropriate modeling approach for hydropower units. This chapter also includes description of the standard unit status record database, important definitions related to forced outage studies and reliability indices used for benchmarking scenarios. The third chapter describes the methodology adopted for this research and it is divided in two sub-Sections. The first Section describes the development of database followed by the statistical analysis that determines the properties of outages that have to be modeled. At the end of this Section, a probabilistic model for the occurrence and duration of outages is developed. The second Section describes the scenario generation algorithms that were developed using the Markov Chain processes and Monte Carlo Simulation. Chapter four of this thesis describes the case study taking a BC Hydro generating station as an example. The results of statistical analysis and scenario generation algorithm are presented. Chapter five summarises the main conclusions and provides recommendation for future work.

## **Chapter 2: Literature Review**

Modeling forced outage involves statistical analysis of data and application of suitable reliability methods. The purpose of this chapter is to provide background information in both these topics and summarize relevant research work on stochastic modeling of outages. Starting with a description of the database system for generating units, all operating states of a hydro power generating unit are described. Past work done in statistical analysis of hydropower unit outages is presented to show how database systems have been used to obtain specific information. The next part delves into reliability concepts as applied to forced outage studies. It includes basic concepts of reliability, hazard rate, development of reliability indices and application of Markov process to model forced outages. Some studies are presented to show how researchers have modeled forced outages for different purposes. Finally, the key takeaways from this literature review are summarized to help in designing appropriate methods for a stochastic representation of forced outages for the BC Hydro system.

### **2.1 Database Management System for Generating Unit States**

In system operations data is collected primarily for two purposes: for the assessment of past performance and for the prediction of future system performance. Both purposes are complimentary in nature. To accurately predict future performance or improve upon past performance, it is essential to transform past experiences into suitable models. This makes data collection a crucial part of the whole exercise. The methodology of future predictions and data collection evolve simultaneously and this process is iterative. The database has to be comprehensive enough to reflect the needs of the predictive methodology, at the same time, it's

important for utility companies not to collect unnecessary data which can lead to computation of irrelevant statistics (Billinton and Li 1994).

To standardize common reporting procedure and realize a practical database, all major electric power utilities in Canada and USA, have developed a uniform data collection and analysis system. In Canada, comprehensive generating unit outage databases are maintained by the Canadian Electricity Association (CEA) and in USA, by the North American Electric Reliability Council (NERC). The CEA's Equipment Reliability Information System (ERIS) and the NERC's Generating Availability Data System (GADS) contain a wealth of important information. GADS maintain complete operating histories on more than 7,700 generating units, representing over 90 percent of the installed generating capacity in the United States and Canada (NERC 2015). The objective of this reporting system is the identification of the state of a generating unit at each hour throughout the year. ERIS started collecting data in 1977. It is structured into 3 basic components: the Generation Equipment Status Reporting System, the Transmission Equipment Outage Reporting System and the Distribution Equipment Outage Reporting System (R. Billinton 2001). For the purposes of this thesis, only the generating equipment status is relevant for dealing with generating unit availability.

## **2.2 Unit States in the BC Hydro Unit Status Record**

BC Hydro maintains the unit status for each unit since 1977 in its Unit System Recording (USR) database, in line with CEA's and NERC guidelines. This database has information of all the operational status changes (state codes) accurate to one-minute interval for all BC Hydro's generating units since 1977. Originally, this status record was only needed to satisfy the Canadian

Electric Association's (CEA) reporting requirements. Today, the record and the statistics available from it are also required for a growing number of internal and external applications including: British Columbia Utility Commission (BCUC) rate hearings, monthly executive meetings, BC Hydro capital planning, and various other internal and external reporting purposes (Unit Status Recording Primary Reference, 2009).

USR is the record that identifies a change in unit status. There are 30 state codes to cover various operating scenarios. Broadly, these states are: Available, Forced Outage, Maintenance Outage, Planned Outage and Forced Extensions of planned and maintenance outage. The following sub-Sections describe the major operating states and tabulate different state-codes, under which a unit can be classified in USR.

### **2.2.1 Available States**

Available States refers to operating and not operating states of units when a generating unit is not in an outage state. The gross maximum electrical output (in megawatts) which a generating unit has been designed for and/or shown by acceptance testing to be capable of producing continuously is called the Maximum Continuous Rating (MCR). Units can either generate energy at full MCR, reduced MCR or run in synchronous-condense (SC) mode.

A derating is a reduction below MCR of generating unit capacity more than 2% of its MCR resulting from a component failure or other condition. A derating may be forced or scheduled. If some condition requires that the generating unit to be derated at once or, as soon as possible up to and including the very next weekend, then it is called forced derating. Reduction in MCR resulting

from a planned outage of a piece of equipment is called scheduled derating. Synchronous Condense (SC) state refers to the case when a hydropower unit is synchronized with system and operated as a motor, spinning freely in air and drawing power from the grid to provide reactive power/voltage support. The operating units may run at its Maximum Continuous Rating or reduced capacity (de-rated) or run in condensing mode.

Table 1-Table 6 summarizes the different available states as classified by CEA (Canadian Electricity Association 2016):

1) Available On-Line States:

**Table 1: Available On-Line States**

<b>State</b>	<b>Name</b>	<b>Definition</b>
11	Generating	The generating unit is spinning and synchronized with the system and is capable of operating at MCR under normal operating procedures. (Breaker Closed)
11-1	Condensing (turbine coupled)- <i>Hydro only</i>	
11-2	Speed-No-Load (SNL) Breaker Closed	
12	Generating under Forced Derating	The generating unit is spinning and synchronized with the system but not capable of carrying its MCR due to a Forced Derating being in effect. (Breaker Closed)
12-1	Condensing under Forced Derating - <i>Hydro only</i>	
12-2	SNL Breaker Closed under Forced Derating	
13	Generating under Scheduled Derating	

13-1	Condensing under Scheduled Derating - <i>Hydro only</i>	The generating unit is spinning and synchronized with the system but not capable of carrying its MCR due to a Scheduled Derating being in effect. (Breaker Closed)
13-2	SNL Breaker Closed under Scheduled Derating	

2) Available Off-line States:

This represents the set of cases similar to available states with the only difference that the unit is not electrically connected to the transmission system for economic reasons.

**Table 2 Available Off-line States**

<b>State</b>	<b>Name</b>	<b>Definition</b>
14	Available but not Operating ( <b>ABNO</b> )	The generating unit can carry its MCR but is not being operated to supply system load. The unit may be spinning with the breaker open.
14-1	Condensing (Turbine Un-coupled) - <i>Thermal Only</i>	
14-2	SNL Breaker Open - <i>Hydro only</i>	
15	ABNO under Forced Derating	The generating unit can deliver only part of its MCR due to a forced derating but is not being operated to supply system load.
15-1	Condensing under Forced Derating - <i>Thermal Only</i>	
15-2	SNL Breaker Open under Forced Derating - <i>Hydro Only</i>	
16	ABNO under Scheduled Derating	The generating unit can deliver only part of its MCR due to a scheduled
16-1	Condensing under Scheduled Derating - <i>Thermal Only</i>	

16-2	SNL Breaker Open under Scheduled Derating - <i>Hydro Only</i>	derating but is not being operated to supply system load.
------	--	---

### 2.2.2 Forced Outage State

For the purpose of systematic record keeping, forced outages are defined as the occurrence of a component failure or other condition which requires that the generating unit be removed from service immediately or up to and including the very next weekend. The sub categories of forced outages are: Sudden Forced Outage Unit Trip, Immediately Deferrable Forced Outage, Deferrable Forced Outage, and Starting Forced Outage. These are described in more detail in Table 3 below.

**Table 3 Forced Out States**

<b>State</b>	<b>Name</b>	<b>Description</b>
21-1	Sudden Forced Outage. Unit Trip	The occurrence of a component failure or other condition which results in the unit being automatically or manually tripped.
21-2	Immediately Deferrable Forced Outage	The occurrence of a component failure or other condition which requires that the unit be removed from service within 10 minutes
21-3	Deferrable Forced Outage	The occurrence of a component failure or other condition which requires that the unit be removed from service from 10 minutes up to and including the very next weekend.
21-4	Starting Failure Outage	The unsuccessful attempt to bring a unit from a shutdown state to synchronism with the electric system within a specified time interval.

Unit outages can occur due to several reasons. The ERIS database system mandates the assignment of a 5-digit component cause code that identifies the specific component or part thereof that caused the unit outage. Many of these outages are organized together under major groupings. These major groupings with examples are shown below (Canadian Electricity Association, 2016):

- a) Buildings and Structures: draft tubes, channels, tunnels, sluice gates, penstock, etc
- b) Power Generation Facilities (Turbines): runner, hub, blades, bearing, wicket gate, etc
- c) Power Generation Facilities (Generators): stator, core iron, slip rings, commutator
- d) Electrical Power System: bus duct, cable, switching equipment, transformers
- e) Instrumentation and Controls: governor, excitation equipment, power output system
- f) Plant Auxiliary Processes and Services: fire protection, water depressing system,
- g) External Conditions: storms, floods, fire, staff shortage, transmission line outage.

### **2.2.3 Maintenance Outage**

A Maintenance Outage (MO) is defined as an outage that can be deferred beyond the end of the next weekend. A MO can occur any time during the year, has a flexible start date, may or may not have a predetermined duration, and is usually much shorter than a planned outage. The removal of a generating unit from service is generally done in order to perform work on specific components which could have been postponed past next weekend. This work is done to prevent a potential forced outage and which could not be postponed from season to season.

**Table 4 Maintenance Outage State**

<b>State</b>	<b>Name</b>	<b>Description</b>
24	Maintenance Outage	Event that can be deferred beyond the next weekend but not beyond the season

#### **2.2.4 Planned Outage**

A Planned Outage is the removal of a generating unit from service for the inspection and/or general overhaul of one or more major equipment groups (e.g. five-year turbine overhaul, annual boiler overhaul). Scheduling of these planned outages for maintenance is a complicated task in itself. Constraints on resources, such as working crews, hours, and budget, have to be taken into account apart from assessing the impact of the scheduled outages on operations (Dalal et al. 2018). After completing the scheduling process, the outage period is assigned well in advance and can be postponed from season to season if needed. From a unit availability point of view, planned outages are deterministic in nature, although there could be some uncertainty in start or end dates and in duration of the outage.

**Table 5 Planned and Upgrade Outages**

<b>State</b>	<b>Name</b>	<b>Description</b>
25	Planned Outage	Removal of unit from services can be deferred beyond the next the season
26	Upgrade Outage	Removal of a generating unit from service for prolonged work to make modifications that will alter its performance beyond the original design and/or provide life extension through rehabilitation

### 2.2.5 Forced Extensions

Sometimes the duration of Maintenance and Planned Outages are unexpectedly exceeded during repair work. The period of time exceeded beyond the originally scheduled period is called Forced Extension. The different types of Forced Extensions are described in Table 6.

**Table 6 Forced Extension of Outages**

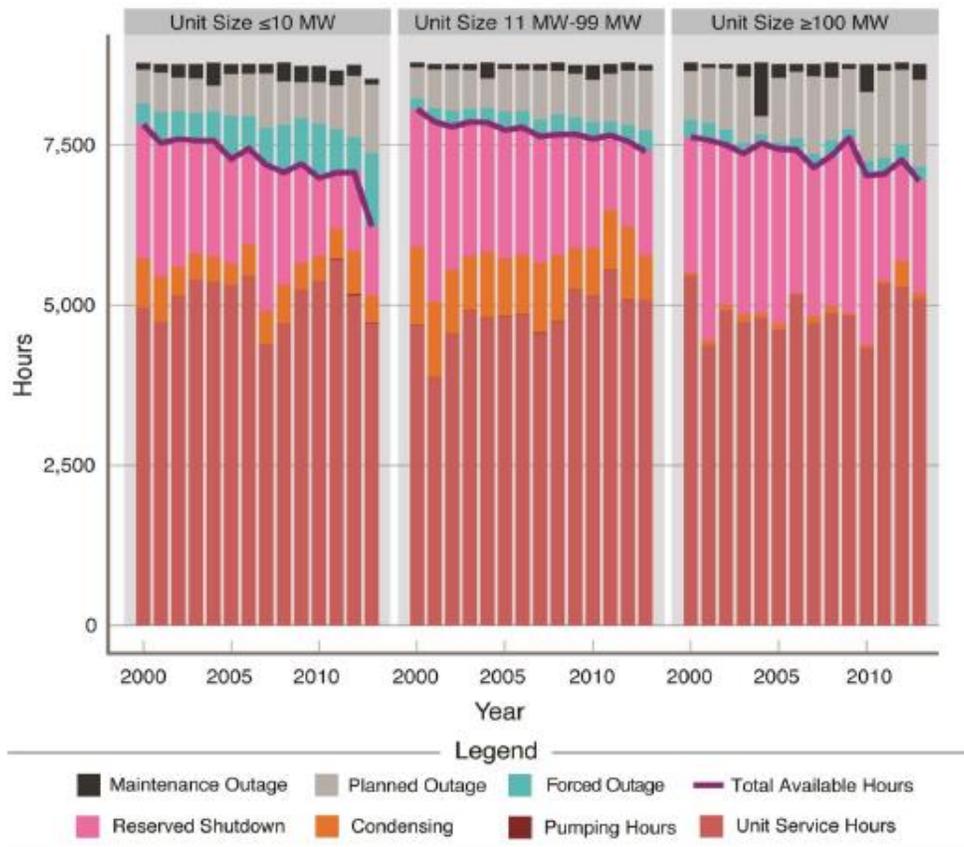
<b>State</b>	<b>Name</b>	<b>Description</b>
22	Forced Extension of a Maintenance Outage	The generating unit has an outage resulting from a condition discovered during a maintenance outage which has forced the extension of the maintenance outage.
23	Forced Extension of a Planned Outage	Outage resulting from a condition discovered during a planned outage.
27	Forced Extension of an Upgrade Outage	Outage resulting from a condition discovered during an upgrade outage which has forced the extension of the maintenance outage.

### 2.3 Statistical analysis of unit states

Data on unit states is collected for two main purposes: first as an assessment of past performance and second as a prediction of future system performance (Billinton and Li 1994). The various sub classification presented in the previous Section are required to obtain operating statistics on any generating unit(s). These statistics are used to make decisions both internally for financial and operational purposes, as well as externally for regulatory reasons and comparison with other electric utilities. The unit operation statistics obtained from USR are used by various departments at BC Hydro for a range of purposes. A water licensing unit may focus on water levels and would

focus on discharge from turbine irrespective of power generation. So, they would not differentiate between on-line and off-line states as long as the turbine is releasing water. The maintenance department may use USR data to optimize the duration and timing of planned upgrades based on past statistics of planned and maintenance outages. The department of system optimization would be concerned about the availability of units without differentiating between available-operating (AO) and available-but-not-operating (ABNO) states. Some examples are presented to show how these databases have been used by researchers to gain various insights about unit operating statistics.

**Figure 1** shows the average hydropower operational status of generating units reporting to NERC. This chart represents cumulative hours spent in specific states by hydropower units reporting to GADS (USA) from 2000 to 2012. The database for this analysis comprised of units with different rated capacities. It has 258 units that have a capacity below 10 MW, 520 units with capacities between 10-99 MW and 126 units with capacities above 100 MW (Oak Ridge National Lab 2014).

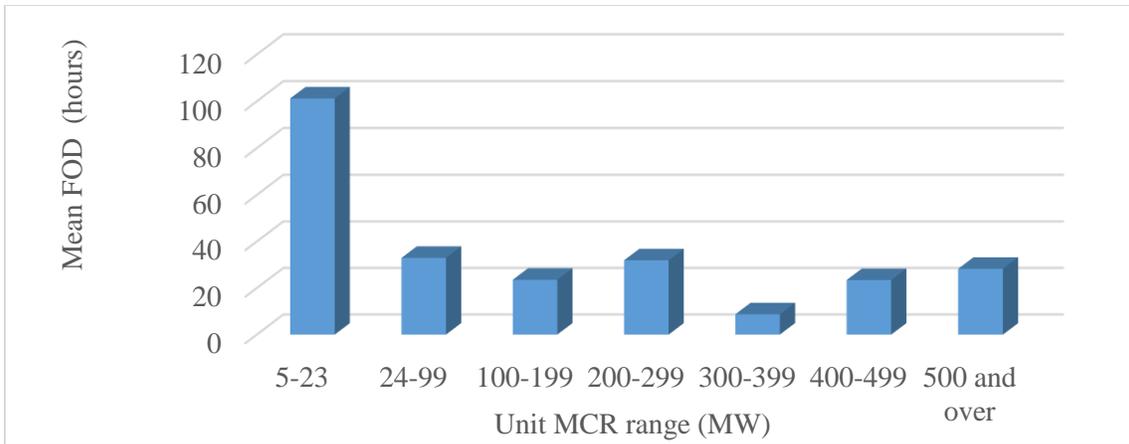


**Figure 1 Average Hydropower operational status  
(Hourly breakdown by unit size classes of units reporting to NERC)  
Source: (Oak Ridge National Lab 2014)**

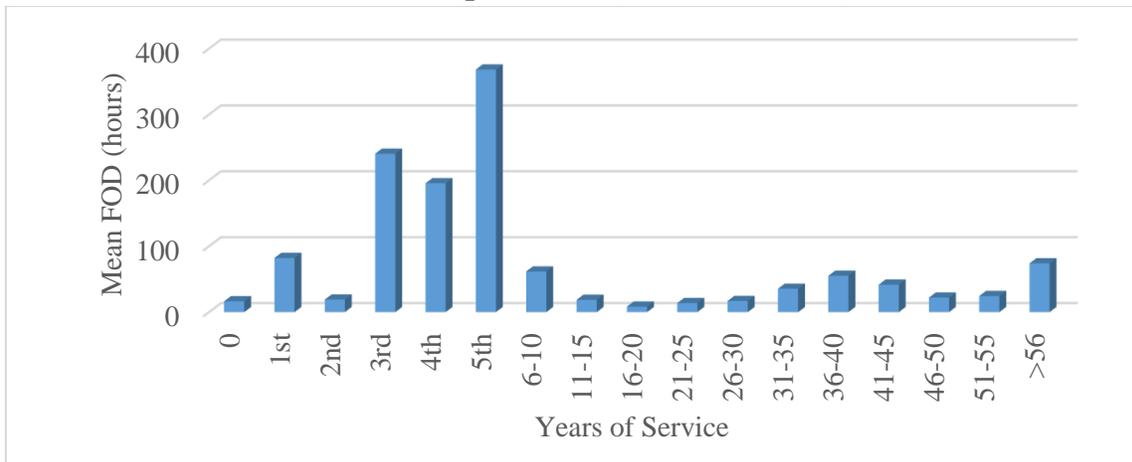
The authors distilled all the operating states into 7 broad categories (4 active and 3 outage states). The active states consist of unit service hours (unit is synchronized to grid), pumping hours (turbine-generator is used as pump/motor), condensing (unit is spinning freely in air to provide reactive power/voltage support) and reserved shutdown hours (Available But Not Operating-ABNO). The outage states consist of Forced Outages, Planned Outages and Maintenance Outages. The sum of hours spent in the active states constitutes the total number of available hours (represented as a purple line in **Figure 1**).

The report concluded that there is a tradeoff between planned and forced outages. For larger units, a significant portion of total outage duration is spent in planned outage state. This may be a reason why larger unit experienced the lowest number of forced outage hour in comparison to smaller units, which experienced an increasing trend in unplanned outages during the last 13 years. Forced outage of a very large unit can be very expensive; therefore, it would make sense for plant operators to invest more in regular inspections to avoid forced outage. This is just one example of how these databases can be analyzed to answer specific questions.

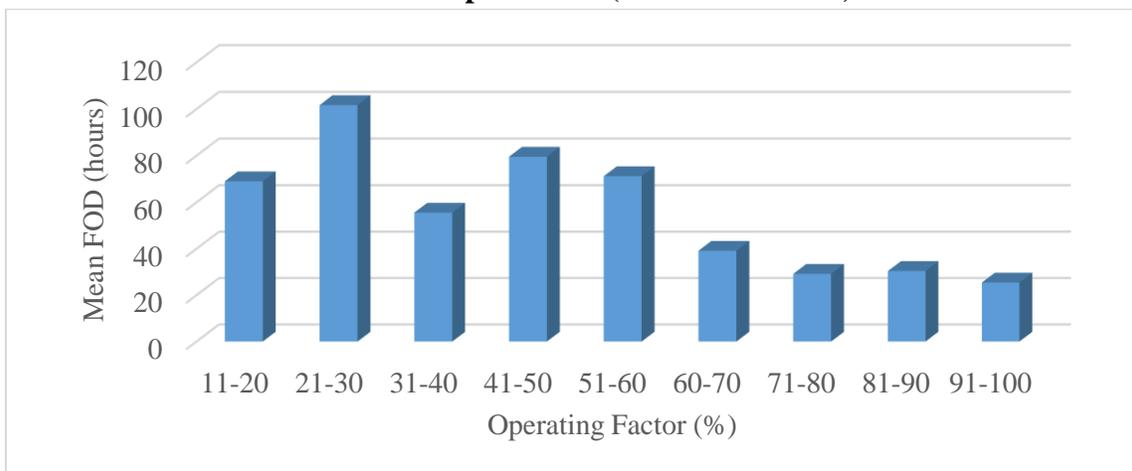
Another example of hydraulic unit outage analysis is found in the generation equipment status report (ERIS-CEA 2012). This report has data on operational performance and forced outages of 456 hydraulic units from 2003-2007. Here the cumulative forced outage duration for all units was 8.4 years with mean forced outage duration of around 46 hours per outage event. This paper further defined Incapability Factor (ICbF) as the ratio of Total Equivalent Outage Time, in hours, to total number of hours the unit is in service times 100. The ICbF due to unplanned outages (forced outages, forced derates and maintenance outages) for the 456 units was computed to be ~2.4% and due to planned outage was computed to be ~6.3%. The report analyzed forced outage information in many different ways. The charts of mean forced outage duration (FOD) under different MCR ratings, years in service and operating factors (unit loading in percentage) are shown in Figure 2 to Figure 4.



**Figure 2 Mean FOD for units with different MCR ratings**  
**Source: Adapted from (ERIS-CEA 2012)**



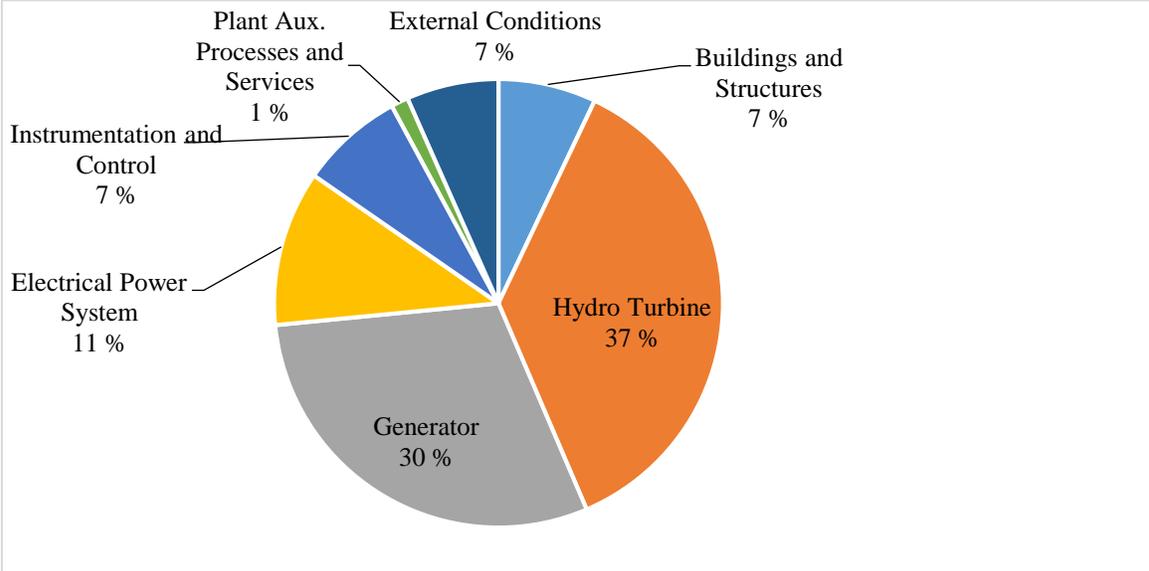
**Figure 3 Mean FOD for units at different age**  
**Source: Adapted from (ERIS-CEA 2012)**



**Figure 4 Mean FOD for units operated differently**  
**Source: Adapted from (ERIS-CEA 2012)**

Figure 2 shows that bigger units have relatively shorter duration per outage and are brought back into service quicker than smaller units. Figure 3 suggests there are higher outage impacts in the first years when a new unit is installed which then stabilizes during the useful life of the unit and starts increasing again as they age. Figure 4 shows units mean FOD for different operating factors. Operating factor is the ratio of actual output of generating unit to the total potential output of unit if it was operated at full capacity for the entire duration. It was observed that units that have a higher operating factor undergo less severe outages. Units that have a lower Operating Factor are operated less frequently and have reported a much higher mean FOD. This can be because frequent start-stops of unit can increase chances of starting failure.

The report also analyzed the contribution of major components to hydraulic unit ICbF due to unplanned outages, as shown in Figure 5. It can be noted that component failure related to turbine and generator are the main cause of forced outage in hydro power units.



**Figure 5 Major Component contribution to hydraulic unit ICbF due to unplanned outages**  
**Source: (ERIS-CEA, 2012)**

The CREATE report (Simonoff et al. 2005), statistically analyzed electricity outages over the period of January 1990 to August 2004 using data obtained from NERC. The authors looked at seasonality and annual trend of several attributes such as: number of incidents, average outage duration, capacity lost, and customers affected. They used three-month averages to analyze seasonal characteristics and twelve-month averages for annual trends. Winter was defined as December through February, spring as March through May, summer as June through August, and autumn as September through November.

The report found that the number of incidents is increasing annually at a rate of 8.3%. However, this increase in number of events did not translate into duration of outage hours. From 1990 to 1993 durations of outages were getting shorter on average but this trend changed in the mid-1990s when the average duration started to increase, and this trend became more pronounced after 2002. Seasonal analysis suggested there were 65-85% more incidents in summer than the other seasons. The winter, spring, and autumn estimated rates were found to be similar to each other, with autumn having a rate that is slightly lower. This was presumably attributed to weather effects such as: snow and ice in the winter, thunderstorms in parts in spring and summer, and most importantly intense heat with corresponding air conditioner use in the summer. Autumn had lesser load and seasonal weather affects. However, no significant seasonal affect in duration of outages was observed as there was in number of incidents of outages.

Apart from these reports, researchers have also emphasized the need for statistical analysis of operating states to understand the characteristics of a system before selecting a modeling approach for outages. (Koval and Chowdhury 1994) highlighted the importance of statistical analysis of unit

operation patterns and its relevance in the assessment of reliability of units. In their work, they argued that past patterns can help in answering relevant questions such as are generating units behaving similarly? Are there distinct outage patterns for each one of them? Are there more outages in one season versus others? What is the best probabilistic model to represent repair duration and failure interval, among others?

The main takeaway from these past works is that analysis of outage starts with collection of relevant data on unit operating states from standard databases. A database has to be defined for specific need of the modeler and then statistical analysis has to be carried out to find the properties relevant for modeling. Once the statistical properties of outages are obtained then suitable mathematical approaches can be selected for modeling the uncertainty in forced outages. In the next Section, some of mathematical tools found in the literature are discussed in detail.

#### **2.4 Reliability analysis in Forced outage modeling**

Reliability analysis methods have been increasingly used in academia and industry mostly due to availability of historic database and low cost computational power. In analysis of electric power systems, Fault Tree Analysis (FTA) and Event Tree Analysis (ETA) have been widely used to quantify major factors influencing total system reliability (Endrenyi 1979; Billinton & Allan, 1992; Zhou, 2015).

ETA is a bottom-up analysis approach that defines all potential accident sequences/chains associated with particular initiating events. (Begovic, Perkel, & Hartlein, 2006) obtained reliability data of generating units and used Monte Carlo Simulation to quantify system reliability. On the

other hand, FTA is a top-down failure analysis where a system failure event is first identified and then all combinations of component failures are analyzed that can cause that system failure event. (Van Casteren, Bollen, & Schmieg, 2000) used FTA methods in forced outage analysis where they identified system failures and quantified it using component outage values. Reliability Block Diagram (RBD) is a diagrammatic method to visualize how component reliability contributes to the success and failure of complex systems.

Forced outages are essentially failure events so they have been investigated by researchers and utilities using concepts of reliability engineering (Barroso and Conejo 2006; Roy Billinton and Allan 1992; Roy Billinton and Ge 2004; Bornak 2013; Cepin 2011; Curley 2013; Finger 1979; Scully et al. 1992). There are two main reasons of why reliability analysis has been employed for forced outage related studies. First, to compute operating reserves required for reliable system operations given outages are uncertain events. It was realized that use of deterministic criteria, such as “peak load percentage” or “loss of largest unit”, could not quantify the risk of supply shortages in the system and also failed to quantify the worth of added reliability by using reserves. Therefore, stochastic methods have to be explored to quantify reliability of generating resources (Prada 1999). The second reason was to quantify the optimum cost of preventive maintenance by analyzing the trade-off between the cost of maintenance and the cost of plant unavailability due to forced outages for different cases of failure and repair rates.

To better understand of the research work in the literature, some of the basic concepts and traditionally used reliability indices are discussed with an aim to attain required background in reliability analysis of power systems.

### 2.4.1 Basic concepts of Reliability Analysis

The classical index to measure reliability has been the probability of not failing. Reliability, as the characteristic of an item, is often defined by the probability that it will perform a required function under stated conditions for a stated period of time. Simply put, it is the probability of system staying in the operating state without failure. Mathematically, reliability can be expressed as:

$$R(t) = P(T > t) \quad \text{Equation 1}$$

where  $T$  is a non-negative random variable denoting the failure time, and  $t$  is the designated period of time given the operating conditions. If  $f(u)$  is the probability density function (PDF) for the failure time  $T$ , the reliability function can be calculated as:

$$R(t) = \int_t^{\infty} f(u) du \quad \text{Equation 2}$$

The reliability function  $R(t)$  can also be defined as the complement of the cumulative distribution function (CDF) for failure probability,  $F(t)$ , corresponding to  $f(u)$ :

$$R(t) = 1 - F(t) = 1 - \int_0^t f(u) du \quad \text{Equation 3}$$

The reliability function gives the inverse cumulative probability distribution of failure time for components. Failure time, also referred to as Time to Fail (TTF), is the duration of time the component remained in-service/ available state before going out of service. For different values of TTF,  $F(t)$  or  $[1 - R(t)]$  gives the probability of exceedance of component failing after that time.

Since generating units are a continuously operated system that can tolerate failure, a slightly modified term is used, called availability. Availability is defined as the characteristic of an item

expressed by the probability that it will perform a required function under stated conditions in a stated moment of time (Roy Billinton and Allan 1992). This refers to probability of finding the system in operating state at some time into the future, which means that either the component has not failed at all till time  $t$  or it has already been repaired after failure so that it is fully operational again at time  $t$ . The operative state of a facility at time  $t$  is assured, either because it has not failed till time  $t$  with the reliability  $R(t)$ , or because it has been functioning properly since the last repair which occurred at time  $u$ , where  $0 < u < t$ . Hence, the instantaneous availability is expressed as:

$$A(t) = R(t) + \int_0^t R(t-u)f(u) du \quad \text{Equation 4}$$

Since the instantaneous availability is not easy to evaluate, the steady state availability is more commonly used (Zhou, 2015). The steady state availability gives the long-term operational performance of a repairable system and is mathematically defined as the limit of the instantaneous availability function as time approaches infinity:

$$\lim_{t \rightarrow \infty} A(t) = \left( \frac{\sum_{i=1}^N TTF_i}{\sum_{i=1}^N (TTF_i + TTR_i)} \right) \quad \text{Equation 5}$$

where  $(TTF_i, TTR_i)$  are the alternating sequences of time to failure and time to repair being simulated, and  $N$  is the number of samples. The estimated steady state availability can be used to calculate forced outage rate of generating units (Billinton and Allan 1992).

#### 2.4.2 Hazard Rate and Bath Tub Curves

Hazard functions  $h(t)$  are another method to describe reliability of components. For a small time  $dt$ , hazard  $h(t)dt$  is the probability that a component that has survived until time  $t$  will fail in the next  $dt$  interval (Endrenyi 1979). Mathematically hazard functions are described as,

$$h(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P[t < T < t + dt | T > t] \quad \text{Equation 6}$$

where  $T$  is the lifetime of the component, after which it fails.

Endrenyi also showed that  $h(t)$  can be expressed as:

$$h(t) = -\frac{d}{dt} (\ln[R(t)]) \quad \text{Equation 7}$$

Or,

$$R(t) = e^{-\int_0^t h(t) dt} \quad \text{Equation 8}$$

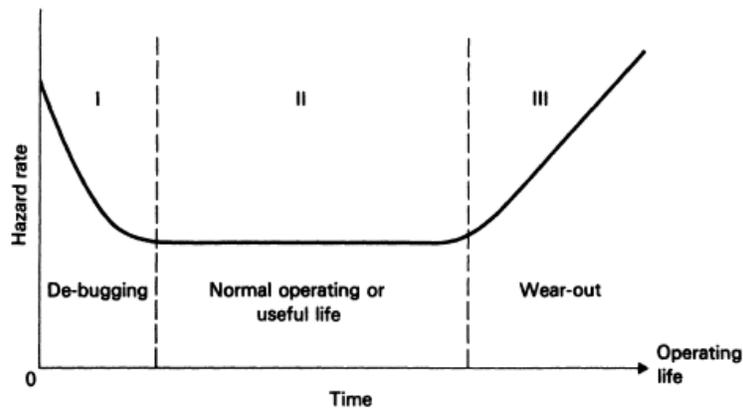
Hazard functions tend to increase, decrease or remain constant. An increasing hazard function over the lifetime of the unit indicates components are becoming more prone to failure as they age. A decreasing hazard rate indicates continuous reduction in chances of imminent failure as time passes. A constant hazard rate applies to components where the chance of failure in some time  $t$  remains the same for any other  $t+dt$ .

In reliability analysis, electrical/mechanical components are treated as either repairable or non-repairable. The life of non-repairable components lasts up to its first failure where the repair is uneconomical or infeasible. A constant hazard rate is often assumed to simplify analysis of non-repairable components. Due to this assumption of constant hazard rate, Poisson process has been used to describe the occurrence of failure events for non-repairable components which have approximately constant hazard rate (Endrenyi 1979). A Poisson process describes the probability of an isolated event (for say accidents, telephone calls, etc.) occurring a specified number of times in a given interval of time or space when the rate of occurrence (hazard rate) in a continuum of time or space is fixed (Billinton and Allan 1992). For non-repairable system, Poisson distribution

is used to find probability of single arrival of failure event. Mathematically, the reliability of a non-repairable component over a specified period  $t$ , having constant hazard rate  $\lambda$ , is given by:

$$R(t) = e^{-\lambda t} \quad \text{Equation 9}$$

On the other hand, repairable components are those that can be brought back to service after failure events. The life histories of repairable components consist of alternating operating and repair periods. Generating units come under repairable categories. They experience “failure” due to forced outages but can be brought back into service after maintenance and/or repair. Many repairable electrical components with preventive and corrective repairs exhibit a bath-tub hazard curve as shown in Figure 6 (Endrenyi 1979).



**Figure 6: Typical bath tub curve for component failure.**

**Source: Roy Billinton & Allan, 1992**

As shown in the above figure, the hazard curve has three distinct parts each with its own hazard rate and hence its own probability distribution for failure events. The first phase is called the burn-in phase or infant mortality, where the failure occurs likely due to design issues. Issues are detected and fixed in this initial period, so the hazard rate decreases rapidly. This phase is followed by the useful life period where the chance of failure is low and relatively constant. In the ‘old-age’, wear

out failures becomes the predominant cause of failure and results in an increasing hazard rate (Endrenyi 1979).

A generating unit, like any other electrical or mechanical system, can be made to remain within their useful life period for the bulk of their installed period by constant and careful preventive maintenance and corrective maintenance. Preventive maintenance is performed in order to keep the unit in a condition that is consistent with the required levels of performance and reliability. This is achieved by regularly checking all the operating systems, cleaning, adjusting, lubricating all the components, replacing the components nearing a wearout condition and checking and repairing failed redundant components. Corrective maintenance is required after forced outage when the system malfunction. Its purpose is to restore system operation as soon as possible after failure by replacing, repairing or adjusting the components which have caused interruption or breakdown of the system (Roy Billinton and Allan 1992).

### **2.4.3 Probability distributions for different hazard rates**

The mathematical representation of these different types of hazard functions is done by selecting appropriate model from the Weibull family of distributions. The advantage of using the Weibull family of distributions is that these have a specific characteristics shape and any experimental data can be fit by choosing appropriate parameters. The general form of Weibull hazard function is

$$h(t) = \frac{\beta t^{\beta-1}}{\alpha^\beta} \quad \text{Equation 10}$$

where  $\alpha$  and  $\beta$  are constants. If  $\beta > 1$ , the hazard rate increases; if  $\beta = 1$ , the hazard rate is constant and if  $\beta < 1$  less than 1 the hazard rate is decreasing (Endrenyi 1979). The reliability function (i.e., inverse distribution of failure time) for the Weibull hazard function is represented by:

$$R(t) = e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad \text{Equation 11}$$

The Gamma distribution has similar properties to those of the Weibull distribution, as it is a two-parameter distribution having a shape parameter  $\beta$  and a scale parameter  $\alpha$ . By varying these parameters, a gamma distribution can be fitted to a wide range of experimental data. The reliability function is represented by:

$$R(t) = \int_t^\alpha \frac{t^{\beta-1}}{\alpha^\beta \Gamma(\beta)} (e^{-\frac{t}{\alpha}}) dt \quad \text{Equation 12}$$

Where  $\alpha$  and  $\beta$  are constants and  $\Gamma(\beta)$  is the Gamma function defined as:

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} e^{-x} dx, \quad n > 0 \quad \text{Equation 13}$$

The exponential distribution is a special case for both Weibull and Gamma distribution functions. When the parameter  $\beta = 1$ , the hazard function is constant and is given by  $(1/\alpha)$ . Interestingly, Poisson distribution is special case of exponential which evaluates the reliability until the first failure.

The inverse distribution of failure time or reliability function is given by:

$$R(t) = e^{-\frac{1}{\alpha}} \quad \text{Equation 14}$$

While much of the theoretical work has been done on failure time and reliability functions, the repair time is simply checked for suitable fits using trial and error. (Roy Billinton and Allan 1992)

states that a lognormal distribution can be a good fit to the distribution of component repair times and hence it is finding acceptance in academia for the assessment of repairable systems. The cumulative probability distribution of repair times, using lognormal distribution is given by:

$$Q(t) = \int_0^t \frac{1}{t\sigma\sqrt{2\pi}} e\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right) dt \quad \text{Equation 15}$$

Where,  $\mu$  is the mean and  $\sigma$  is the standard deviation of values of repair times.

#### 2.4.4 Development of reliability indices for forced outages

Generating unit availability is affected by many factors other than forced outages. Equipment aging, operational changes from environmental regulations and changes of the relative priority given to different water uses in multipurpose projects are all likely contributors to steady state availability factors (Oak Ridge National Lab, 2014). To quantify the impacts of forced outages on generating units, specific reliability indices have been used by building onto the basic definition of reliability and availability.

Loss of Load Probability (LOLP) is a popular reliability index that incorporates a probabilistic approach. It represents the probability of net load on the system exceeding its available generation capacity under the assumption that peak load of each day lasts all day (Endrenyi 1979). In this method, all possible combinations of available units and unavailable units are taken to evaluate net system availability and their associated probabilities. Any load value can be checked against the calculated system capacity to find cumulative probability of that load not being met. This measure does not exactly stand for loss of load but rather a deficiency of installed available capacity. A modified form of LOLP is Loss of Load Expectation (LOLE) which gives the duration

of time, rather than of a probability measure in percentage, for which a certain load would not be met (Cepin 2011). LOLE is used to analyze outages at the consumer's end, for example 5 hours of load shedding in 1 year. In order to accommodate de-rated states and non-operating states for peaking units, specialized indices such as the Derating Adjusted Utilization Forced Outage Probability (DAUFOP) have been developed (Wang, Ramani, and Davies 2004). This index gives the probability of a generating unit (including de-rated states) of not being available when needed (Roy Billinton and Ge 2004).

The problem with LOLP, LOLE and DAUFOP is that they do not give any indication of the frequency of occurrence or the duration for which an insufficient capacity condition is likely to exist. It can provide a probability measure associated with every possible plant capacity due to individual unit outages. For example, for a plant with 2 generating units with capacities of 100 MW each, the plant capacity can be 0, 100 or 200 MW depending upon how many units are forced out. Using the above indices, such as LOLP, a probability measure can be assigned for each increment of plant capacity to quantify the chances of the load not being met. However, in a multi-reservoir system, the loading decisions on individual unit are based on local inflows and the marginal cost of water in the reservoir that are decided by system optimization studies. Simply put, LOLP and similarly derived indices provide a lumped measure of reliability and does not answer specific questions like how many outages occurred in a particular period, or what was the expected duration of each event in a year.

Hydro power companies use simulation studies and optimization models to plan and operate multi-reservoir systems. To account for the uncertainty in system capacity, scenarios of outages are

required that can realistically represent the stochastic nature of time between failures and duration of outages as observed in historic data. Frequency and duration of outage events contain these additional physical characteristics and hence provide valuable information for the purposes of modeling (Billinton & Li, 1994). This method uses discrete unit states and transition probabilities to find ‘when’ and for ‘how long’ the unit will remain in a particular state (Roy Billinton and Allan 1992). The frequency of encountering a certain state is the probability of being in that state multiplied by the rate of departure from that state (Cepin 2011).

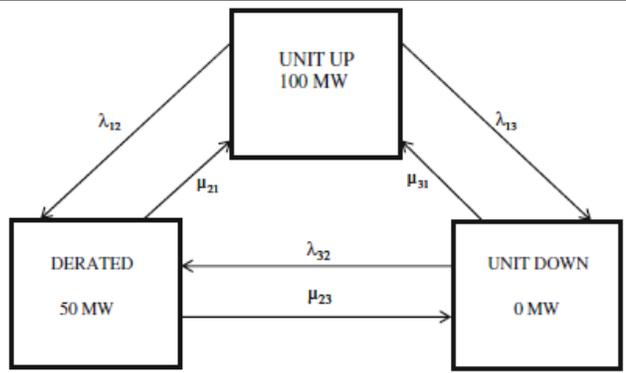
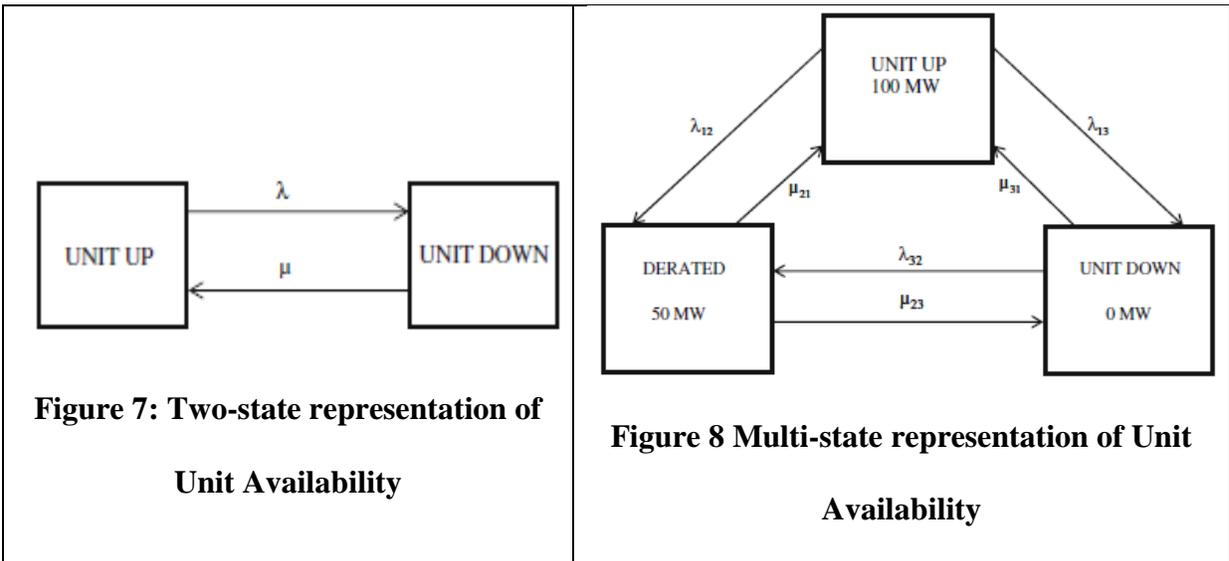
To summarize, LOLP and related indices provide the probability of load exceeding available generation. However, from an operational modeling perspective, the average number of occurrences and duration of interruptions per time period are needed and are important. Basic Markov modeling process has been applied to generate scenarios and obtain reliability indices of frequency and duration. The application of a Markov process for assessment of reliability indices is explained in the following sub-Section.

#### **2.4.5 Use of Markov process in Forced Outage modeling**

Generating units can be run at full capacity, can run at partial capacity when it is in derated state or can have 0 capacity during forced outages. Markov process is often used to describe the process of a system changing its state. For a Markov process to be applicable, the system behavior should have two characteristics: a lack of memory and stationarity. For generating units, these criteria are generally assumed to be true (Roy Billinton and Ge 2004). Generally, in a Markov process the probability of being in one state at time step  $t+1$ , depends on the state of the system at time  $t$ , but not on the states occupied earlier (Finger 1979). This change of state is governed by transition

probabilities which in case of outages, take the form of failure rate ( $\lambda$ ) and repair rate ( $\mu$ ). Failure rate is the number of failures per unit available hour and repair rate is number of repairs per unit unavailable hour.

Depending on the research question, either two state models is used (Figure 7) or multi-state discrete capacity models are used (Figure 8)



Multi-state discrete states have been used to describe the uncertainty in generation for a distributed system having wind and solar components as the system capacity is affected by sudden changes in weather as well as from individual component failures (Yan-fu Li and Zio 2012; Yan Li, Cui, and Lin 2017). In case of hydropower or thermal generating units, multi-state approach can be used to model different de-rated states. However, it is more common to adjust the duration of de-rated state into equivalent full forced outage (Roy Billinton and Ge 2004).

The application of Markov method requires the computation of transition probabilities in terms of failure and repair rate. The standard method of quantifying the transition probabilities for generating units was discussed by (Endrenyi 1979) who defined availability of any power system component undergoing normal repair and preventive maintenance. Endrenyi defined duration of outage events as Time to Repair (TTR) and duration from the start of unit operation to unit outages as Time to Fail (TTF). Endrenyi further showed that if the rate of occurrence of outage is constant, then in a long-time horizon the probability density function of TTF is given by an exponential distribution.

The parameter that characterizes this exponential distribution is the failure rate  $\lambda$  or Forced Outage Rate (FOR). FOR is defined as total number of outages divided by total duration of time the unit remains in an operating state. FOR is inverse of Mean Time to Failure (MTTF). Hence,

$$\lambda = \frac{1}{MTTF} \quad \text{Equation 16}$$

The probability density function of TTF is then given by the expression

$$f(t, \lambda) = \lambda e^{-t/\lambda} \quad \text{Equation 17}$$

Assuming a constant repair rate  $\mu$ , given by inverse of Mean Time to Repair (MTTR)

$$\mu = \frac{1}{MTTR} \quad \text{Equation 18}$$

The probability density function of TTR becomes

$$f(t, \mu) = \mu e^{-t/\mu} \quad \text{Equation 19}$$

The use of exponential distribution makes the process a homogenous Markov process. A Markov process is called homogeneous if transition probability from  $t_n$  to  $t_{n+1}$  is independent of 'n'.

Mathematically this can be written as:

$$P[ X_{n=j} / X_{n=i} ] = P[ X_{1=j} / X_{0=i} ] \quad \text{Equation 20}$$

The use of an exponential distribution enables analytical solutions for a homogenous Markov process and so it has been used widely in power system reliability application. There have been other good reasons in the past to assume exponential distribution for values of TTF. Limitations of the availability and quality of data made it difficult to verify an exact distribution. The parameter of exponential distribution is the failure rate which is simply the inverse of mean time to failure. And most importantly, the assumption of constant failure rates and hence exponential distribution simplified complex problems in power system reliability (Roy Billinton and Allan 1992).

A Markov process can be used to find the state of the unit in every time step in a time series. An aggregation of these states creates a realistic scenario of outage events. Hence the required reliability indices such as frequency and duration of outage can be computed from the generated data. The generated data can be used as input to energy planning and simulation models, the details of which are provided in the methodology Section.

It is important to mention here that the use of an exponential distribution is not an inherent limitation and the Markov process techniques shown above applies equally well to any other distributions that would fit time to failure values. The essential difference is that the integration for some distributions is rather complex or even impossible analytically and additional numerical integration techniques must be used. This can make the evaluation tedious or too difficult for hand

calculations and computer solutions are then required. (Barroso & Conejo, 2006, Hall & Ringlee, 1968) used exponential distributions as described above to model both TTF and TTR. The use of exponential distribution is specifically attributed to the fact that it can make complex models very tractable. However, some researchers have argued against use of exponential distribution and proposed Weibull distribution as it is bell-shaped and provides more modeling flexibility (Anderson and Davison 2005; Van Casteren, Bollen, and Schmieg 2000)

Additional difficulty in using exponential distribution also arises from the use of MTTF as parameter. Ideally, TTF should be computed from the moment the unit begins to operate to the moment it fails (Roy Billinton and Allan 1992). This definition tends to subtract the periods of planned outages and reserve shut down states from TTF durations. However, when the scenario generation algorithm is to be run, the modeler has no information on how the unit would be loaded. Forced outage scenarios are supposed to an input in simulation models and Unit commitment is an output of those models. Hence defining TTF from start of unit loading to beginning of forced outage is not useful to obtain a sequence of outages that is supposed to be input for simulation studies.

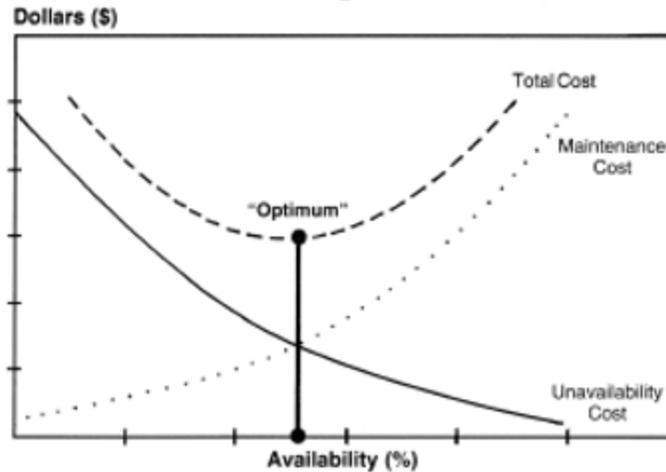
Also, assumptions of exponential distribution for Time to Repair have not been found to be realistic (Roy Billinton and Allan 1992). Due to unavailability of parts or manpower, certain repair many take very long time thereby making lognormal distribution a more suitable choice (Anderson & Davison, 2005).

## **2.5 Research literature on Modeling of forced outages**

Reliability analysis and scenario generation of forced outages for generating units have been carried out in industry and academia for two major purposes. One is to quantify the cost of preventive maintenance by analyzing the tradeoff between cost of maintenance and cost of plant unavailability due to forced outages at different levels of failure and repair rates (Begovic, Perkel, and Hartlein 2006; Binder et al. 1991; Das and Wollenberg 2012; Parrish 2015; Prada 1999; S. Ryan and Mazumdar 1990). The other is to compute operating reserves required to meet the load based on reliability criteria specified by regulators (Roy Billinton & Ge, 2004; Boomsma, Kristoffersen & Denmark, 2008; Bornak, 2013; Scully et al., 1992). Some past works are reviewed in the following Sections.

### **2.5.1 Studies on cost of forced outage and preventive maintenance**

(Binder et al. 1991) presents top down methods of analyses for predicting unit availability via different case studies of electricity utility firms. The authors argue that in bottom-up analyses, retrofits are identified, and most cost-effective repairs are conducted sequentially until the funds are exhausted. On the other hand, in top-down analyses, historical trends of unit availabilities along with their overall spending are used for benchmarking unit's performance. In top down analyses, different maintenance scenarios are examined to obtain a trade-off curve between availability and cost of maintenance as shown in Figure 9 .



**Figure 9 Trade off curve for Total cost of unit maintenance  
(Binder et al. 1991)**

In order to increase system availability, the cost of maintenance has to increase. Optimum cost is that point which minimizes the sum of cost of unavailability and cost of maintenance. In the report, the authors reviewed the methods used by utility firms to quantify cost of unavailability. In one case study for Houston Lighting and Power Company (HL&P), they looked at impacts of upgrade spending, unit aging and plant load on unit availability. They developed a forced outage model based on the GADS database to compute duration of forced outages every year. They used regression analysis and the key explanatory variables for annual forced outage hours were found to be annual service hours, service hours per start and number of starts.

In another case study, Southern Company Services used records of 85 fossil fuel units to find the factors affecting availability, using multiple regression on their database. They found that forced outage rates of current year were impacted by: previous year's forced outages, current year's planned outages and current year's spending on maintenance, among others.

(NERC 1992) conducted a study on coal-fired thermal plants using planned and forced outage data from 1982-1988. It concluded that there are increased chances of forced outage in the week after a long duration planned outage due to boiler tube leaks and turbine vibrations. The study showed that utilities can reduce impacts of forced outages after long planned outages by looking at historical records of specific component failures. While the results may not hold for hydro turbine units, the study emphasized that it is important to understand the behavior of forced outages following planned outages.

(Das and Wollenberg 2012) investigated the financial risk in day-ahead market associated with forced outages and how that risk varied with change in bids and location of generators. To compute the Value at Risk (VaR), the authors looked at a random outage scenario after the generators have been scheduled. The work focused on computing costs of outage for low bids, high bids and normalized bids to help in investment decisions. Outages were simulated just by considering probability of failure of a generator every hour, in a 24-hour period.

### **2.5.2 Forced Outage modeling for reserve computation**

(Finger 1979) developed a simple 2 state Markov process to model forced outages and used Loss of Load Probability (LOLP) to quantify reliability of electric power system due to load demand and generator operating characteristics. The power output of the plant was modeled as a Markov chain with parameters being mean TTR and mean TTF. The generated plant output and load demand were analyzed to derive equivalent demand curve that shows the load demand not being met (LOLP) for different possible cases of failure rates.

(Begovic, Perkel, and Hartlein 2006) used individual component probability of failures to obtain system reliability. Past historical data was used to obtain parameters of Weibull distribution for performance of the overall system. Monte Carlo Simulation technique was used to randomly sample and synthesize failure events. The authors state that use of historic data to quantify parameters of distribution is more robust approach and use of Monte Carlo simulation aids in extracting confidence intervals around the forecasted parameters.

(Roy Billinton and Ge 2004) described the IEEE proposed four-state model and modified it to make it more realistic based on historic data and modeling using transition probabilities. He states that the basic two state model, with one state being available and another being forced out, is good for base load units but not for units used for peaking loads. Peaking units have large shut down time and thus simple 2 state models give unreasonably high forced outage rates. Four-state representation of unit characteristics overcomes: In service state, Reserve shut down state, Forced out needed and Forced out not needed state. The authors argue that the IEEE 4 state model does not capture transitions from the reserve shutdown state to the forced out but not needed state. They argue that unit operating characteristics should be analyzed from historic data for computation of transition rates and proposed a modified 4 state model to capture all possible transitions. A Markov process was used to obtain probability of failure. DAUFOP was used as a reliability measure because it provided practical and realistic unit reliability indicators for peaking units. (Rondla 2012) similarly developed 2-state and 4-state models to differentiate forced outage rates in base load and peaking units.

(Scully et al. 1992) used a semi-guided Monte Carlo simulation method to obtain scenarios of forced outages of generating units which are ‘statistically balanced’ at the system level. A 2-unit system was considered with a given failure and repair rate. Since there are two units in the system, the system capacity can be 100%, 50% or 0% depending on how many units are out. The authors first considered a simple two state Markov process to find the unit state in daily time step. Next, they developed a model to guide the selection of forced outage periods such that outage-periods are statistically reasonable from a ‘system’ standpoint. The difference in this method, from that of simple Markov process, is that before scheduling the forced outage in a period the program checks the total capacity on outage for each day in the period. If too much outage capacity is scheduled in any day of the period, the draw is considered biased. That particular random draw of outage duration is discarded and loop back to obtain a new outage period. The authors claimed to obtain statistically balanced forced outage schedules and faster convergence of Monte Carlo iterations required to produce reliable results. They further showed the use of this method to quantify benefits of reducing forced outage rates at system level.

(Van Casteren, Bollen, and Schmiege 2000) used concepts of Fault Tree Analysis (FTA) to derive system state probabilities using component state durations and transition probabilities. They further compared two reliability assessment methods: homogenous Markov model using exponential distribution and a Semi-Markov model using Weibull distributions. They argued that use of Weibull distribution is more realistic and can be made mathematically tractable using a Semi-Markov process. A Markov process moves from one-time step to another and finds state of system in every time step. In Semi-Markov model the duration of time for which the system remains in a given state is also a random variable. So instead of moving from one-time step to

another, the Semi-Markov process moves from one state to another. The authors show that using a homogeneous model for reliability-based calculations overestimates the interruption costs by overestimating the fraction of longer outages. They claim that Semi-Markov method alleviates most of the shortcomings of homogenous Markov method.

(Boomsma, Kristoffersen, and Krogh 2008) developed a prototype model for estimation of reserve power needed to cover forced outages of power plants and deviations from forecast of wind and solar power generation. While modeling outages, they considered planned outages to be deterministic and forced outages as stochastic processes in a simple 2-state model. Instead of using a Markov process and generating scenarios using exponential distribution, the authors used a Weibull distribution for TTF and TTR values. They applied Semi-Markov process to obtain a sequence of TTR and TTF values for a whole year. The scenario generation algorithm had an exclusion rule to prevent overlap of generated forced outage with an existing planned outage. An index called Forced Outage Rate (FOR) was defined as the ratio of forced outage hours to sum of forced outage hours and available hours. A generated scenario was accepted when the FOR for the generated scenario was close to FOR calculated from historic data.

In his master's thesis (Bornak 2013) used similar methods used by (Boomsma, Kristoffersen, and Krogh 2008) for generating forced outage scenarios to calculate reserves. A Weibull distribution was used for TTR and TTF values. Semi-Markov methods were used with Monte Carlo Simulations to generate forced outages. However, in this case the author did not use any heuristic exclusion rule to prevent overlap of planned and forced outages. He applied this method to obtain outages for generating units in the South Africa power system which were then used to analyze

the implementation of reserves. (Barroso and Conejo 2006) also described a scenario generation algorithm comprising of successive sampling from TTF and TTR distributions till the end of time horizon to obtain outage scenario.

## **2.6 Suitable methods for hydropower units**

Based on past works, it is evident that modeling of forced outage should start with collection of required data on unit operations. Statistical analysis of outage data is important to understand the properties of outages that needs to be modelled. This analysis should answer questions about choice of distribution for TTR and TTF. It should also provide information about trend, seasonality and independence of outage events since application of Markov/Semi-Markov process requires events to be stationary and memoryless. Frequency and duration are appropriate reliability indices to compare historic outage data with generated outage scenarios. Monte Carlo Simulation is being used in the industry to generate scenarios of outages.

In this thesis, BC Hydro's generating units are considered base load units and the units undergo periodic planned maintenance. To the best of the author's knowledge, the impact of planned outages on forced outages for hydropower units have not been statistically investigated. At best, heuristic methods have been applied to prevent overlap of forced and planned outages in scenario generation (Bornak 2013). Appropriate statistical tests should be carried out to quantify impacts of planned outage on forced outage.

In the literature, either a simple two state model has been used for base load units or a 4-state model has been used for peaking units. Planned outages are scheduled well in advance and are

known to the modeler during scenario generation of outages. Two-state models move from one forced outage to another and are not designed to account for planned outages. It appears that past researchers had subsumed the planned outage state within “available” state or did not consider it altogether. Four-state models for peaking units are not suitable for modeling impact of planned outages because “reserve shut down” state is not equivalent to periodic “planned outage”. Moreover, BC Hydro’s optimization models aims to load the available units in such an order that maximizes revenue based on marginal cost of water and market prices for electricity. The forced outage scenarios that would be generated would be an input to these optimization models at a stage when loading on units is unknown. Hence, the states like “reserve shut down” and “Forced Out Not Needed” have little relevance in context of simulation and optimization studies. This problem necessitates modification of the two-state model to account for any impact of planned outages on forced outages.

In the next chapter, the methods used for statistical analysis of database are described. Based on the statistical analysis of data, an appropriate algorithm to generate scenario of forced outages is developed. Detailed description of Markov processes, Semi-Markov processes and Monte Carlo Simulation is provided. A two-state base case model for scenario generation is developed, which is extended to include impact of planned outages. The results of statistical analysis and scenario generation methods are presented in the case study Section.

## **Chapter 3: Methodology**

This chapter presents the method used to analyze the forced outage dataset and outlines the rationale behind the scenario generation algorithm developed and used in this research. The first Section describes available generating unit state data and its classification and reclassification to obtain two important outputs: Time to Fail (TTF) and Time to Repair (TTR) values. The second Section explains the methods used in statistical analysis of the TTR and TTF data. Finally, the third Section presents the scenario generation algorithm developed using Markov and Semi-Markov processes.

### **3.1 Sources of Unit Unavailability**

Statistical analysis of outage data and the development of probabilistic models for scenario generation requires the creation of a database. The standardized data recording system for various operating states of generating units was discussed in Section 2.2. A number of unit unavailability states can be identified, including:

- forced outages,
- maintenance outages,
- planned outages,
- upgrade outages,
- forced and scheduled de-rates, and
- forced extensions of maintenance and planned outages.

For all other operating states, the unit may or may not be generating power but is still available when needed. Since the analysis is only concerned with failure of generating units, the various

classifications of available states described in Section 2.2.1 is irrelevant. Hence, data can be collected only for the unavailable states. Quantification of uncertainty in unit availability is of paramount importance for use in system energy studies and in operations planning and optimization. The next sub-Sections outline the assumptions that were made to simplify generating unit's unavailability states.

### **3.1.1 Classification and reclassification of outages, forced extensions and de-rates**

Generating unit upgrade outages, planned outages and scheduled de-rates are known causes of unit's unavailability as described in Section 2.2.4. Information on planned outages is usually made available well in advance, as they require the allocation of manpower and other resources needed to complete the work. A major system upgrade implies that the unit will be scheduled to be out of service for an extended period of time, and the state of the units will be known to the energy studies modelers and operations planners as well. Therefore, planned outages, scheduled deratings and upgrade outages can be treated as deterministic inputs in operations planning and energy studies.

Maintenance outages are different from planned and upgrade outages as explained in Section 2.2.3. By definition, maintenance outages are those events where the unit has to be taken out of service to prevent a forced outage in the near future (NERC 2015). During these outages, the removal of a generating unit from service is generally done in order to perform work on specific components which could have been postponed past next weekend. The plant operator can schedule maintenance outage during weekdays following the next weekend, but it cannot be transferred to the next season. The problem identified, must be fixed in the current season implying that the plant operator has no information about maintenance outages that may occur in the next season or seasons

thereafter. This means that if a system optimization study is being carried out for a planning horizon of three to five years, then the system modelers would have no information about future maintenance outages. Therefore, for the purposes of energy studies, maintenance outages can be classified as uncertain as forced outage in terms of occurrence and duration.

Forced extension of maintenance outages was defined in Section 2.2.5. It is the period of time for which maintenance outage duration was extended as the repair work needed could not be completed in its allocated time. Now, if maintenance outages are to be treated as forced outages, there is no point of differentiating maintenance outage period and a forced extension of maintenance outage period. Therefore, it is logical to merge the data sets on forced extensions of maintenance outage with their corresponding maintenance outages.

Forced extensions of planned outages were defined in Section 2.2.5. Decision on including these outages for model studies should be based on preliminary data analysis on its impact on system. If the forced extensions of planned outages are not significant in number and duration in comparison to other sources of unit unavailability, then it can be eliminated from the database or merged with the corresponding planned outages.

Forced de-rates are operating states that are caused by a sudden malfunction of parts, similar to forced outages but do not require the unit to be shut off completely. During forced de-rates the unit availability can be anywhere between 0-98% of its Maximum Capacity Rating (MCR) as defined in Section 2.2.1. A common method used to simplify forced de-rates is to convert de-rates to an equivalent outage (Roy Billinton and Allan 1992; NERC 1992). For a derating event it be

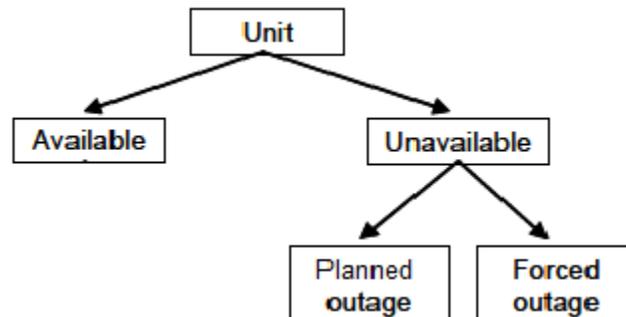
calculated as follows: Equivalent forced outage hours = Forced De-rate hours x (MCR – Final de-rated capacity)/MCR. For example, a 100MW unit was de-rated to 50MW for 10 hours during a de-rating event then the equivalent forced outage hour for that derating would be 5 hours. From an operations perspective, a 10-hour derating of 50MW is not equal to 5-hour outage of 100MW. However, for medium to long term energy planning studies, the conversion method can be used to simplify the problem if de-rates are not a major contributor to system unavailability.

However, if data analysis of different unit states shows that de-ratings are frequent and contribute significantly to unit's unavailability then, based on historic data, some de-rated states can be considered beyond the ON and OFF states. This would lead to a more realistic multi-state modeling of unit states and availability. For example, if a unit of 100MW capacity has incurred many deratings that reduced its capacity by half then a unit can be represented by 3 states: 0, 50 and 100MW. The discretization of unit states should be based on historic data and in consultation with the modeling group that would use forced outage scenarios in their models.

### **3.1.2 Defining Time to Repair and Time to Fail**

Based on the classification presented in the previous Section, all unavailable unit states can be grouped into two main categories: planned outages and forced outages. Planned outages, upgrade outages and scheduled deratings are deterministic inputs to modeling studies and are simply referred to as planned outages in this thesis. The forced extension of a maintenance outage is merged with its corresponding maintenance outage. Forced de-rates are converted to equivalent forced outages. All these are collectively called forced outages as they represent the source of

uncertainty for unit availability. Such simplified representation of unit availability has been used in (Boomsma, Kristoffersen, and Krogh 2008; Bornak 2013) and shown in Figure 10.



**Figure 10 Simplified chart for unit availability**

**Adapted from (Boomsma, Kristoffersen, and Krogh 2008)**

Stochastic representation of outages requires the statistical analysis of Time to Fail (TTF) and Time to Repair (TTR) values using the forced outage database. It should be noted that the use of TTF and TTR values fits well with the index of reliability, i.e., the frequency and duration of outages.

TTF values provide information about the frequency of outages and TTR values reflect the duration of outages. TTF should be computed from the moment the unit begins to operate to the moment it fails (Roy Billinton and Allan 1992). This definition of TTF tends to differentiate between the various Available states, as defined in Section 2.2.1, by allowing the inclusion of unit committed state and excluding the reserve shut down state, when the unit is not generating power. In the historic data set, there is information on unit's operating status and the reserve shut down period can therefore be eliminated from TTF values. But, looking forward, planners do not have any information about when the unit would be committed or in shut down state. So, the ideal way

of defining TTF from the historic dataset it is not useful. A more useful method to define TTF would be to move from one outage to another so that outage scenarios can be derived irrespective of all other deterministic operating states. Hence, TTF is computed from end of one forced outage to the beginning of next forced outage. TTR is the duration of the outage event, as shown in Figure 11.



**Figure 11 : Description of TTR and TTF**

A TTR and TTF database is created from historical records of generating units in order to perform statistical analysis as discussed in the next Section.

### **3.2 Statistical Analysis**

After obtaining the database on forced outages, statistical tests are used to understand the properties of forced outages that need to be accounted for in generating scenarios. Units and system performance levels can be examined in a variety of ways, most of which are appropriate for a specific purpose (Binder et al. 1991). If outage events are memoryless, independent and stationary then, homogenous Markov process can be used to model such events (Roy Billinton and Allan 1992; Perrica, Goldoni, and Raimondi 2009). The memoryless property assumes that the operating state of a unit in the next time step would be dependent only on the current state of the unit and not on the previous state. Independence of outage events assumes that the duration of TTR is independent of the duration of TTF. Stationarity assumes that the behavior of the system is not

changing over time, i.e., the conditional probability of failure or repair during any fixed interval of time is constant. In addition to these properties, the seasonality of outages and the impact of planned outages are important questions that can be evaluated using statistical tests. The methods used in this analysis are described in the following sub-Sections.

### **3.2.1 Trend in Occurrence and duration of outages**

The TTF and TTR depends on many factors, including basic system design, operating conditions, type of repairs, quality of repairs, materials used, etc. (David 1996). Generating units that take preventive maintenance measures have a bath-tub like hazard rate as discussed in Section 2.4.2. The first few operating years of any unit represent the period of infant mortality zone, therefore data for the initial years should not be included in the analysis as it would cause an unrealistically high rate of failure for a plant that is already in its useful life period. The useful life period of the unit is a period of constant rate of hazard, i.e., assumption of stationarity of the random process is valid in this zone. However, if there is an increasing trend in occurrence and duration of outage then it is indicative of an ageing unit with increasing wear and tear. Thus, it is essential to analyze trends in outage data to check if the unit is entering a wear out phase.

The evaluation of trends in TTF values was done by computing the number of outages per year and the mean annual TTF values. For analyzing trends in TTR values, the mean and median duration of outages per year were used. Mean TTF and mean TTR have been used in past works for trend analysis (Ellis and Gibson 1991; Koval and Chowdhury 1994). Median values have been used for additional verification because a single large outage event can skew the mean of outages.

To test the hypothesis that the time series data is stationary, Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test can be conducted on both the datasets of mean TTF and mean TTR. In the KPSS test, the time series under investigation is represented as the sum of a deterministic trend, a random walk, and a stationary error. For the time series to be stationary the variance of the random walk has to be equal to zero. The technical details about this test can be found in (Kwiatkowski et al., 1992).

### **3.2.2 Seasonality of Forced Outages**

The load demand and inflows in hydropower generation have seasonal variations. Since unit operation is highly dependent on these seasonal factors, it is crucial to check for seasonal patterns in outages. These factors have not been covered in much detail in past works. In (Simonoff et al., 2005) the authors look at the number of incidents and average generation capacity lost in each season for assessing seasonality of outage occurrences and durations. Analysis of seasonality is subjective depending on the question being asked. Outages can span across different seasons, so the generation capacity lost in each season does not provide information on the frequency and severity of outages in those seasons. For hydropower units, the real question to test seasonality is whether statistical distributions of TTR and TTF are significantly different from one season to another.

To study the impact of seasonality on TTR and TTF values, the dataset can be divided into four seasons. Winter season from Dec.-Feb., Spring from Mar.-May, Summer from Jun.-Aug. and Autumn from Sep.-Nov. Empirical Cumulative Distribution Functions (ECDFs) of TTR and TTF can be plotted for each of the seasons along with the ECDFs for all the months taken together.

Based on the plot, assumptions can be made regarding seasonality, with inputs from system modelers. If the individual ECDF plots of each season are outside the 95% confidence interval for ECDF of all months taken together, then this would indicate seasonality in data.

The reason for directly evaluating ECDFs instead of comparing point estimates like mean, median or variance is to prevent any bias and have a clear comparison of data values in every percentile of TTR and TTF. The median would only reflect a single point of the ECDF. The variance would explain the dispersion only. Mean value would be impacted by outlier values of extremely high TTR and TTF. However, direct analysis of ECDF overcomes these problems and shows how the comparison fares throughout the range of TTR and TTF.

### **3.2.3 Independence of TTR and TTF values**

Outage events are basically a series of transitions from TTF to TTR and back. Hence, the impact of TTR and TTF on each other has to be analyzed to correctly model failure events. Duration of an outage can depend on the preceding TTF value, and TTR can impact the subsequent TTF. If this phenomenon is demonstrated in the historic records of units, then it must be accounted for in the scenario generation algorithm. The Pearson correlation coefficient was used to study dependence between TTR and preceding TTF values, and between TTR and succeeding TTF values (Perrica, Goldoni, and Raimondi 2009). Scatter plots can be used to visualize these types of relationships. Autocorrelation in TTR and TTF values should also be checked by using the same Pearson correlation coefficient for different lags intervals.

### **3.2.4 Impact of planned outage on forced outages**

Planned outage events are basically preventive maintenance activities which are carried out to keep the system in a condition that is consistent with the required levels of performance and reliability. The objective is to keep the system failure rates from increasing above the design levels. Since the generating units are undergoing regular maintenance, it is worth looking at the impacts of planned outages on the occurrence and duration of forced outages.

In the previous analyses, planned outages have not been accounted for quantitatively. As shown in, Figure 11 earlier, a simple two state model has been used. In this analogy, a forced outage was tagged as an “unavailable” state and all other states, including planned outages, were incorporated within the “available” state. Use of this model is based on the assumption that planned outages and forced outages are completely independent of each other. However, planned outages can have an impact on forced outages as discussed earlier.

To quantify the impact of planned outages on forced outage, the database should be reclassified based on new definition of TTF. Until now, TTF has been defined as the time from one forced outage to the next forced outage, without consideration of planned outages. This has led to TTF values to be included in planned outage periods. Therefore, TTF was redefined as the time from the end of ‘last outage’ (planned/forced) to ‘beginning of the next forced outage’. This redefinition helps in classifying forced outages based on whether the preceding outage was a forced or planned outage. Two sets of TTR and TTF can then be developed, one for forced outages that occur just after planned outages (set 1: PF) and forced outages that occur after forced outages (set 2: FF), as

shown in Figure 12. These sets are thus called TTF\_PF, TTR\_PF and TTF\_FF and TTR\_FF for comparison purposes.



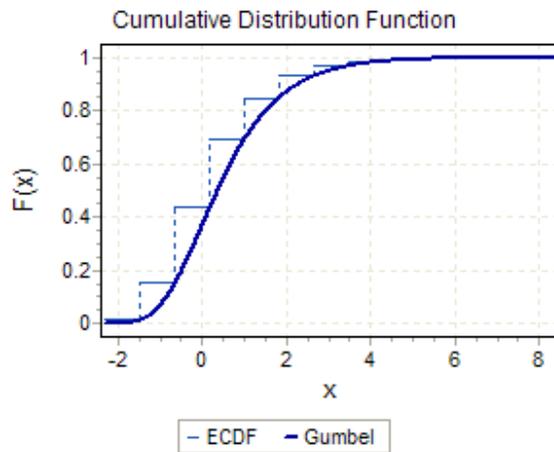
**Figure 12 TTR/TTF definition considering Planned Outages**

Comparison of ECDFs can then be done to see if the distributions of occurrence and duration of outages is impacted by a preceding planned outage. One plot can be created having ECDFs for TTR\_PF, TTR\_FF and TTR obtained using the older definition. Similarly, another ECDF plot of TTF\_PF, TTF\_FF and TTF using older definition can be created and analyzed. Based on these plots and in consultation with system modelers, suitable assumptions can be made regarding impacts of planned outages on forced outages for modeling and scenario generation.

### 3.2.5 Probabilistic model of outages

The statistical analysis of historic data is based on the premise that future values of TTR and TTF would be consistent with the past behavior of generation units. These past records should be used to obtain suitable probabilistic distribution for TTR and TTF. Probability distributions can be used to describe the outcomes of a random variable, which, in this case are the values of TTR and TTF. The probability distribution of a random variable is represented using Cumulative Distribution Function (CDF). ECDFs have been used in the statistical analysis Section for analyzing the properties of outages. However, parametric distributions are required for scenario generation

algorithms. In statistical analysis of seasonality, the purpose of using ECDFs was to compare one ECDF of TTR/TTF with another TTR/TTF. This comparison using ECDF is more robust than comparing point estimates. However, ECDFs is not well suited for use in mathematical models. Furthermore, ECDF is usually bounded at its lower and upper ends by the lowest and highest value in historic dataset but parametric CDFs are not. ECDFs are stepped functions but parametric CDFs are smooth as illustrated in Figure 13.



**Figure 13 Difference between ECDF and parametric CDF**

This warrants development of a parametric CDF representing historical TTF and TTR values that can be used to sample future TTR and TTF values.

In Section 2.4.5, it was explained that the exponential distribution is the most commonly used probabilistic model to fit TTF values in electric power systems (Barroso and Conejo 2006; Finger 1979; Scully et al. 1992). However, Weibull and Gamma distributions provide more flexibility as they use shaping parameters and thus have been used by many researchers (Boomsma, Kristoffersen, and Krogh 2008; Bornak 2013). The Lognormal function is also being considered an important tool to describe TTR values (Billinton & Allan 1992; Anderson & Davison 2005).

The advantage of Weibull, Gamma and Lognormal distribution functions is that they don't have any characteristic shape and can be shaped to represent many distributions that best fit the actual data set.

For forced outage events, the best-fit parametric distributions of TTR, TTF, TTF\_FF and TTF\_PF can be obtained from empirical data instead of assuming some distribution. A MATLAB code can then be used to find the best fit distribution for TTR and TTF values. The parameters of the distribution are obtained by Maximum Likelihood Estimation (MLE) (Raychaudhuri, 2008). (Sheppard, 2012) developed a code that can fit all the 18-parametric distributions defined in MATLAB to a time series dataset. The best fit among these distributions is obtained using Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

AIC is derived from information theory and is designed to pick the function that produces a probability distribution with the smallest discrepancy from the true distribution. It is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model. A lower AIC means a model is considered to be closer to the data. BIC is derived from a large sample asymptotic approximation to the full Bayesian model comparison. It is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup. A lower BIC means that a model is considered to be more likely to be the true model. More details about these criterion can be found in (Kuha 2004). For the purpose of finding the best distribution, it is sufficient to know that the distribution which minimizes AIC and/or BIC is the most suitable fit for that dataset.

In this Section, the statistical properties of forced outages were established and probabilistic model for TTR and TTF dataset were developed. Analysis of these methods will allow selection of suitable methods for the stochastic representation of forced outages as discussed in the following Section.

### **3.3 Methods of sampling from forced outage distributions**

Reliability based methods for stochastic modeling can be applied using an analytical, simulation or an optimization approach. The difference between analytical, simulation and optimization approaches is in the way in which reliability indices are evaluated. In an analytical approach, the system is represented by a simplified mathematical model and reliability indices are directly evaluated from solutions of equations. On the other hand, simulation and optimization approaches treat the problem as a series of real experiments conducted in simulated time to model the random behavior. Reliability indices are estimated from the results of simulation or optimization studies, by counting, for example, the number of outage events.

(Billinton and Li 1994) provides a good justification for using simulation methods rather than analytical solutions. In the past, analytical methods had the advantage of being less time consuming compared to simulation methods. However, due to modern computing resources, solution time is no longer an issue. Mathematical models can sometimes be an over-simplification of the system, thus becoming unrealistic. Simulation techniques can provide a wide range of output parameters including all statistical moments and probability density functions, whereas the output from analytical methods is usually limited to expected values. (Cepin 2011), on the other hand, cautions that the results in both approaches are only as good as the model of the system, the

appropriateness of the evaluation technique, and the quality of the assumptions and input data used in the models. (Endrenyi 1979) argued that analytical methods are more suitable for components with instantaneous repairs. However, for components with normal repair times and preventive maintenance, the simulation approach is preferred.

The usefulness of modelling a stochastic process rests heavily upon the ability to adequately model uncertainty. Simulation method aims to capture uncertainty through scenarios. In the following Sections, Monte Carlo Simulation methods are described followed by description of its application in forced outage scenario generation. Optimization methods are beyond the scope of this study and will be investigated in future studies.

### **3.3.1 Monte Carlo Simulations**

Monte Carlo Simulation (MCS) is a simulation method that uses repeated sampling of the random variable from a probability distribution to generate scenarios of the event as it might happen in real life thus, enabling computation of desired output. In simple terms, MCS is a methodical way of doing a *what-if* analysis in experimental situations where the results are not known in advance. It allows for the investigation of the complete range of risk associated with each uncertain input variable (Raychaudhuri 2008). Because of this property, MCS is being used in many applications of modeling studies ranging from natural sciences, social sciences, engineering, finance, etc. The application of MCS in any application involves four steps. First, a base case is established which is a deterministic model of the system that uses the most likely value of the input parameters. The next step involves identification of a suitable statistical distribution of the random variable. Next, a set of random variables is sampled from this distribution. One set of random numbers, consisting

of one value for each of the input variables, is generated to provide one set of output values. This represents one scenario. This process is repeated by generating more sets of random numbers, to obtain different sets of possible output values (i.e., scenarios). Finally, a statistical analysis is carried out on the generated output to obtain required statistics on the response of the system.

Failures in power systems are random events, hence they can be simulated using MCS. Sequential time MCS is applied when system behaviour depends on historic data (Alkuhayli, Raghavan, and Chowdhury 2012). In case of generating units, MCS can be used to randomly sample values of TTR and TTF from their respective probability distributions to generate scenarios of forced outages which can then be compared against historic data in terms of the frequency and duration of outages. In (Billinton and Li 1994), the authors described basic principles of three different simulation approaches. These approaches are: a state sampling approach, a state duration approach and a system state transition sampling approach. These methods are described in the following Sections.

### **3.3.1.1 State Sampling Approach**

A system is made up of many individual components. The combination of each component state decides the final state of the system. A power-plant can be thought of as a system consisting of individual units as components. In a state sampling approach, the behavior of each generating unit can be modeled to get plant level reliability indices.

On assuming individual generating units as components there are two states: available and forced out. Let  $PF$  denote the probability of failure of one unit. To obtain the unit state, a random binary

number is first generated, with values of either 0 or 1, from a uniform distribution and then the results are compared with  $PF$ . The planning horizon is discretized into smaller time-steps. If the generated number is less than  $PF$ , then the unit is assigned forced out state and if it is larger than  $PF$  then it is assigned available state for that time-step. The sampling is done for the next time-step till the end of planning horizon for all units. The main advantage of this approach is that the sampling method is relatively simple as it entails comparing a uniformly generated random number against the value of  $PF$ . It requires little input data as only component state probabilities are required. The main disadvantage of this method is that frequency and duration of outages have to be back calculated from the generated scenario of outage and available states.

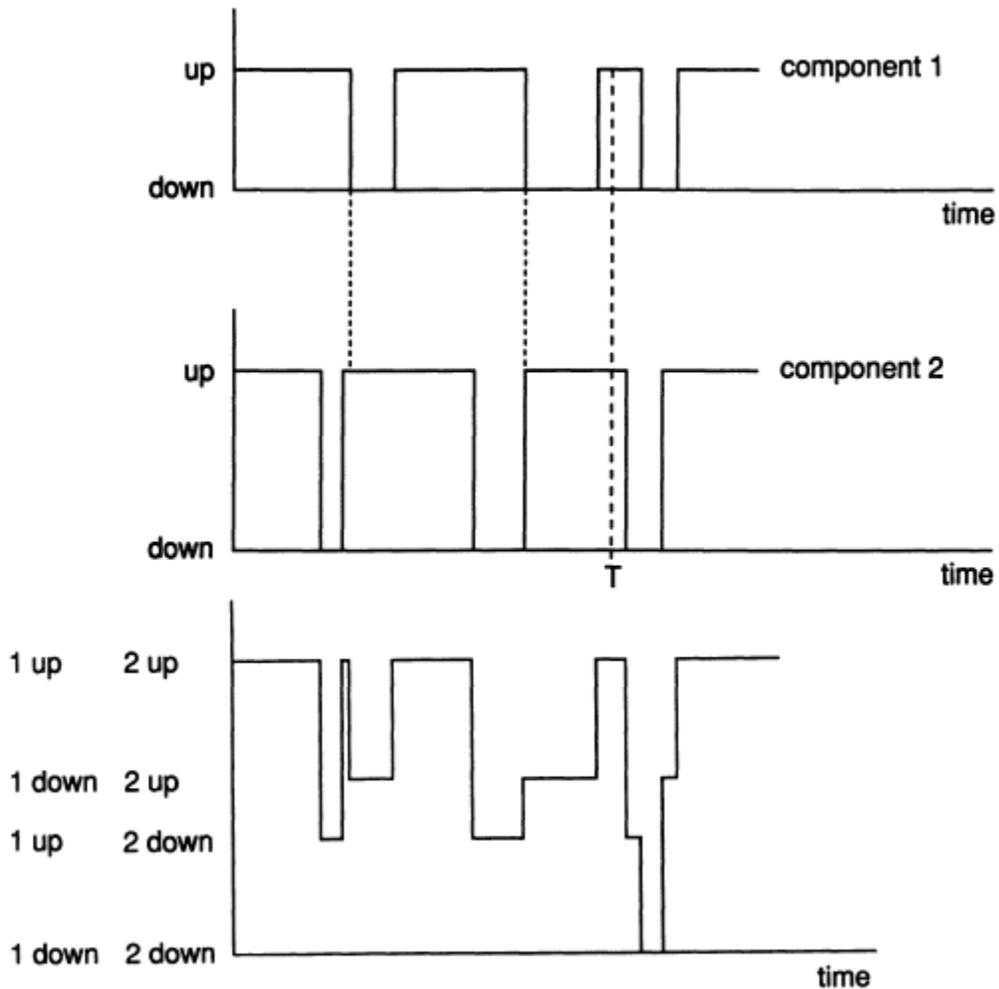
### **3.3.1.2 State Duration Sampling Approach**

The state duration sampling approach is based on sampling the probability distribution of component state durations. To apply this method, the unit state duration distribution functions are obtained. Here, duration distribution can be the probability distribution of any state duration like TTR or TTF. Next, unit states are sampled chronologically for all units. The chronological system state transition process is then created by combining individual unit state patterns generated. For the two-state representation of generating unit, component state duration distribution functions can be obtained for their available state (TTF) and repair state (TTR). The following steps can then be followed:

- 1) The initial state of each unit is assumed to be Available/UP state.
- 2) The duration of each component residing in its present state is sampled from its respective distribution. If unit is UP, a value of TTF is sampled. At the end of TTF period, the unit switches to the forced-out state and a value of TTR is sampled.

- 3) Step 2 is then repeated for the period of planning horizon for all units separately.
- 4) The individual unit state can then be used to obtain system state transition scenario.

This method is illustrated for a two-component system in Figure 14. First, individual unit states are developed, which are then combined to give system state.



**Figure 14 System State Duration Method (Billinton and Li 1994)**

The main advantage of this approach is that it allows any state duration distribution to be used for sampling. It also easily provides frequency and duration indices of outage patterns. Its disadvantages being computational time and memory required because of the large number of

individual component's states that must be covered by the sampling algorithm. This method also requires detailed inputs in the form of probability distributions of TTR and TTF for each component.

### **3.3.1.3 System State Transition Sampling Approach**

This method focuses on sampling state transitions of the system rather than sampling individual components in the system. For a power plant with  $n$  generating units, there can be  $n+1$  possible states that can be attained by the plant, i.e., 1 unit OFF, 2 units OFF, etc. If the individual component state durations are assumed to be represented by exponential probability distributions then the state duration probabilities of the system can be derived as follows (Billinton and Li 1994). To sample the system state durations, a random binary number can be generated (0, 1), and compared against state duration probabilities for the plant. A system state transition sequence can be obtained by sampling of each time step until the end of planning horizon. The main advantage of this method is that the reliability indices for the entire system are calculated directly. The major disadvantage of this approach is that the system state probability distributions can only be represented by exponential distribution if component states durations have exponential distribution.

In summary, there is a trade-off between the state sampling approach and state duration sampling approach in terms of preparation of database, computational expense and post processing of output results. The state sampling approach requires few inputs and has a simple sampling algorithm but requires computation of reliability indices like frequency and duration of outage events from generated scenarios. The state duration sampling algorithm requires detailed input distributions for

component TTR and TTF. Although it is computationally expensive, it can model any statistical distribution and easily provide the required reliability indices. The system state transition approach comes with very restrictive assumptions in that it can be used only if system states follow exponential distribution. Furthermore, it does not provide outage patterns for individual units. Both state sampling and state duration sampling approach are applied to generate scenarios of forced outages in the next Section.

### **3.4 Methods for scenario generation of forced outages**

Finally, after analysing the statistical properties, fitting a probability distribution and selecting suitable sampling approaches; appropriate modeling algorithms can be developed to generate scenarios of forced outages. The scenario generation methodology is described using a two-state model based on a Markov Process and a Semi-Markov Process. The Markov Process method uses state sampling approach to generate scenarios of outages and the Semi-Markov Process uses state duration sampling approach. The Semi-Markov method can be further modified to include impact of planned outages on scenario generation.

#### **3.4.1 Markov Process-State Sampling Method**

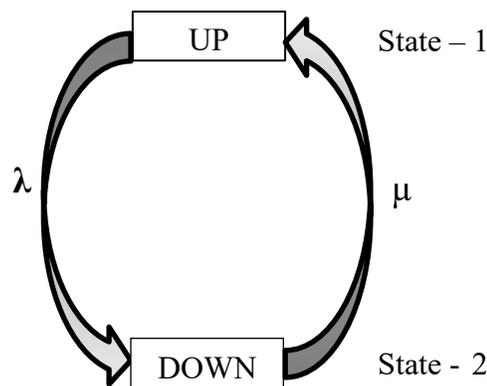
As has been briefly mentioned in the literature review, in Markov processes, the state of a system in the next time step is only dependent on the state of the system in the current time step and not on any past events. The dependence is quantified using transition probabilities for every possible change of state. If these transition probabilities are stationary over time, then the resulting Markov chains are called time-homogenous Markov chains (Konstantopoulos, 2009). The applicability of this method depends on the assumption that outage events are independent and identically

distributed. An outage scenario can be considered a series of ON-OFF states at every discrete time steps. A series of ON makes up TTF and a series of OFFs makes up TTR. Since this pattern involves the ‘transition’ from one state to another, a Markov Chain – Monte Carlo Simulation (MC-MCS) process is suitable for scenario generation. Initial research on scenario generation of outages has extensively used Markov processes (Li & Wang 2017; Moatti 1988; Finger 1979; Scully et al. 1992; Endrenyi, 1979 ; Barroso & Conejo, 2006).

Figure 15 shows a two-state representation of generating unit where the ‘Up’ state is when the unit is available and ‘down’ state is when unit is under forced outage state. Markov model for forced outage assumes that the switch from Up to Down and vice versa, is constant in time irrespective of duration of existing state (Scully et al. 1992).

**Figure 15 Two-State model for generating units**

The state changes are also shown in Figure 15. The duration of the transition from Up to Down is



called as ‘failure rate’ or  $\lambda$ , defined as mean number of transitions from up to down state per unit time in the Up state. Similarly, ‘repair rate’,  $\mu$  is defined as mean number of transitions per unit

time in Down state (Scully et al. 1992; Endreyeni, 1979). The following equations describe both these parameters.

$$\lambda = \frac{n_{12}}{T_1} = 1/(T_1/n_{12}) = 1/ (\text{Mean Up Time}) \quad \text{Equation 21}$$

$$\mu = \frac{n_{21}}{T_2} = 1/(T_2 / n_{21}) = 1/ (\text{Mean Down Time}) \quad \text{Equation 22}$$

Where,  $n_{21}$  is the number of transitions from State 2 (UP) and 1 (DOWN) and is numerically equal to  $n_{12}$ , number of transitions from State 1 (DOWN) to 2 (UP).

If  $F(1 \rightarrow 2)$  is the frequency of transition from Up to Down, and  $P_1$  is the steady state probability of system remaining in state Up, then

$$F(1 \rightarrow 2) = P_1 * \lambda \quad \text{Equation 23}$$

And similarly,

$$F(2 \rightarrow 1) = P_2 * \mu \quad \text{Equation 24}$$

In steady state conditions, the frequency of entering an outage state is same as exiting it and  $P_1 + P_2 = 1$ .

Therefore,

$$P_1 = \mu / (\mu + \lambda) \quad \text{Equation 25}$$

and

$$P_2 = \lambda / (\lambda + \mu) \quad \text{Equation 26}$$

$P_2$  is also called Forced Outage Rate (FOR) as it gives the steady state probability of the unit remaining in a forced out state (Rondla 2012; Scully et al. 1992).

The scenario generation methodology using Markov Chain – Monte Carlo Simulation involves 3 steps. First is the preprocessing of historical data and the computation of transition probabilities. Second, scenario generation using repeated sampling, and finally, post-processing of output unit states for deriving the time and duration of outages.

In pre-processing of data, the historical data is taken, and pre-processing is done to obtain transition probabilities. The TTR value for every outage event is rounded off to the nearest integer. This is done because the planning time horizon for future forced outage scenarios is to be discretized in hourly time-steps. This step might make all TTR less than 0.5 hours equal to 0 and modelers can either eliminate such values or round those values to 1 hour. However, the failure time (TTF) values cannot be rounded off directly. Outages with TTF values less than 0.5 hours cannot be eliminated or merged. This is because a small duration outage state may be quickly followed by a large duration outage due to various reasons. Merging different types of outages is not a realistic representation of all possible cases. Hence, all TTF values less than 1 hour are made equal to 1 hour and other TTF values are rounded off to their nearest integer values. This is followed by the computation of the failure rate, repair rate and FOR in the dataset.

Having quantified the FOR,  $\lambda$  and  $\mu$ , outage scenarios are created. The planning time horizon is discretized in units of 1 hour. The initial state of the system can be an input or assumed to be UP or can be randomized,  $x$ , between 0 and 1. To illustrate, if  $x \geq \text{FOR}$  (the  $P_2$  defined in Equation 26), then the unit is considered available at the start of the study; otherwise it is considered Forced Out at the start (time='t'). For the next time step (time='t+1'), another random variable is generated between 0 and 1. The following rule is used to assign the state of a unit:

- 1) If at  $t=1$ , unit is UP and  $x \geq \lambda$ ; then the state of the unit at  $t+1$  remains UP.
- 2) If at  $t=1$ , unit is UP and  $x < \lambda$ ; then the state of unit at  $t+1$  changes to Down.
- 3) If at  $t=1$ , unit is Down and  $x > \lambda$ ; then the state of unit at  $t+1$  changes to UP.
- 4) If at  $t=1$ , unit of Down and  $x \leq \lambda$ ; then the state of unit at  $t+1$  remains Down.

Once the state of the unit at  $t+1$  is obtained, the steps are repeated until the end of time horizon. This gives a series of UP and DOWN states. Multiple scenarios can be created for stochastic modeling purposes. However, the goal of this thesis is to test the accuracy of scenario generation methods, so to ensure statistical significance, scenarios for a period of 1000 year is generated. This is based on the assumption that a 1000-year period would remove the impact of sampling variability and help in assessing the adequacy of the algorithm used.

The scenarios generated are time series of unit states being in either UP and DOWN state in every hour. Post processing is also done on the generated data to obtain TTR and TTF values. The TTR and TTF values computed from outage scenarios are then compared against historic data. Additional statistics can be computed, such as the annual duration of outages, mean TTF, annual number of outages above certain value, etc. This scenario generation method can be further refined to model partial Derated states if those states are significant (Lisnianski et al. 2012).

### **3.4.2 Semi-Markov Process for FO-FO model**

Many authors have pointed to the inefficiency of homogenous Markov processes in modeling TTR and TTF, as it can give unrealistic results (Van Casteren, Bollen, and Schmieg 2000). A Semi-Markov process has been proposed as an alternative to homogenous Markov chain processes (Bornak 2013).

The distinguishing feature of the Semi-Markov model is the addition of a random variable that represents the duration in which the system remains in each state. In Markov processes, the planning horizon is discretized, and the algorithm moves from one time-step to another to evaluate the state of the system in every time step depending on the realization of the transition probability. In Semi-Markov processes, the unit remains in a particular state for the duration of time given by TTR/TTF. The values of TTR/TTF are derived from a pre-defined distribution, and the transition between states is modeled using those values. This feature allows the use of any distribution to represent the characteristics of the process. If the TTF and TTR of units are uncorrelated, then Semi-Markov process becomes suitable to generate outages for long time periods.

If  $X(t)$  is the state of the unit at time  $t$  and  $S_n$  represents the time of the  $n$ th transition, then the duration between the transitions can be calculated as  $U_n = S_n - S_{n-1}$ , which is also a random variable.  $U_n$  distribution is chosen to best fit historic data. Hence,  $U_n$  depends only on the states switched from  $X(S_n)$  and  $X(S_{n-1})$  and is independent of  $X(t)$  for  $t < S_{n-1}$ . To put it into the context of forced outage, there are two states: Available and Forced Out. After a transition from the available state to the forced-out state, the generating unit will remain in a forced-out state for the duration given by TTR value. Once it is in available state again, it will remain available for the duration given by TTF value (Boomsma, Kristoffersen, and Krogh 2008).

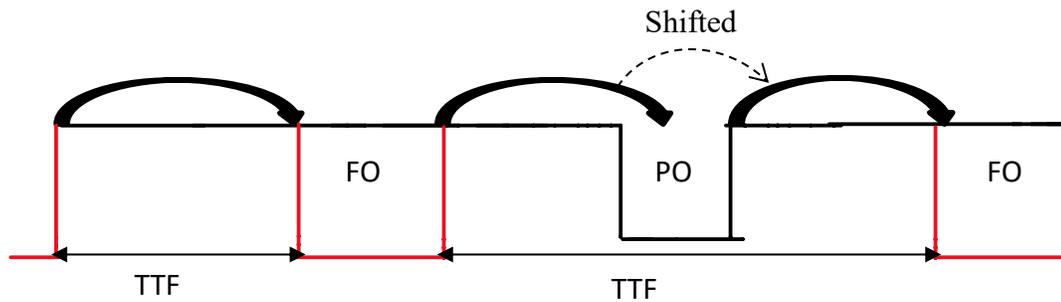
The application of Semi-Markov processes to generate outage scenarios does not involve any special pre- or post- processing of data. The algorithm, outlined below, is straightforward if the system is assumed to have the following states: Forced Out state and Available state. For the sake of convenience, this method is called FO-FO model.

- 1) Starting state of the unit can be assigned as UP or DOWN by user. However, as discussed above instead of multiple scenarios, algorithms would generate one scenario of 1000 years; so, there is no point in drawing the starting state randomly based on FOR values.
- 2) If the unit state at starting time step is UP, a TTF value is randomly sampled from the TTF\_FF distribution. The time pointer is moved to the end of TTF value.
- 3) If the unit state at the start is DOWN, a TTR value is randomly sampled from TTR\_FF distribution. The time pointer is moved to the end of TTR value.
- 4) Successively, TTR and TTFs are generated till the end of time horizon.

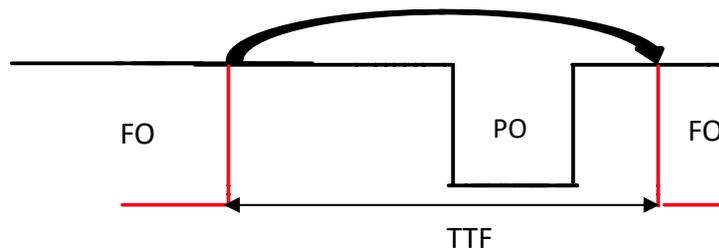
This method does not account for scheduled outages for the period under investigation. While it is likely that most of the planned and forced outages won't overlap, there must be rules to check and avoid overlaps between planned outages and forced outage patterns. The basic model was slightly modified using a heuristics rule to prevent overlaps. A similar approach was used by (Boomsma, Kristoffersen, and Krogh 2008). Basically, three things can happen with the generated TTR and TTF values. Either, the end of TTF falls within a planned outage period or a TTF period can completely cross over a planned outage period and/or a TTR period can overlap an existing planned outage period. The following heuristic rules were used in the scenario generation algorithm to avoid overlap between outages:

- a) When a value of TTF is sampled from the end of a forced outage, such that the end of TTF period enters a period of planned outage, then the time pointer is shifted to the end of planned outage period, as illustrated in Figure 16. Another value of TTF is then sampled from the end of planned outage period and the TTF period is defined as the duration between the end of last forced outage to the beginning of current forced outages.

**Figure 16 Preventing overlap of sampled FO with existing PO**



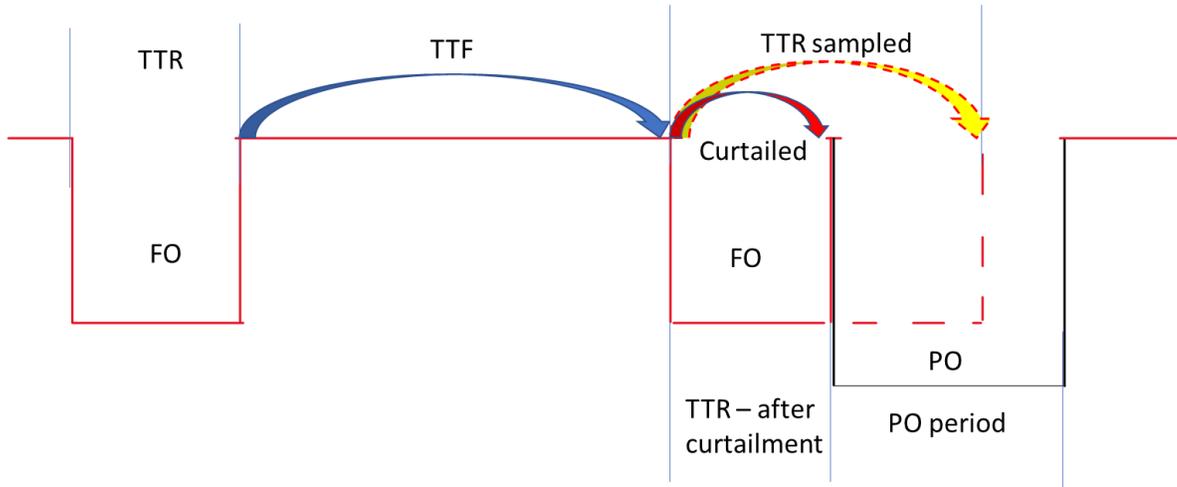
- b) If a TTF period crosses a period of planned outage, then TTF value is accepted without any changes, as illustrated in Figure 17. This is because the definition of TTF from historic data included periods of planned outage which are in between the two outages. Moreover, if a TTF is entirely crossing a planned outage period, there is no overlap between planned and forced outages.



**Figure 17 Case of TTF period crossing PO period**

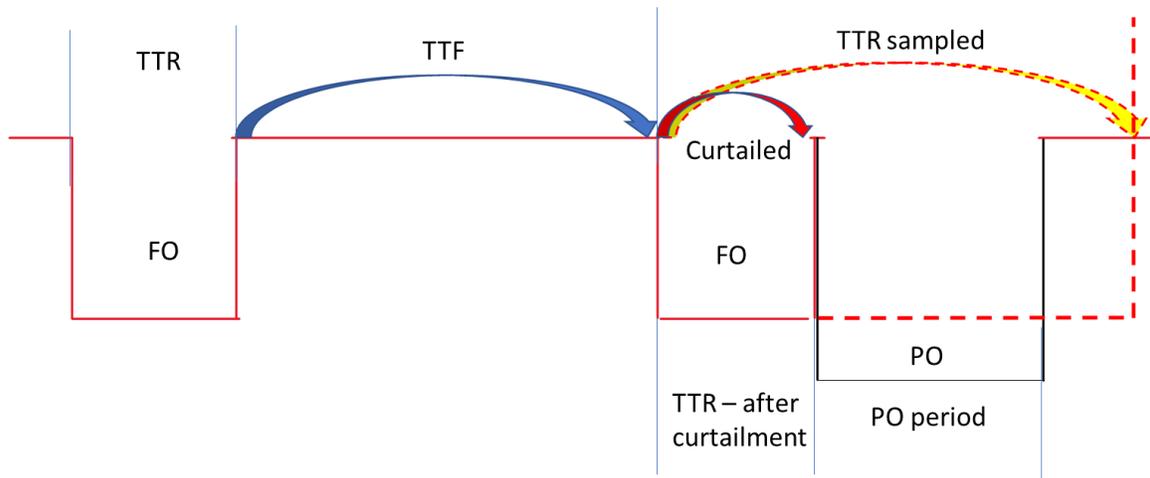
- c) If a drawn sample of TTR extends into the period of planned outage the forced outage period is curtailed to the beginning of the planned outage, as shown in Figure 18. To the energy planning modeler, this case would appear as a forced outage immediately followed

by a planned outage. During scenario generation, there is information about the cause of forced outage or the reason of planned outage, so they cannot be merged into a single outage. A new TTF is sampled after the end of planned outage period.



**Figure 18 TTR- curtailment to prevent overlap with PO**

d) If TTR completely covers the period of planned outages and extends beyond it then the procedure outlined in step (c) is followed. The TTR is curtailed and new TTF is sampled from end of planned outage as illustrated in Figure 19.



**Figure 19 TTR curtailment when FO exceeds PO period**

Some authors suggested another additional step in the scenario generation algorithm. They suggested that choosing scenarios where the FOR of generated pattern is within a tolerable limit of historic FOR (Boomsma, Kristoffersen, and Krogh 2008; Bornak 2013). In this work, this step is not included because the goal is to find a scenario generation algorithm that gives realistic distribution of TTR and TTF. Apart from that, annual hours lost and mean annual TTF are also used for evaluating scenarios.

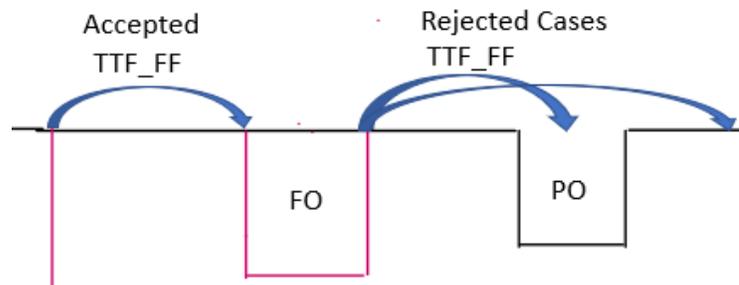
### 3.4.3 Semi Markov Process for PO-FO model

The heuristics used in the simple Semi-Markov method does not incorporate any impact of planned outages on TTF or TTR. In Section 3.2.4, it was argued that TTF for the next forced outage can be impacted by preceding planned outages. To quantify any impact, the database was reclassified to differentiate distributions of TTR and TTF for outages that occur after planned outages and those that occurred after forced outages. For this, TTR\_PF, TTR\_FF, TTF\_PF and TTF\_FF were defined, where PF means that a forced outage is preceded by a planned outage and FF means that a forced outage is preceded by forced outage. To make use of this differentiation, another scenario generation method is proposed to extend beyond simple heuristics and to comprehensively account for impact of planned outages on forced outages. Hence, this modeling method is referred to as PO-FO model to differentiate it from the FO-FO model as described in the preceding Section.

This method basically assumes a system-reset once a planned outage is encountered. In the FO-FO model, TTF was reset only when it ended within (*overlapped*) in a PO period (Figure 16). If the TTF value *crossed* over a PO period, it was accepted (Figure 17). The major difference in PO-FO model from FO-FO model is that in both the cases, when TTF period *crosses over* or *overlaps*

a planned outage period, the time pointer is set to the end of planned outage. Furthermore, this method uses TTF\_FF and TTF\_PF instead of TTF to distinguish between TTF distributions after forced outages and after planned outages. Similarly, TTR\_PF and TTR\_FF can be used to sample TTR values after a planned and forced outage. But for simplicity, it is assumed that both TTR distributions are similar, therefore a single TTR distribution is used in this case. The scenario generation steps are explained in the following steps.

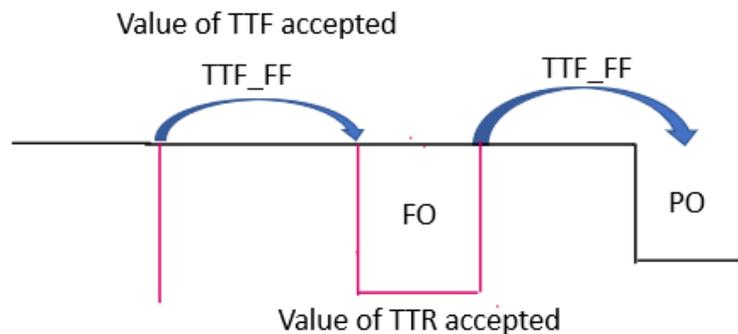
- 1) TTR, TTF\_FF and TTF\_PF distributions are obtained. Scenarios are to be generated for a period of 1000 year just for the purpose of testing the statistical significance of the algorithm used. Therefore, historic values of planned outages for the 20-year period are repeated sequentially to get data of planned outage for 1000 years.
- 2) The unit state is entered by the user for the starting time step. It is also assumed that the last outage was a forced outage.
- 3) At the starting time step, a value of TTF is drawn from the TTF\_FF distribution. A check is done to see if TTF overlaps/crosses a Planned Outage period as shown in Figure 20.



**Figure 20 Check for TTF value crossing a PO period**

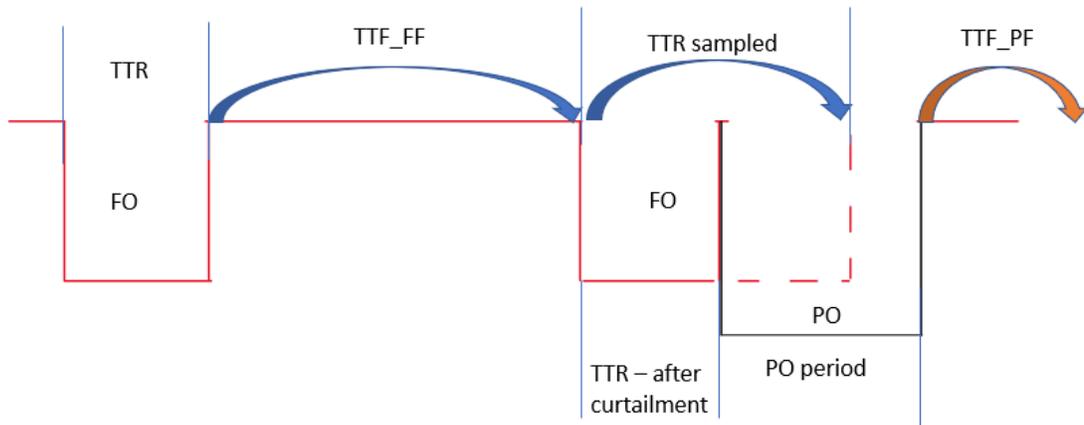
Two cases can occur. First, the TTF value does not cross or intersect any PO period or it might overlap/cross or intersect a planned outage period. Each case requires different set of succeeding steps as explained in the steps 4 and 5.

- 4) If the TTF value generated does not intersect or cross a planned outage period, then the following steps are carried out:
- The unit is assigned Available/ON state for the duration of TTF.
  - The time pointer is moved by the value equal to the duration of sampled TTF.
  - TTR value is sampled from TTR\_FF distribution.
  - The TTR value obtained is checked if it is intersecting with planned outage period or not. If TTR value does not overlap with planned outage then, the unit is assigned a Forced Out/OFF state for the duration of TTR. The time pointer is moved by a value equal to the duration of TTR. TTF\_FF is used to sample another TTF and the algorithm moves back to step 3. These steps are illustrated in Figure 21.



**Figure 21 Accepted value of TTF and TTR**

- However, if TTR ends up within a planned outage or crosses a planned outage period, the TTR is curtailed and reset to the beginning of the planned outage and the time pointer is moved to the end of planned outage. TTF\_PF is used to sample another TTF and the program moves back to step 3. This is illustrated in Figure 22.

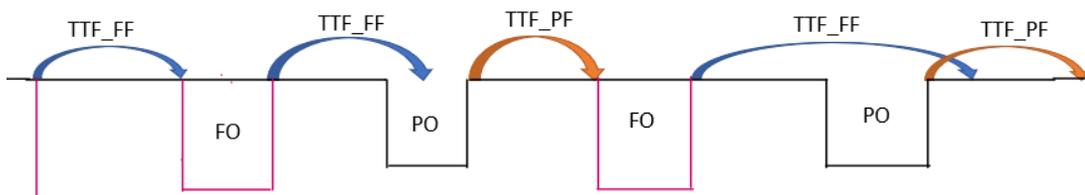


**Figure 22 TTR curtailment and sampling from TTF\_PF**

5) If a TTF value, obtained from TTF\_FF, lies within *or crosses* a planned outage period then the following steps are performed:

- a. Unit remains in the ON state only until the beginning of the planned outage. The time pointer is then moved to the end of planned outage period.
- b. TTF is sampled from TTF\_PF distribution and algorithm moves back to step 3.

This is illustrated in Figure 23.



**Figure 23 Rejecting sampled TTF\_FF and sampling from TTF\_PF**

6) The process is continued until the end of time horizon to obtain an outage pattern. From the output pattern, values of TTR, TTF\_FF and TTF\_PF are obtained for comparison with historic data.

### **3.5 Summary of Methods**

The first part of this chapter summarized the assumptions to create the database of TTR/TTF. The second part explained the statistical tests to analyze the seasonality trend and independence of outage events. Three different scenario generation algorithms were developed for generation of forced outage scenarios: Markov Process, Semi-Markov Process FO-FO model, and Semi-Markov Process PO-FO model.

The methods discussed in this chapter for statistical analysis and scenario generation are applied to a major power plant in the BC Hydro system. Taking the power plant as a case study, the results obtained after applying the methods developed here are presented in the next chapter.

## **Chapter 4: Case Study and Analysis of Results**

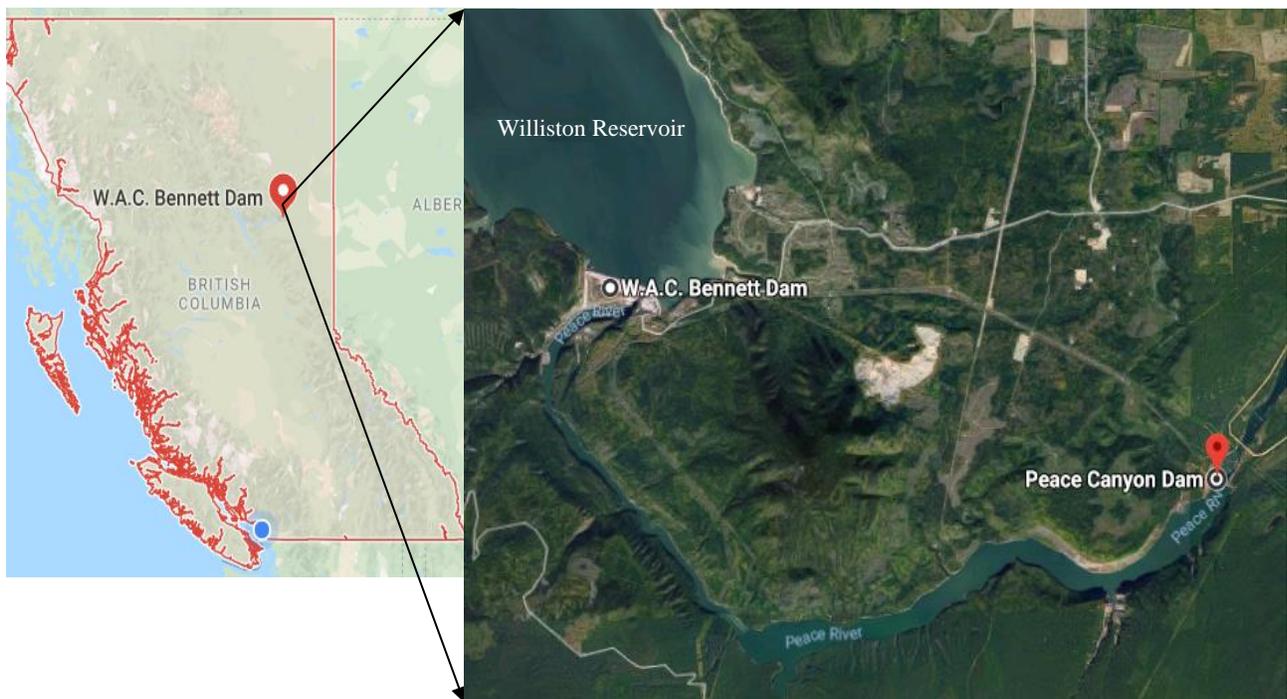
This chapter describes application of the methodology developed in the preceding chapters for a BC Hydro power plant. The first Section describes the power plant and develops database of TTR and TTF. In the next Section, statistical analysis of data is carried out and probability distribution of TTR and TTF is obtained. Finally, three scenario generation algorithms are used to generate scenarios of outages and the outputs are compared against historic data. In every step of this application, the modelers were consulted to verify assumptions and get their feedback. Parts of this chapter including the statistical analysis of data and Semi-Markov process were published in proceedings of Canadian Dam Association -2017 conference (Agrawal et al. 2017).

### **4.1 Gordon M. Shrum Generating Station**

Gordon M. Shrum (GMS) is the generating station on W.A.C. Bennet Dam. It is located on the Peace River, 23 kilometers upstream from the Peace Canyon Dam and about 160 kilometers upstream from the BC/Alberta border near Hudson's Hope. The dam project commenced in November 1961 and reservoir filling began in December 1967. It impounds water on Williston lake and releases it downstream in the Dinosaur reservoir for power generation through Peace Canyon Dam. The Peace River system is shown in Figure 24.

The reservoir gets inflow from winter runoff and snowmelt. The dam height is 183 meters with a reservoir operating range of 17 meters. The reservoir area spans 177,000 hectares. The GMS power plant has 10 generating units with a combined generating capacity of 2730 MW. It has Francis type turbines with annual average generation of 13,225 GWH. The facility began operating in 1968. Units 1 to 5 were installed in 1968 and 1969 and have a capacity of 261 MW each. Units 6

to 8 were installed in 1971 and upgraded in 2004 and have a capacity of 275 MW each. Units 9 and 10 were installed in 1974 and 1980 respectively and each has a capacity of 300 MW. Currently it serves 23-25% of the BC Hydro's total load. The powerhouse is 271 meters long, 47 meters high and 20 meters (wide. GMS is one of the oldest and largest of BC Hydro's generating facilities. At the time of completion, GM Shrum was the largest underground powerhouse in the world. It is the sixth-largest hydroelectric power plant in North America by capacity and the second-largest underground station by physical size.



**Figure 24 Peace River System in British Columbia**

**Source: Google Maps**

Since the plant is over 40-year old, the units are under periodic maintenance to keep the plant as a whole in its 'useful life' as described in Section 3.2.1. These maintenance procedures are extensive multi-year endeavors that include electrical upgrades, transformer replacements and unit

improvements with the goal to improve the plant efficiency in the most cost-effective manner. In the last two decades BC Hydro has undertaken many plant upgrades for GMS such as: Units 1-10 control system upgrade; Units 1-5 turbine replacement, rotor pole rehabilitation and exciter transformer replacement; Units 1-4 stator replacement; Units 6-8 capacity increase (from 275 MW to 305 MW each); Units 6-10 governor controllers; Units 7-8 exciter transformer replacement; and Units 9-10 excitation replacement. The common services projects include station service rehabilitation, transformer replacements (three phases), fire protection and alarm system replacement and an HVAC upgrade. (Armstrong et.al., 2016).

However, even with continuous maintenance, many forced outage instances occur in GMS. One of the most severe event occurred in March 2008 that resulted in a forced outage for duration of 14 months. GMS Unit 3 experienced a significant damage to the runner and water passage components, and operation of the unit ceased until repair work was completed. However, it is not just these extreme rare outage events that impacts energy studies. All forced outages of different duration occur throughout the planning horizon, impacting the production of the plant and the results of energy studies in varying degrees as described in 0. GMS, with its 10 units, has the largest capacity and energy storage in the BC Hydro system. Records of the operating status of GMS units were available for a long period of time. Hence, GMS was selected as a case study for modeling forced outage scenarios.

#### **4.2 Database of forced outages**

BC Hydro has maintained records on the unit status for each unit since 1977 in its Unit System Recording (USR) database, in line with the guidelines of CEA. Forced outage data for G.M.S was

obtained from USR administrator for analysis. GMS has 10 generating units. 8 units were already active since 1977 when the data recording was started. The last two units were added in 1980. Data was obtained for Planned Outages (PO), Forced Outages (FO), Maintenance Outages (MO), forced extensions of planned and maintenance outages, Forced De-rates (FD), and Scheduled De-rates (SD) for the period from January 1997 till December 2016. Outages that started in 2016 and extended beyond December 2016 were eliminated.

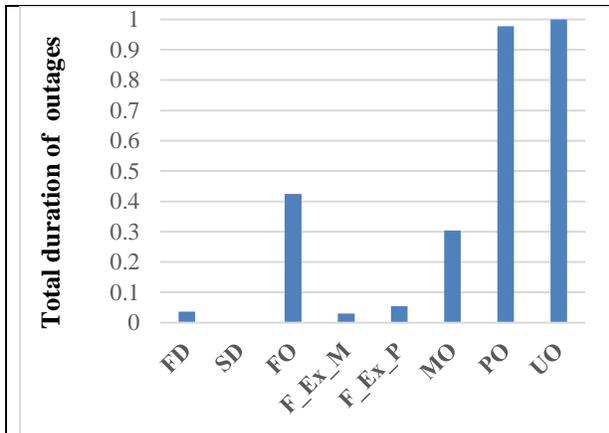
#### **4.2.1 Preliminary Analysis**

Initially, all the unavailable states of the units were individually analyzed in different ways. One approach of data analysis was to look at *impact of outages* by aggregating capacity lost due to forced outage events in daily/monthly time step. While such analysis indicated impact of outages on system but provided no help in understanding the frequency and duration of outage events. On discussion with BC Hydro modelers, it was decided to model outage *events* for units, rather than modeling *loss of generating capacity* for the entire plant. The problems with modeling at a system level were discussed in Section 3.3.1.3 on the system state transition sampling method. Another significant reason for not looking at capacity lost at the plant level was the fact that there is loss of information in aggregating data at daily or monthly time step. The actual event that occurs at a point of time is broken down into the impact the event caused in individual time steps. For example, an outage of 20 hours can occur in such a way that its impact is spread out equally into two days. Another instance can be case when two 10-hour outages occur on two consecutive days. Both these cases would result in the same impact in a daily time-step discretization of capacity lost and the information of frequency of outages/duration between outages will be lost. Correct system representation with such information would require modeling the possible transition from capacity

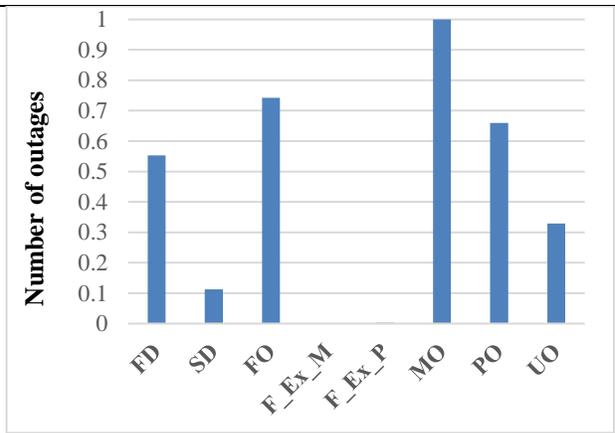
of one day to the other. Simply put, it would be far simpler and realistic to generate outage events that can be fed into models to represent unit availability.

Modeling outage events creates another problem because in the database, there are 8 different state codes that represent outages and derates. In Section 3.1.1, it was argued that planned outages, upgrade outages and scheduled derates are deterministic and hence can be combined as planned outages. Forced extensions of maintenance outages were merged with respective maintenance outages. All maintenance outages were classified as forced outages. For simplicity all forced derates were converted to equivalent forced outages as described in Section 3.1.1. Further analysis was also conducted on different categories of outages to verify those claims quantitatively and to gain some preliminary understanding of the outages occurring in GMS.

Figure 25 shows the normalized total duration of equivalent hours for which a generating unit remained out of service due to different types of outages. Equivalent hours refer to the duration of derating that has been converted to equivalent outage hours as described in Section 3.1.1. It shows the net impact that each category of outage had on the system in 20 years. The values are normalized to highlight relative impact. Figure 26 shows the number of events for each outage category over the same span of 20 years.



**Figure 25 Normalized Total Duration of Outages**



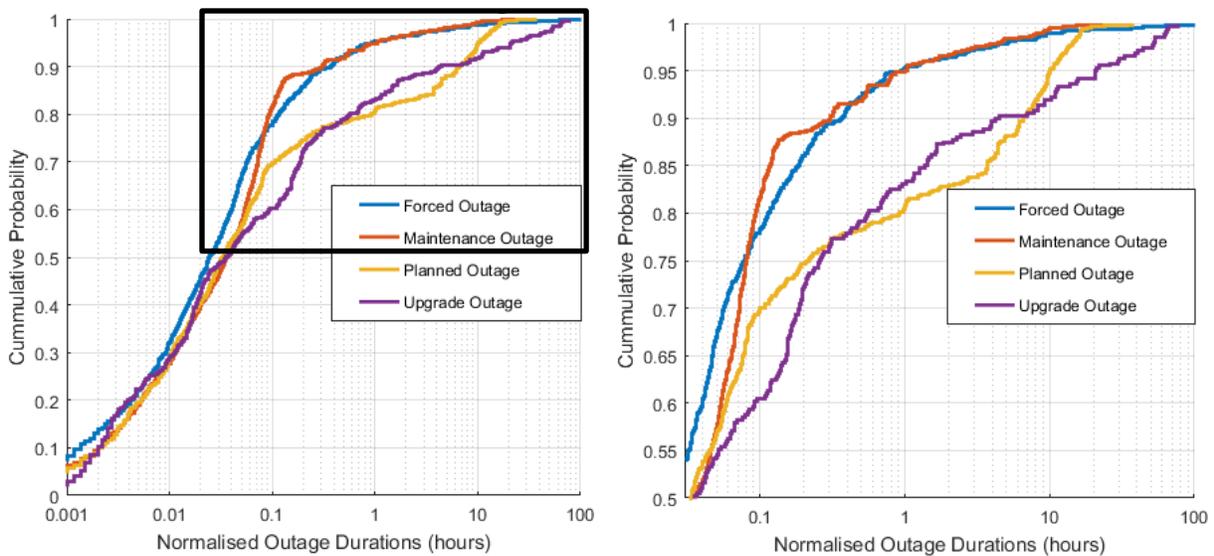
**Figure 26 Normalized Number of Outage Events**

**Legend:** **FD:** Forced Derates, **SD:** Scheduled Outage, **FO:** Forced Outages, **F\_Ex\_M:** Forced Extension of Maintenance outage, **F\_Ex\_P:** Forced Extension of Planned Outage, **MO:** Maintenance Outage, **PO:** Planned Outage, **UO:** Upgrade Outage.

It is evident from Figure 25 that the relevant outages are only FO, MO, PO, and UO. The derates and forced extensions have relatively negligible impact on the system performance. Furthermore, Figure 26 shows that there are many derating events (**FD**) and negligible forced extension (**F\_Ex\_M**) events. This indicates that, if forced derates are modeled, then it would lead to modeling many extremely small duration equivalent forced outage, and it could further complicate definition and modeling of Time to Fail (TTF) between outages and derates. Keeping both these things in mind, it was decided to drop forced-derates from modeling altogether. The same argument were made with regards to forced extensions and hence it was also dropped from further analysis.

Figure 27 shows the cumulative distribution of Time to Repair (TTR) values for different categories of outage events. The x-axis represents the normalized values of TTR which are obtained

by dividing the TTR values by a normalising factor. The values of TTR are shown on a log scale to visualize the entire range of outage durations. Y-axis represents cumulative probability in a linear scale and starts with 0.5. This is because smaller outages were found to have similar distributions for all categories and were not adding value to the discussion. Higher duration outages have greater impact on the management of the system and hence it is logical to look at difference in the curve in that region. It is obvious that the statistical distribution of maintenance outage is similar to that of forced outage and is distinctly different from that of planned outages. It is therefore safe to assume that the underlying processes responsible for forced outages are also responsible for maintenance outages. The MO and FO distributions are distinctly different from PO and UO distributions. This analysis further supports the argument presented in 3.1.1 that MO can be treated as FO and therefore it is logical to merge forced extension of maintenance outages with respective maintenance outages.



**Figure 27 Statistical distribution of TTR of different Outages**

In summary, the following steps were carried out to obtain database:

- 1) Forced de-rates, scheduled de-rates and forced extensions of planned outages were eliminated from dataset.
- 2) Forced extension of maintenance outages were merged with respective maintenance outages.
- 3) Maintenance outages and forced outages are together referred to as forced outages without any distinction.
- 4) Time to Fail (TTF) was defined as duration between end of one forced outage to the beginning of next forced outage. Time to Repair (TTR) was defined as duration of the outage interval, as discussed in Section 3.1.2.
- 5) TTR and TTF values were computed for each of the 10 units of GMS for outages in the 20-year period extending from 1997 to 2016.

The statistical properties of TTR and TTF are discussed in the following Sections.

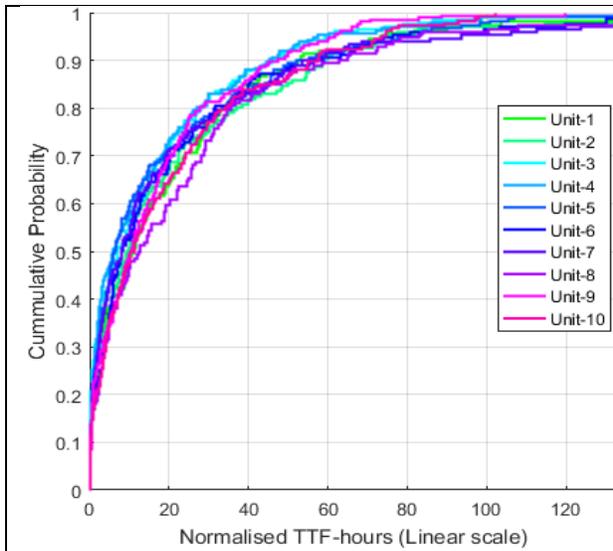
### **4.3 Statistical Analysis of data**

The goal of modeling forced outages in this work is to obtain a set of outage scenarios that realistically represents the historic data. TTF and TTR values for 20 years were computed for each unit without filtering out any recorded outage incident. Filtering data below or above certain outage duration introduces subjectivity in the analysis without adding value to the method of scenario generation. For example, a ten-hour outage may be insignificant in low load periods but can be crucial during peak load winter periods. Therefore, a more logical approach was to use all the data for scenario generation and then let the system planner decide on any filtering that needs to be done of the generated scenarios. In the following Sections, the data is statistically analyzed to find properties that have to be modeled and to develop representative probabilistic distribution models.

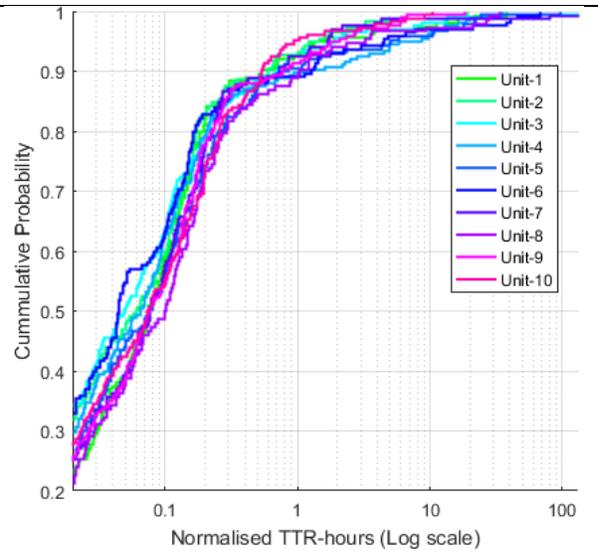
Note that the actual values of TTR and TTF cannot be disclosed as they contain commercially sensitive information. The graphs of TTR/TTF had to be normalised using the mean duration of all outage events in the historic data. This factor was applied for all datasets of TTR and TTF shown in all the graphs. Use of a constant value for normalising all datasets does not change the nature of the distributions and makes comparisons between them more informative.

#### **4.3.1 Homogeneity of Units**

Homogeneity refers to the similarity of outage related performance indicators of generating units. If all the 10 units in GMS behave similarly, then data for all units can be combined and a single distribution can then be used for all units. To check this assumption, Empirical Cumulative Distribution functions (ECDF) of individual units were plotted for both TTF and TTR as shown in Figure 28 and Figure 29. The two figures show the normalized values of TTF and TTR. It can be seen that the CDF of all units are very similar and crosses each other at several points, and appears to follow similar distribution.



**Figure 28 ECDF curves of TTF for each unit**



**Figure 29 ECDF curves of TTR for each unit**

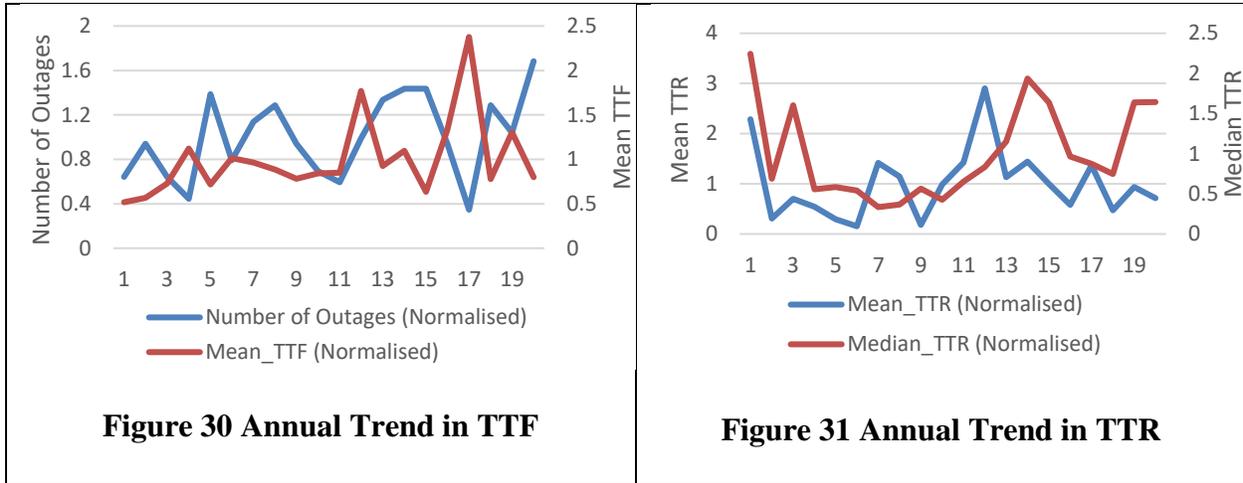
Statistical tests can be used to compare two sets of values to verify if they are derived from the same parent distribution. Kolmogorov-Sminrov two sample test is one such statistical procedure to test the hypothesis that the datasets are derived from the same distribution. This test was chosen because it does not assume any distribution for the sample data and compares the empirical cumulative frequency distribution of the two data sets (Karson 1968). Both TTR and TTF of each unit were compared with every other unit at the 95% confidence level. Results indicated that most unit outages are statistically similar. Given that such homogeneity of the TTF and TTR distribution exists, irrespective of how units are operated, it is reasonable to assume that the distributions are similar.

### **4.3.2 Trends in Forced Outage Data**

Having established that all units in the plant have similar outage distribution, it is important to check for trends in outage data. The data for the most recent twenty years was taken under the assumption that GMS plant would be in a stable region of a typical bath-tub curve, as discussed in Section 2.4.2. The interval between successive events and the duration of events must be independent and identically distributed, otherwise the trend in data would have to be accounted for in modeling studies (Lindsay 2001). For forced outages, a trend in outage pattern is indicative of an aging system with increasing wear and tear. Note that the smaller duration of forced outages recorded in the USR system are not because of component wear and tear. Smaller outages are caused by many factors that are not necessarily related to mechanical wear and tear. The cause codes and comments in the USR system indicated that smaller outages were caused by failure to switch from different states, such as start-ups, sync-condense failure, or incomplete start sequence failures, etc. For this reason, it was decided that smaller outages had to be screened out to check for trends in data and to have a clearer picture of the outage frequency. Therefore, outages above 9 hours were included in the analysis. The value of 9 hours was decided upon following some trail analysis with different limits, and in consultation with planning modelers.

Evaluation of trends in TTF values was done by computing the number of outages per year and the mean annual TTF values. As shown in Figure 30, the number of outages and mean TTF vary every year and do not exhibit any increasing or decreasing trends. The high peaks in TTF values are due to major planned upgrades in those years. If units are in planned outage state, then the units cannot go into a forced outage state, and this will result in an increase in time between failure events.

For analysis of TTR, the mean and median duration of outages per year were used. Mean duration per outage has been considered in past works (Koval and Chowdhury 1994). Here, median values have been used for additional verification because a single large outage event can skew the mean. Figure 31 shows that there is no statistically significant trend in the TTR data set.



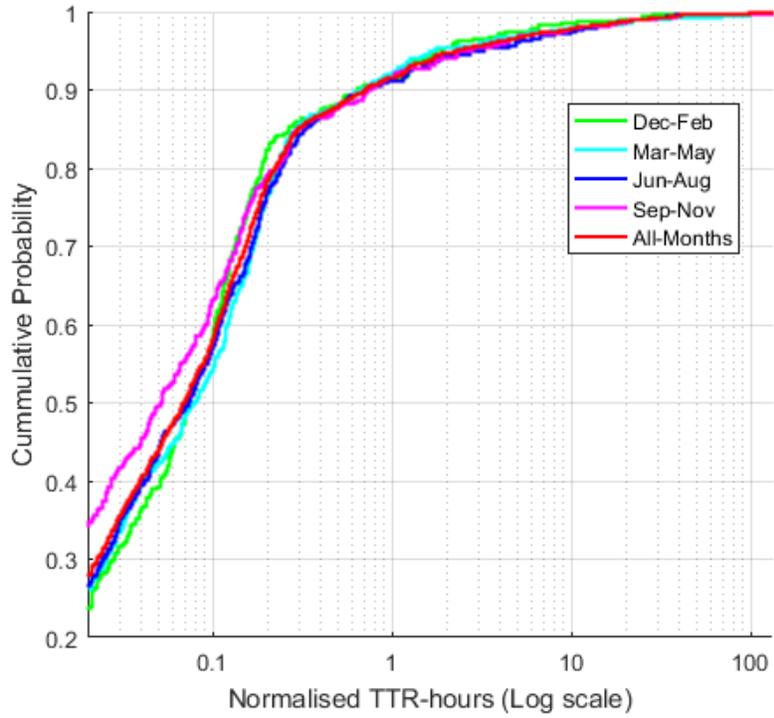
This analysis confirms the assumption that outage pattern is stationary for the plant as a whole. The annual number of outages and mean repair times may increase or decrease individually for each unit, but the plant witnessed a trendless outage frequency and duration due to periodic maintenance and upgrade of its components.

### 4.3.3 Seasonality of outages

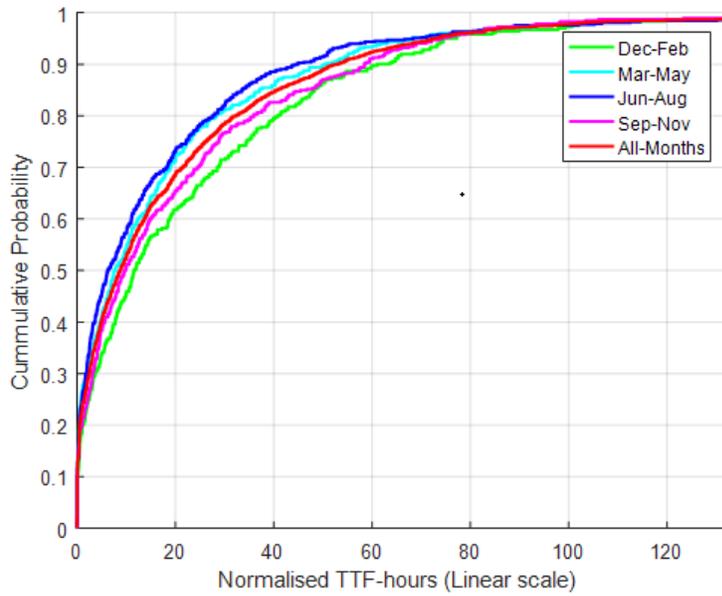
The load and inflows in hydropower generation have seasonal variations. Since unit operation is highly dependent on these seasonal factors, it is crucial to check for seasonal patterns in outages. This aspect has not been covered in much detail in past works. Simonoff et al. (2005) investigated the average generation capacity lost in each season for seasonality. Analysis of seasonality is

subjective, depending on the question being asked. In this work, the “effect” of outages in terms of generating capacity lost is not being investigated. Outages can span across different seasons, so generation capacity lost in each season does not provide information about the frequency and severity of seasonal outages. For hydropower units, the real question on seasonality is whether distributions of TTR and TTF are significantly different from one season to another i.e., whether the sampling algorithm should use the same distribution for the entire year or use different distributions for different seasons. The data set was divided into four seasons. Winter: Dec.-Feb., Spring: Mar.-May, Summer: Jun.-Aug. and Autumn: Sept.-Nov. ECDFs of TTR and TTF were plotted for each season along with the ECDFs for all months.

Figure 32 shows TTR distributions in different seasons. It can be seen that the curves for winter, spring and summer are almost overlapping. For autumn, there is divergence in the lower two quartile data, i.e., up to the median value of TTR. In the upper quartile, the autumn plot converges with other seasonal curves. It can also be noted that only one season is diverging and that is occurring for shorter durations of TTR. Therefore, it is reasonable to assume that TTR is not dependent on the season of occurrence and hence an annual random process can be used in this study.



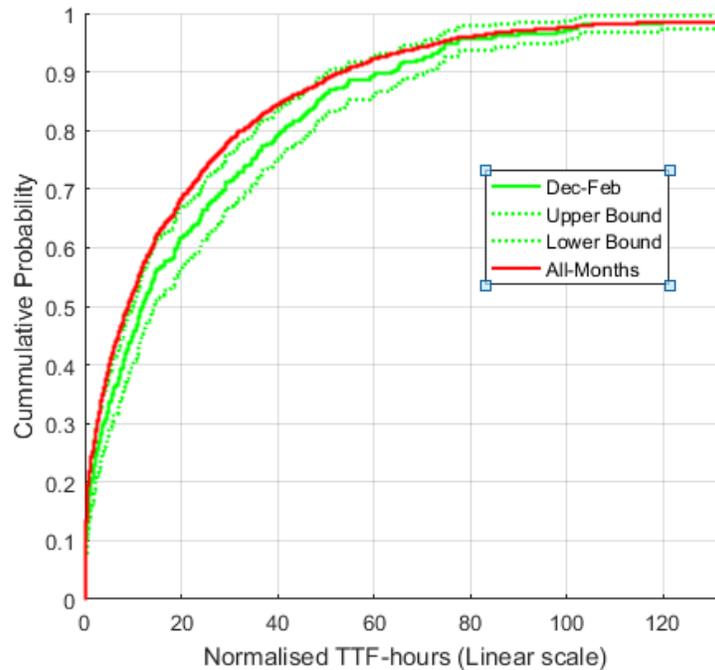
**Figure 32 Seasonality in TTR distribution**



**Figure 33 Seasonality in TTF distribution**

Figure 33, however, shows interesting features. The distribution indicates that TTF frequency is higher in autumn and winter as compared to spring and summer. This can be qualitatively explained by attributing this difference to start/stop of units. In winter months, demand is high and hence units are operated continuously with minimum start and stop events. In spring and summer, demand is low and there are many scheduled maintenance events of units. This leads to frequent start/stop events which are considered as one of the major reasons for forced outages by plant operators. Less number of Starting failures, tripping, and sync-condense operating states, etc. may be the other reasons of lower TTF values in spring and summer.

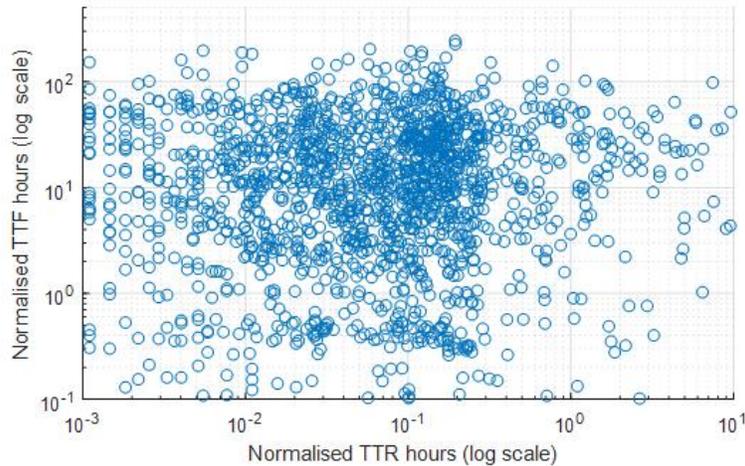
It is worth noting that the TTF plot for the entire dataset (red curve) is close to that of the summer, spring and autumn distributions. It can be also noted that the winter month curve has the highest deviation from the red curve. However, as can be seen in Figure 34, the TTF distribution for the entire dataset is close to the upper bound of the 95% confidence interval around the winter dataset. Therefore, if one distribution is assumed for all seasons, the final result would lead to a conservative estimate of outage frequency in winter months and the modeler would be erring on the side of caution. Considering the fact that the shape and general characteristics of the TTF distribution is the same and the TTR distributions are overlapping, it is reasonable to assume that the TTF distribution has negligible seasonality and one distribution for all seasons will results in a good approximation for modeling the TTF frequency.



**Figure 34 Confidence band on TTF-winter months**

#### 4.3.4 Independence of TTR and TTF

If the duration of outage depends on the preceding TTF or if the repair time impacts the successive TTF then the modeling algorithm must account for such dependence. Figure 35 shows the scatter plot of normalised TTR and their corresponding TTF values. TTR and their corresponding TTF were checked for linear dependence using the Pearson correlation coefficient. The linear correlation coefficient was found to be -0.0212 with a p\_value of 0.35, thereby clearly indicating that there is no linear correlation between TTR and their corresponding TTF values. Similarly, there was no observed correlation between TTR and successive TTF data. Based on this analysis, it can be concluded that TTR and TTF values are independent. Similar results have been obtained using similar method by (Perrica, Goldoni, and Raimondi 2009).



**Figure 35 Scatter Plot of TTR and TTF**

#### 4.3.5 Impact of Planned Outage

In the analysis conducted up till now, Planned Outages have not been accounted for quantitatively. A simple two state model, as shown in Figure 36, was used, where forced outages were classified under the “unavailable” state and all other states were assumed to be deterministically known. Recall that TTR duration covers the period from the end of a forced outage to beginning of next forced outage as discussed in Section 3.1.2. Therefore, all planned outages are incorporated within the TTF period.

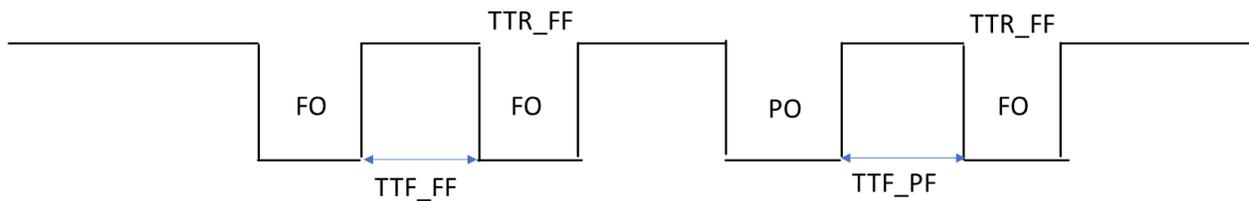


**Figure 36 Two State Unit representation**

Use of this model assumes that planned outages and forced outages are completely independent of each other. However, preventive maintenance should have an impact on forced outages. To quantify such an impact, the database was reclassified based on a new definition of TTF. A

preliminary study of outages at unit level showed that 17% of forced outages were preceded by Planned Outages and 83% were preceded by forced outages. So TTF was redefined from the end of the ‘last outage’ to the beginning of a “forced outage”, as discussed in Section 3.2.4 and shown in Figure 37.

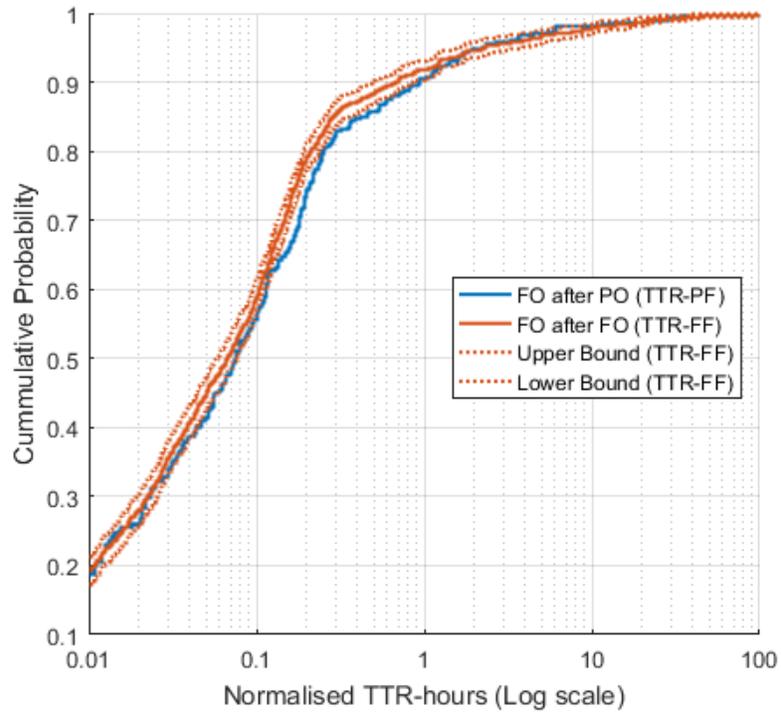
**Figure 37 Impact of Planned Outage on Forced Outage**



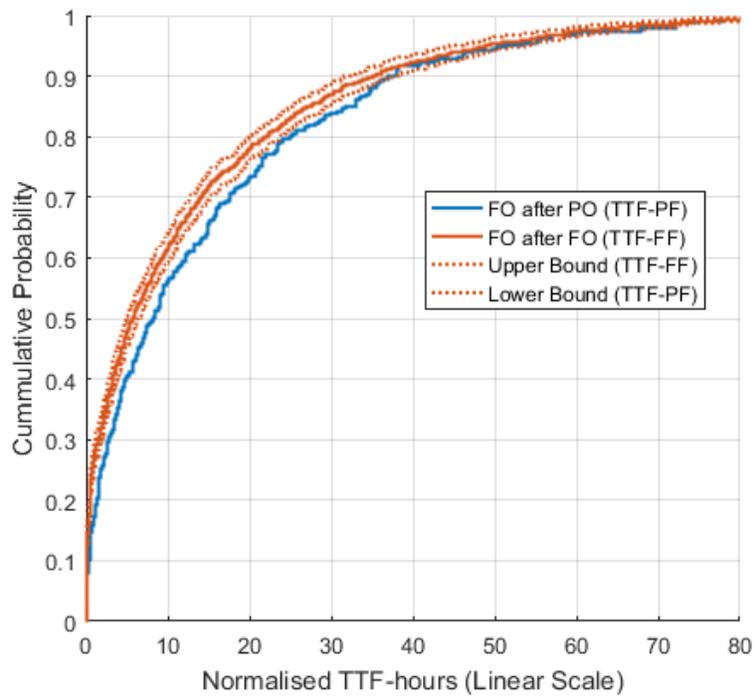
**Figure 37 Impact of Planned Outage on Forced Outage**

This redefinition leads to shorter TTF values. Two sets of TTR and TTF were developed, one for forced outages that occur just after a planned outage (set 1- PF) and the other for forced outages that occur after forced outages (set 2- FF), as shown in Figure 37. These sets were called TTF\_PF, TTR\_PF and TTF\_FF and TTR\_FF for comparison purposes.

Figure 38 shows that ECDFs of TTR\_PF and TTR\_FF cross each other and are overlapping, particularly for higher outages. In addition, it can be seen that most of the TTR\_PF distribution lies inside the 95% confidence band of TTR\_FF.



**Figure 38 Impact of Planned Outages on TTR**



**Figure 39 Impact of Planned Outages on TTF**

Figure 39 shows the ECDF plots for the two TTF distributions. It can be seen that the difference between the two curves is more systematic. TTF\_PF is higher than TTF\_FF for about 80% of the data range. TTF\_PF is also outside of the 95% confidence interval of TTF\_FF. Further it can be noted that the values of TTF in Figure 39 is smaller than Figure 34. This is because of the change in definition of TTF as described in Figure 37.

From the two figures above, it appears that the duration of outages is independent of the preceding planned outage while the time to next forced outage is statistically dependant on the nature of the preceding outage. In general, forced outage takes longer time to occur after a planned outage.

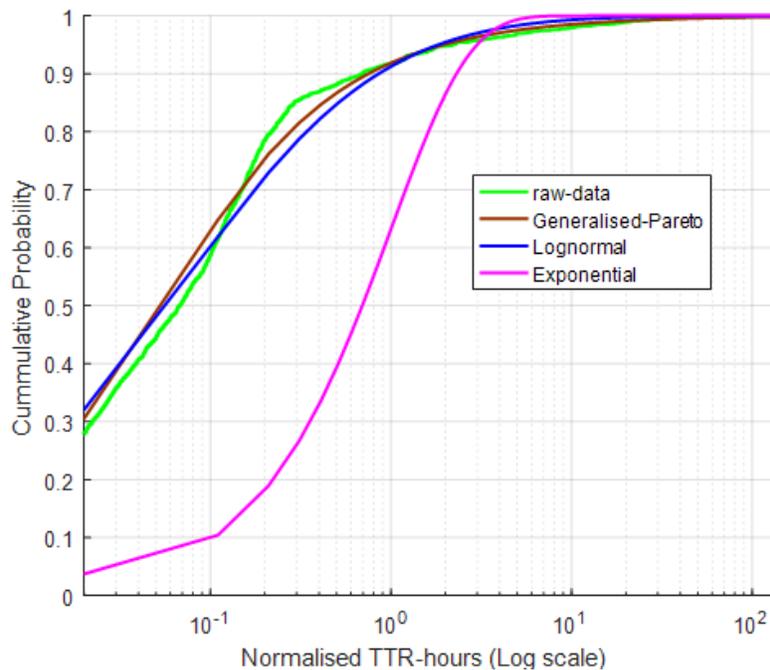
#### **4.3.6 Probabilistic Model for TTR and TTF**

The analysis in the preceding sub-Sections indicate that the probabilistic distribution of both TTF and TTR are independent, homogenous and without significant trends. TTR does not have any seasonal patterns and the assumption of no seasonality in TTF leads to conservative estimates of TTF during peak load periods. It was also observed that TTF for forced outages is longer after planned outages compared to that occurring after forced outages. The data does not indicate that there is any such significant difference between TTR values occurring after forced or planned outage.

Thus, the definition of TTF depends on how the probabilistic model is to be used in future studies. For simple two-state models, the definition of TTF from the end of one forced outage to the beginning of the next forced outage is suitable. But, for models where planned outages are treated separately, TTF can be segregated as TTF\_FF and TTF\_PF. The best-fit parametric distributions

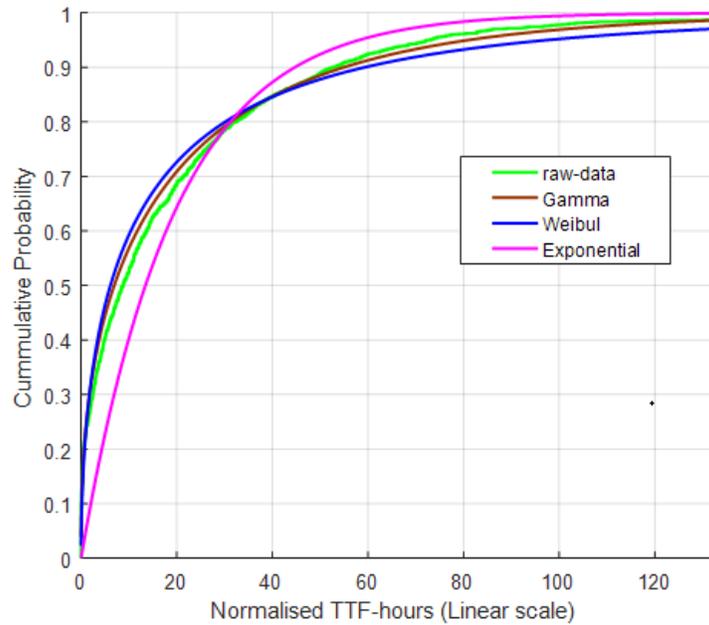
of TTR, TTF, TTF\_FF and TTF\_PF were obtained from historical data as discussed in Section 3.2.5.

Figure 40 shows the fitted probability distribution functions along with the ECDF of raw data of TTR. The green curve shows the ECDF of raw data. It can be seen that Generalised Pareto distribution exhibits the best fit of the dataset, while the Lognormal distribution is second best fit. The Exponential distribution exhibits significant deviations from the historic data and would be erroneous to assume and use an Exponential distribution for outage durations. The X-value of the graph are presented in log-scale to cover the entire range of TTR values. Also, some extremely low values of outage durations are not shown, but were included in the dataset to avoid loss of information in TTF values.



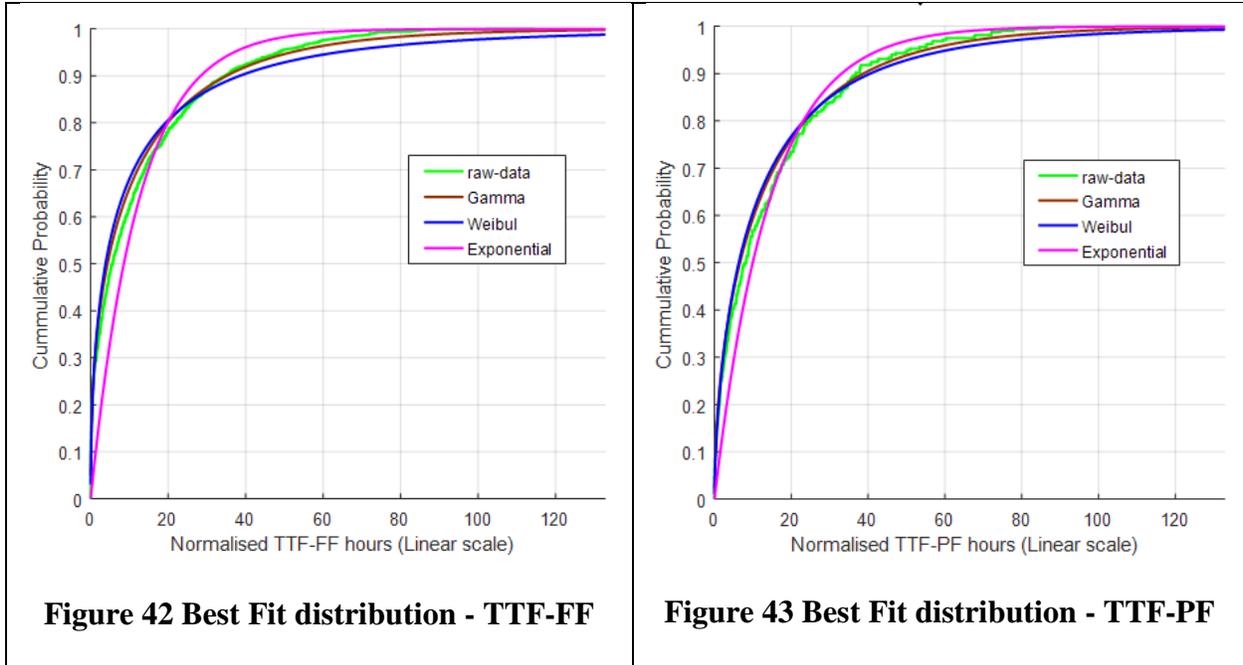
**Figure 40 Best Fit distribution – TTR**

Figure 41 shows fitted probability distribution functions along with the ECDF of raw data of TTF. It can be seen that the Gamma distribution exhibits the best fit of the raw data, with the Weibull distribution being a close fit as well. However, fitting an Exponential distribution exhibits significant deviation from the historic data and would be erroneous to assume and use Exponential distributions for TTF.



**Figure 41 Best Fit distribution - TTF**

Similarly, the Gamma distribution exhibited the best fit for TTF\_FF and TTF\_PF, as shown in Figure 42 and Figure 43.



#### 4.4 Results of Scenario Generation

Having quantitatively verified the statistical properties of forced outages for the GMS power plant, a Markov based scenario generation methods can be applied. As discussed earlier, one distribution can be used for all units instead of having 10 different distributions for each unit. Furthermore, it was found that there is no significant trend, seasonality or correlation in outage data. The impact of planned outage was also quantified, and suitable parametric distributions were selected for TTR and TTF. In this Section, three algorithms, described earlier, are applied to obtain forced outage scenarios which are then compared against historic data. The three methods being investigated are: Markov process, Semi-Markov process (FO-FO model) and modified Semi-Markov process that accounts for planned outages (PO-FO) model.

Given that the objective is to formulate a scenario generation algorithm and evaluate those based on their statistical properties, it was decided to generate one outage scenario for 1000 years instead

of generating multiple scenarios to test the statistical significance of the scenario generation methods. The generated empirical CDF of TTR and TTF were compared against their input distributions. The mean of Annual TTR and Mean of Annual TTF were also compared only after a good match was obtained for TTR and TTF distributions.

#### **4.4.1 Markov Process Model**

A MATLAB code was developed to generate outage patterns via Markov Chain Monte Carlo Simulation (MC-MCS). The raw data was rounded off as mentioned in the Section 3.4.1 to obtain hourly integer values of TTR and TTF. This approximation does not result in loss of information and makes comparison of input and output data easier because the smallest time step used for scenario generation is 1 hour.

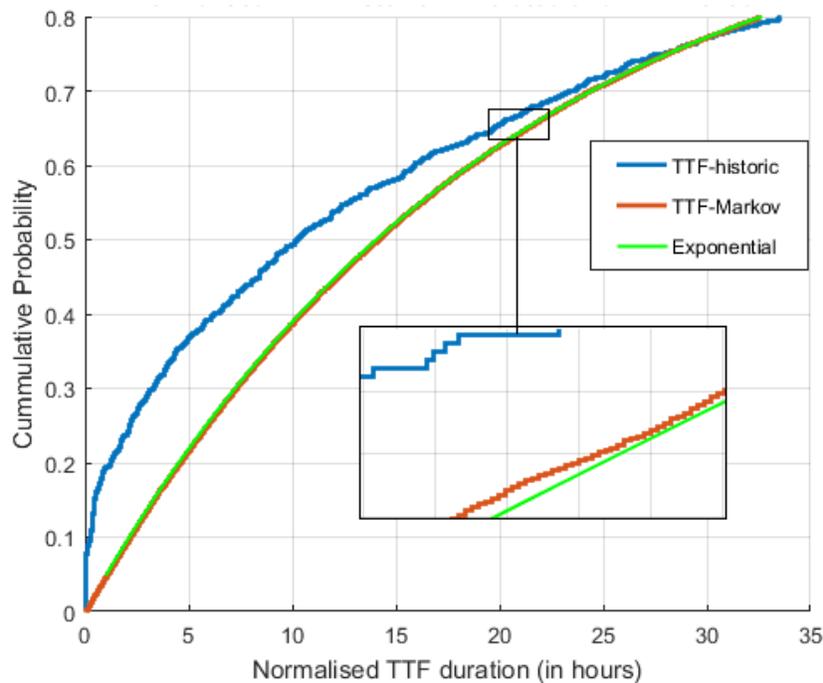
The MATLAB code has three functions. The first function imports raw data from excel. This collects the required state codes, outage time and duration for each unit. TTR and TTF values are rounded off as explained in the methodology chapter and were reloaded into the main code. In the main code, probabilities of unit remaining in UP/DOWN states were calculated from TTR and TTF values using Equation 25 and Equation 26. Then, for each unit, MC-MCS was carried out to obtain unit state for each hour for the entire time horizon. The series of unit states developed for each unit and for each hour was fed into another function. This last function calculates TTR and TTF values from the series of hourly unit states so that these can be used in the analysis of result. The entire process is not time consuming for even a 1000-year period. It took about 10-15 seconds to import data, generate scenarios and produce graphs of output on an 8GB RAM core i5 processor computer.

Empirical CDFs of TTR and TTF were used for comparison between historic values and generated data. As mentioned earlier, none of the outage events were eliminated in this analysis, as it would introduce subjectivity in selection of bounds and creates unnecessary complications in interpretation of TTF. However, for result analysis, it was decided to look at specific Sections of the TTR and TTF that are relevant for operational planning. Filtering at this point is simple and can be changed anytime if the modeler needs to check some specific features or change the bounds; without making significant changes in the scenario generation algorithm procedure.

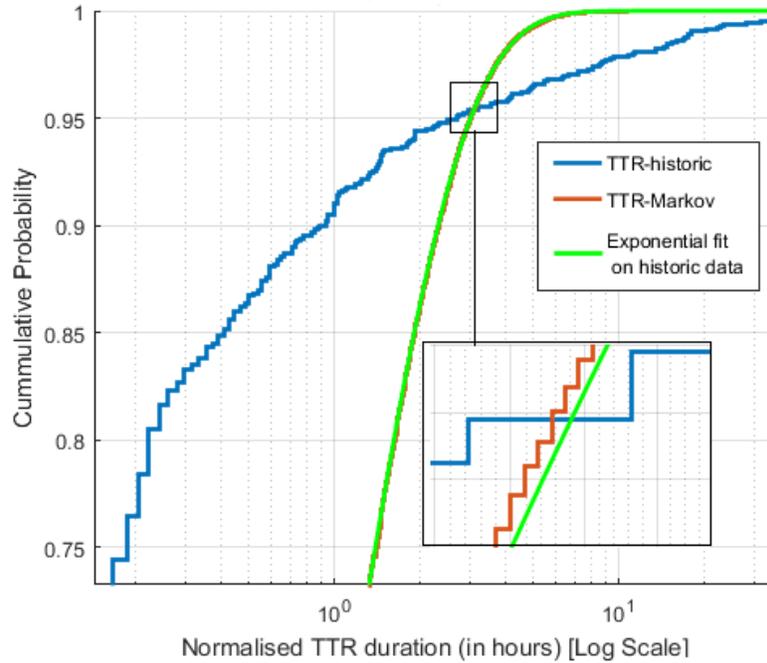
For the present case, the range of TTR from 10 hours to 2000 hours was used after consultation with system planning engineers. It was assumed that outages less than 10 hours are mostly manageable and need not be considered in operational planning models. On the other hand, outages spanning more than 2000 hours (~ 3 months) are catastrophic events arising due to severe damage/unavailability of resources and with no urgent need for the units. These events are very rare in the raw data but result in a flat tail in the cumulative distribution function due to the range of its values. Matching the tail portion is meaningless, so the upper limit of 2000 hours was considered adequate. For TTF, modelers were concerned about frequent outages rather than lengthy TTF durations. A scenario generation algorithm was deemed appropriate for operations purposes, if the distribution of output of TTF is close to the historic input data in the range from 0 to 80<sup>th</sup> percentile. In consultation with energy planning modelers, the bottom 80% of the TTF distribution was deemed most useful for operational planning purposes. It should be noted that that values outside this range of TTR/TTF were not eliminated. This range was just chosen for graphical comparison between

historic data and generated values by simply setting and displaying specific ranges of interest to specifically focus on such ranges.

The normalized results for MC-MCS method is shown in Figure 44 and Figure 45. The blue curve represents the ECDF of historic data. The brown curve is the ECDF of generated data and green curve is the parametric distribution obtained by fitting an exponential distribution on the historic data. It can be seen that the brown curve and green curve are completely overlapping. It is evident that TTR and TTF values obtained by MC-MCS method results in exponential distribution for those quantities. This has also been mathematically shown that an assumption of constant failure/repair rate gives exponentially distributed values for TTR and TTF (Endrenyi 1979). However, in Section 4.3.6, it was shown that exponential distribution neither represents TTR nor TTF for the GMS plant.



**Figure 44 TTF- Markov Process**



**Figure 45 TTR- Markov Process**

The cause of the discrepancy in the fitted distribution lies in the assumption and use of constant transition probabilities in the Markov chain models. It was demonstrated earlier that the failure and repair rates does not follow any rising or falling trends in the years for which the data is taken, however, it is not perfectly constant for each unit individually. There are variabilities within a fiscal year due to many factors like planned maintenance, load, start/stop, availability of man and material, transmission line outages, etc. All these factors make it difficult to model the pattern of outages using failure rates alone, and it can therefore be concluded that Markov process is unsuitable to generate scenarios of outages.

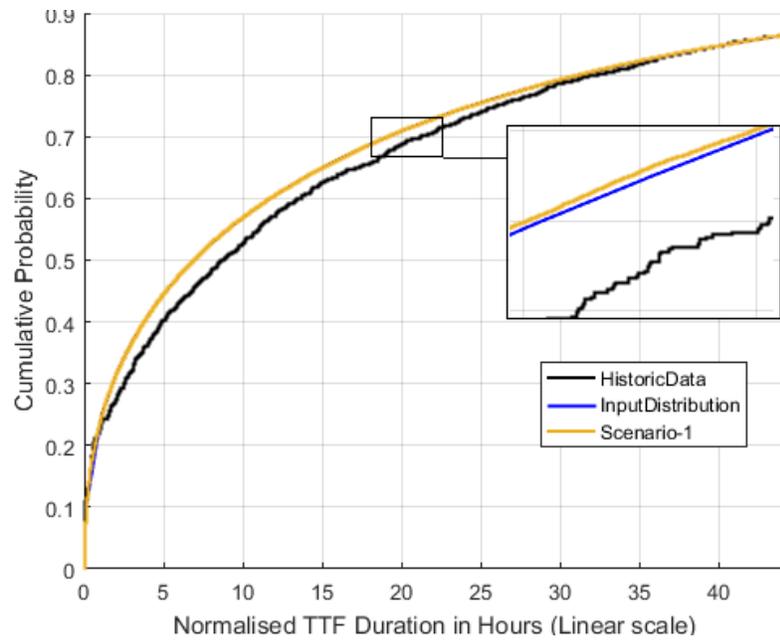
Due to the limitations of Markov Processes in accurately representing TTR/TTF, no further analysis was conducted on the results and the idea to use Markov Process and the system state sampling method was dropped.

#### 4.4.2 Semi-Markov Process (FO-FO) model

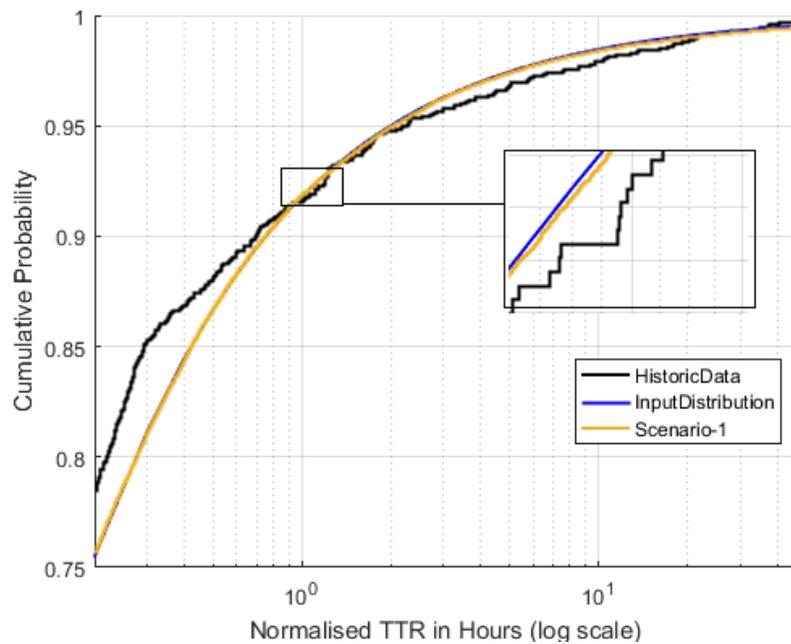
It is difficult to model distributions other than exponential distributions using homogenous Markov processes. Hence Semi-Markov process was applied to generate scenarios. The database of TTR and TTF obtained during statistical analysis were put to use for Semi-Markov process. The methodology was described in Section 3.4.2. The last 20 years of data on TTR and TTF were used without filtering out any values and without any rounding off as discussed in the Markov process model. A 2-state model is assumed; hence this model is named FO-FO model. The most suitable parametric distributions were investigated using the MATLAB code. For the current case, the TTF and TTR distribution obtained in Section 3.2.5 are used.

The model development was done in three stages. In the first stage parametric distributions were used as input. In the second stage, ECDF was used to check if it provided better results. Finally, in the third stage, a heuristic approach was used to prevent overlap of planned outages with forced outages.

**Stage 1:** A very simple model was developed assuming that sampling starts from the end of a forced outage. Repeated sampling of TTF followed by TTR, using the respective parametric distribution, is continued until the end of time horizon. The generated TTF and TTR values were compared against the ECDF of historic data to assess the accuracy of the model developed. The results are shown in Figure 46 and Figure 47.



**Figure 46 TTF FO-FO Model – (Parametric Input Distribution)**



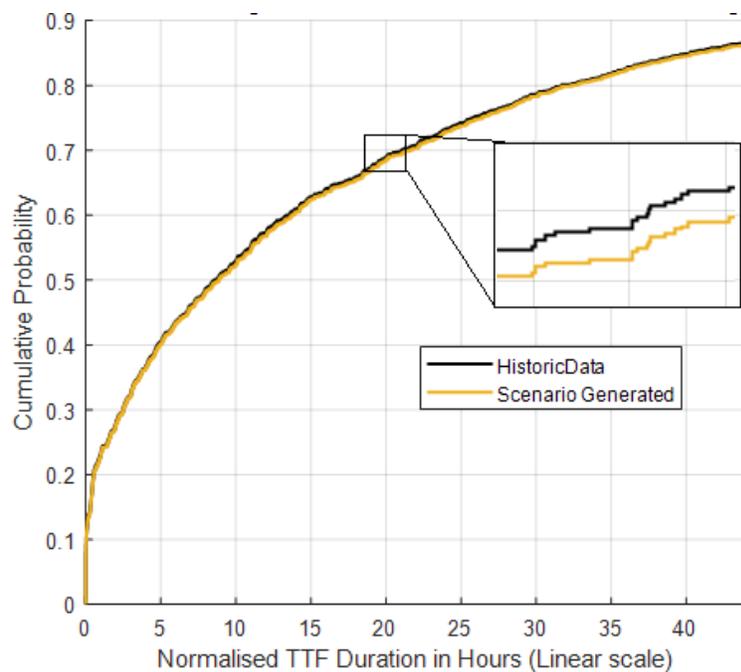
**Figure 47 TTR FO-FO Model – (Parametric Input Distribution)**

The black curve represents the empirical distribution for historic data, the blue curve represents the parametric distribution fit on historic values and the yellow curve represents the ECDF of generated values. For both TTR and TTF, the input and output distributions are exactly coinciding thereby indicating that the algorithm is working properly. After obtaining a good fit, other benchmarking criterion were analysed as follows.

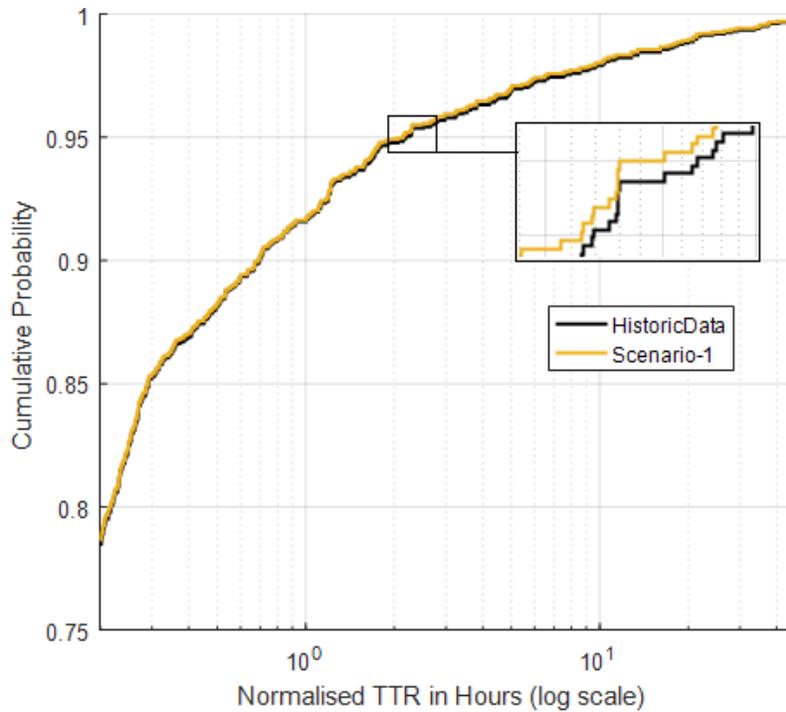
The annual number of major outages and the total annual outage hours (i.e, annual TTR) due to major outages were computed for each of the 20 years from historic data. Major outages are defined as outages with duration equal to or more than 10 hours. The median of these twenty-annual number of outages values was considered as representative of annual number of outages i.e., frequency. The median of the total annual TTR was considered as representative of annual duration of outages on system. The two median values were used as additional benchmarking of generated output scenarios. The same computation was done on generated data as well to get the median of annual TTR and the median of the mean annual TTF. One reason for generating one 1000-year long scenario was the fact that a sufficiently long-time horizon will take care of sampling variability in the generated data set, making it easier to compare median values of historic and generated data.

For the model developed, the median annual duration of outages in generated data was less than historic value by about 9% and the median value of annual number of major outages was more than the historic data by about 13%. The results indicate overprediction of frequency and underprediction of duration of outages.

**Stage 2:** This difference is within tolerable limits and can be attributed to the difference between historic data and input distribution, but the gap on the benchmarking indices can be expected to increase as the heuristics for planned outages are included. The scenario generation methods using non-parametric distribution were tested to eliminate any impact of curve fitting on the final result. To this end, the initial code was modified to sample from empirical CDF instead of parametric distributions. Sampling from ECDF was done by modifying a part of the MATLAB code. The results of using ECDF instead of parametric distribution are shown in Figure 48 and Figure 49.



**Figure 48 TTF FO-FO Model (Non-Parametric Distribution)**



**Figure 49 TTR FO-FO Model (Non-Parametric Distribution)**

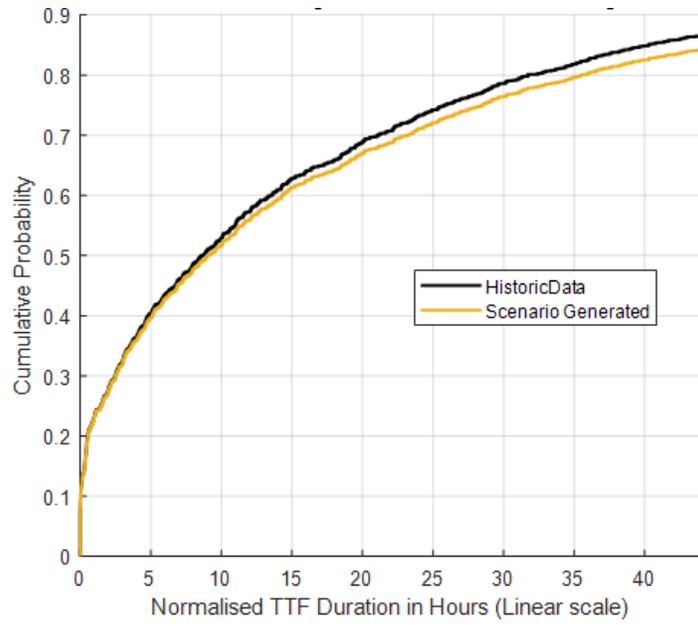
As can be seen in the graphs for both TTF and TTR, use of empirical CDF provides better results. There is complete overlap of input data and output distributions. The gap in the median of annual TTR between historic and annual data reduced from 9% in earlier case to 5%, using the empirical CDFs. The difference in the median annual number of outages in historic and generated data was practically negligible. After some tests, it was found that this improvement in results was obtained by using ECDF of TTR only. Use of ECDF instead of parametric CDF for TTF values did not result in significant changes in results. Based on this, it can be concluded that use of non-parametric distribution for TTR improves accuracy of outage duration in the generated data.

**Stage 3:** Once the basic model to generate scenarios is developed, the planned outage dataset was used and a set of heuristic rules were used to prevent overlap of generated FO with existing PO. The basic concepts of using the heuristic rules can be found in Section 3.4.2. PO data for each unit

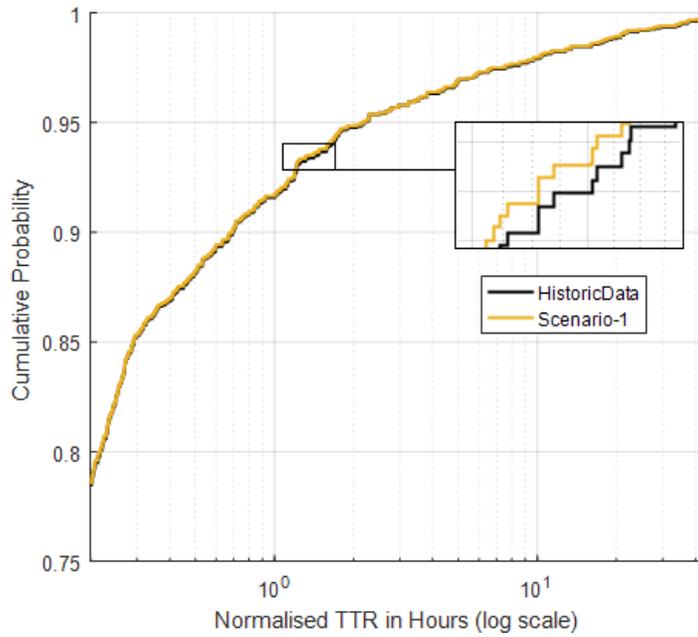
was obtained for 20 years of the historic record. PO corresponds to state codes 25 and 26 in the USR database. These values were replicated for 1000 years for each of the 10 units of GMS to obtain PO dataset.

MATLAB code was modified to prevent the overlap of FO duration and/or the start of FO within designated as PO. This was done by using an exclusion rule and TTR curtailment the scenario generation step, as described in detailed steps in Section 3.4.2. Basically, the heuristics is set so that no FO can originate in a period that is designated as PO (Figure 16) and if a TTR period overlaps with a PO period, then the TTR period is curtailed (Figure 18). However, if a TTF period completely crosses over a PO period then it is accepted (Figure 17). This method is in line with the definition of TTF in 2-state model, where TTF is defined from end of forced outage to beginning of next forced outage. The historic TTF values incorporates many PO periods and so it was necessary that the scenario generation method reflects that while preventing overlaps of FO and PO.

The results of scenario generation with PO heuristics for TTF and TTR are shown in Figure 50 and Figure 51.



**Figure 50 TTF FO-FO Model (PO Heuristics)**



**Figure 51 TTR FO-FO Model (PO Heuristics)**

An interesting but expected result was obtained after adding the PO heuristics. While the TTR distribution was not impacted by the curtailment of FO, the TTF distribution diverged, particularly in the upper quantiles. No change in the TTR distribution indicates that there are hardly any significant overlaps of input PO and FO generated. But a significant diverging trend in TTF starts at CDF value of about 0.6 which then increases until the end of the distribution. It can be seen that the output distribution is shifted to the right of the input distribution indicating that the percentage of higher TTF values have increased in the generated data as compared to historic values. For example, Figure 50 shows that 20% of TTF values in historic data exceeded 30 hours (normalised), but 25% of generated TTF values exceeded 30 hours duration. This result was expected because in this case conditional sampling is used. This conditional sampling process eliminates certain TTF values and replaces them with a larger value of TTF as illustrated in Section 3.4.2.

Other benchmarking indices were also impacted by the heuristics used. The median of annual duration was lower in the generated dataset by about 18% as compared to their historic value. The median of the annual number of outages was also lower in the generated data by about 11% as compared to their historic value. This FO-FO model under predicts both the frequency and duration of outages. One simple explanation for this gap in indices is the elimination of certain outages due to overlap with PO. This leads to slightly fewer outages every year, thus decreasing the mean annual duration in every year. The distribution of TTR is not affected significantly because it represents the statistical distribution of outage durations. The divergence in TTF ultimately leads to under prediction of outage impacts i.e., the annual hours lost to outages and the annual number of outages.

In summary, it can be concluded that a 2-state model is a simple way of scenario generation that can be used to represent stochasticity of historic data. However, the generated scenarios could result in an overlap with PO periods. Correction for these overlaps leads to elimination of certain generated outages adversely impacting the statistical properties of the output distribution, thereby decreasing the confidence of modelers in this method. Furthermore, this method is not designed to model the impact of planned outages on TTF as observed in Section 4.3.5. Therefore, to correctly account for PO and to reduce its impact in modeling, a different framework was adopted as discussed in the next Section.

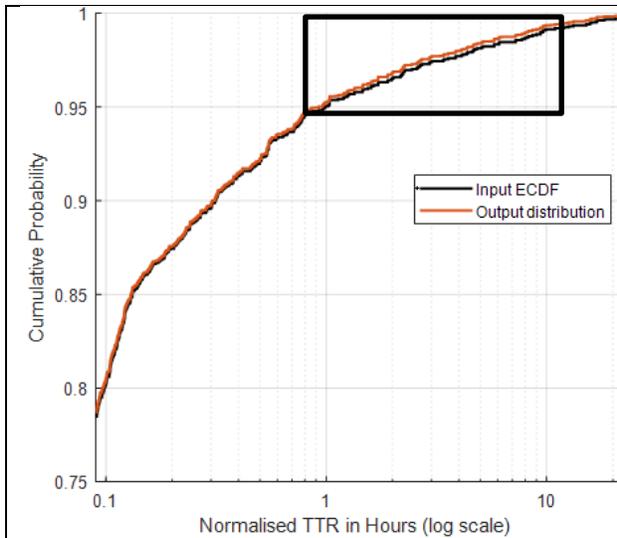
#### **4.4.3 Semi-Markov process PO-FO model**

As evident from the results of FO-FO model in the previous Section, there is a need to account for PO in a more comprehensive way other than by simply avoiding overlaps with FO. To include the impacts of planned outages on TTF, the Semi-Markov process of a simple FO-FO model was modified to a PO-FO model where the input database was revised, and the scenario generation method was altered as per the algorithm described in the methodology Section. To build the database of TTF\_FF and TTF\_PF, planned outage data was obtained from USR. A MATLAB code was written to sort forced outages into two categories. One set consisting of forced outages that were preceded by forced outages and another set of forced outages preceded by planned outages, as described in Section 3.2.4. This code also generated planned outage pattern for input in the scenario generation method. PO patterns in the 20 years of record were replicated for 1000 years for each of the 10 units of GMS to obtain a 1000-year PO dataset.

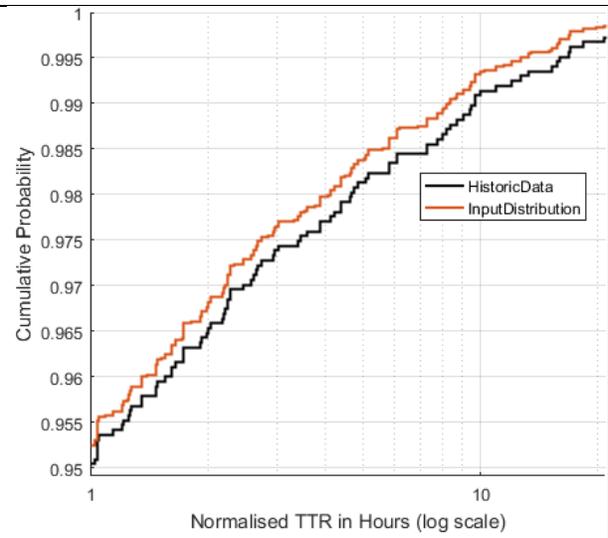
Most of the POs are usually scheduled in a cyclic manner, mostly in low load periods of late spring to early autumn. However, FO can happen anytime. In the historic database, it was observed that 80-85% of the FO for every unit was followed by another FO, while the remaining 15%-20% were followed by planned outages. The TTF definition was modified as explained in Section 3.2.4 to obtain TTF\_FF and TTF\_PF. The best fit parametric distribution on these values was derived in Section 3.2.5.

The algorithm described earlier in Section 3.4.3 was applied using MATLAB code. This code reads planned outage patterns as inputs, ECDF of TTR, TTF\_FF, TTF\_PF and parameters of TTF\_FF and TTF\_PF distributions. This algorithm assumes a system reset after POs and a value of TTF\_FF is sampled from the end of PO period. This was illustrated in Figure 20 and Figure 21. The assumption of system reset on encountering PO and sampling TTF from a different distribution after POs are the major differences of this PO-FO model from the FO-FO model discussed earlier.

The output of the model includes patterns of outages for each unit. From the output dataset, values of TTR, TTF\_FF and TTF\_PF were extracted and the distributions of these datasets are compared against historic values as shown in Figure 52, Figure 53, Figure 54 and Figure 55.

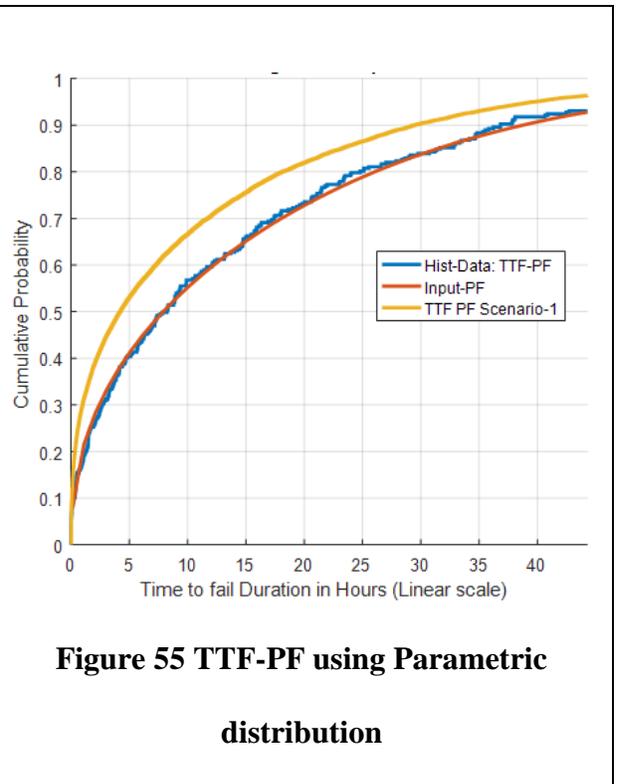
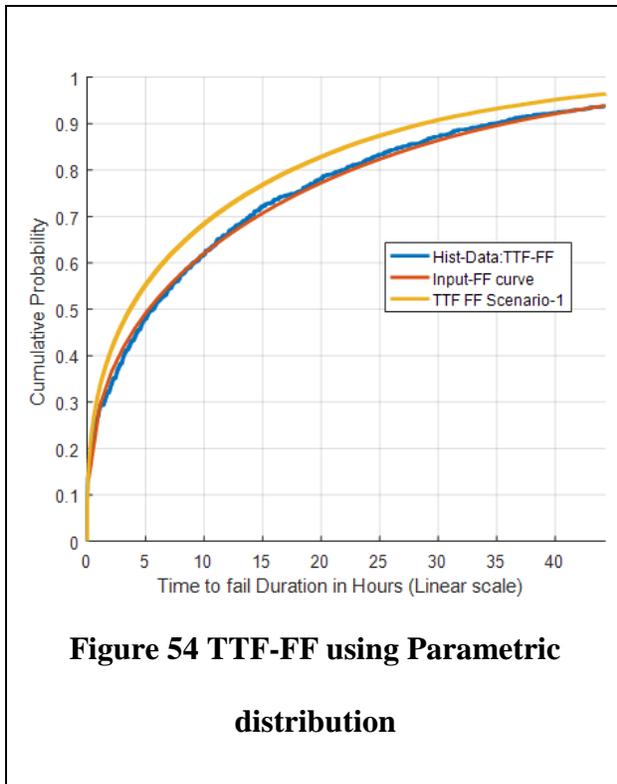


**Figure 52 TTR from PO-FO model**



**Figure 53 Bias in higher TTR values**

It can be seen that TTR values of the PO-FO input and output distributions exhibits a close match, with only a slight bias for very high outage duration as shown in Figure 53. This divergence can be attributed to the fact that a sampled TTR with a higher outage duration has a higher chance of getting curtailed. This leads to a slight shifting of generated ECDF towards left of the historic data ECDF. This small deviation means number of higher values of TTR would be slightly lesser in generated data as compared to historic data. But this difference is small enough for all practical considerations.



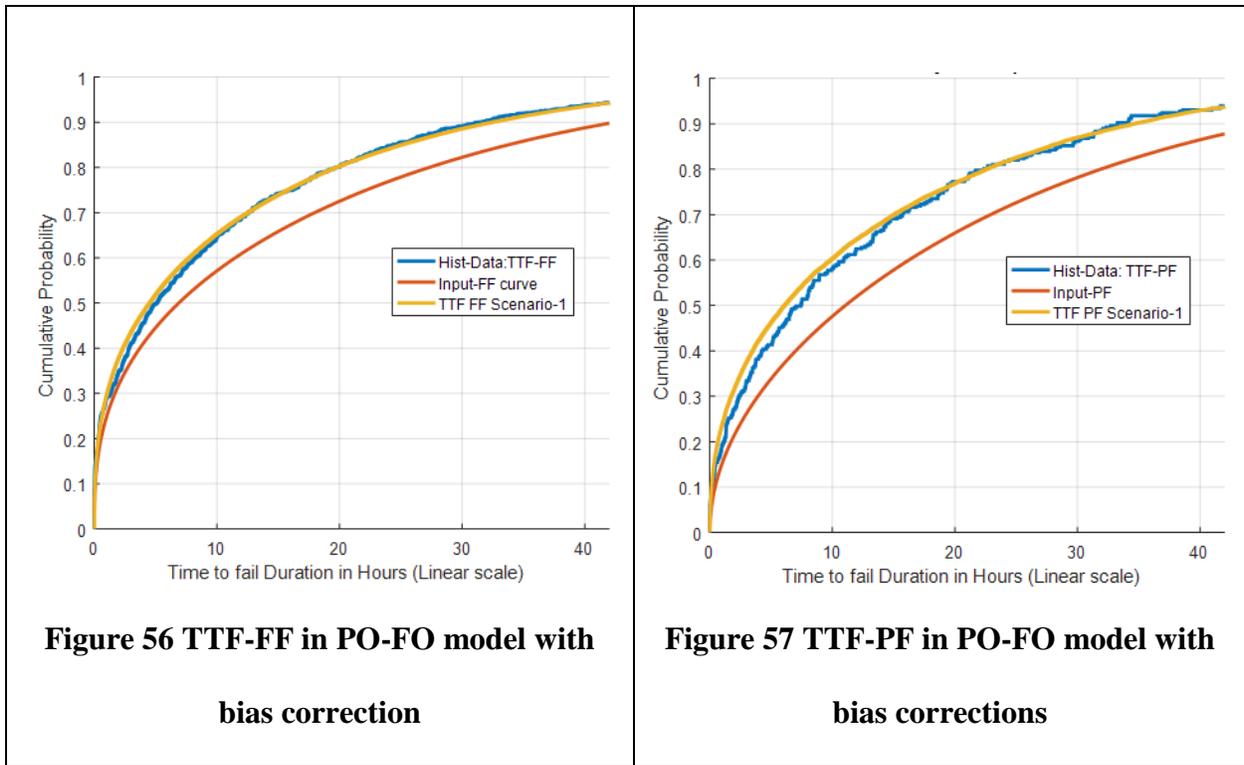
The results of the simulation for TTF\_FF and TTF\_PF are shown in Figure 54 and Figure 55. It can be seen that the results exhibit a systematic shift to the left in the output of the distribution of TTF\_FF scenarios. This shift can be attributed to the fact that the algorithm of checking for planned outage ends up eliminating significant number of TTF values, particularly with higher chance of rejection of higher TTF durations than lower TTF durations.

**Bias Correction:** The systematic bias between input and simulated output could be attributed to the conditional sampling from a distribution. This conditional sampling process resulted in eliminating a significant number of TTF which resulted in a bias in the sampling process. This bias can be easily corrected using a “synthetic input distribution”. From the initial results it was observed that the output and input have very similar distribution albeit, with an offset. This

observation provided reasons to believe that a synthetic input distribution derived from the actual input distribution could be used to obtain an output ECDF that will coincide with the historic ECDF.

Keeping this in mind, different values of parameters were tested on the input TTF\_FF to generate scenarios with the aim of finding input parameters that can best correct the bias and which produce a close overlap of input and output curves. The process can be automated using optimization methods to find the set of parameters of the distribution to minimize the bias. In this work a simple trial and error process was adopted initially to see if the desired objectives of scenario generation can be achieved. After a number of trails, and using difference of input and output at different quartile values as guiding criteria, one set of input parameters was found to perform best for TTF\_FF. The same process was then repeated for TTF\_PF to obtain synthetic input parameters of both TTF\_FF and TTF\_PF.

With these changes, the model was re-run with these synthetic input parameters and scenarios were generated and the results are presented in Figure 56 and Figure 57.



The red curves in the figures represent the synthetic input distribution which helped in obtaining an TTF\_FF and TTF\_PF output distributions (shown in orange) that matches the historic distribution (shown in blue). This was possible because the bias was systematic, and a good match that can be obtained by simply changing the parameters of the Gamma distribution.

After obtaining a good fit on TTF and TTR values, it is worth looking at other benchmarking indices discussed for the previous case. The median of the annual TTR values was lower in the generated data by about 10%. This is an improvement from the difference of 18% in the FO-FO model. Also, the median of the annual number of outages was found to be 8% higher than the historic median.

#### **4.5 Summary of scenario generation methods**

From the results of the three scenario-generation methods described above, PO-FO model appears to best represent the historic outages events in terms of TTR and TTF distributions, median annual duration of outage hours and median annual number of major outages.

A Markov Process model completely fails to represent TTR and TTF values. While the FO-FO model gives better results, the results show significant deviation from historic behaviour when correction for overlaps with planned outages is carried out. These deviations appear to cause the distributions to diverge as seen in Figure 50. However, the Semi-Markov, PO-FO model, exhibited a systemic bias which was can be easily corrected for. The corrected PO-FO model gave satisfactory results for the parameters tested.

## **Chapter 5: Discussion and Conclusions**

This chapter provides the main conclusions drawn from this research and recommendations for future work to be taken up in this field. A summary of the problem is provided including key takeaways for the development of an appropriate outage database and forced outage analysis methods. This is followed by the key takeaways from the statistical analysis and the choice of suitable probability distribution functions to describe forced outage in hydropower systems. Finally, the advantages and disadvantages of the three-scenario generation algorithms developed in this thesis are discussed. The last Section of this thesis chapter provides recommendations for future work to advance the concepts developed in this research.

### **5.1 Key takeaways for building an outage database**

Unit unavailability caused by forced outages of generating units is one of the sources of uncertainty in energy planning and in operational planning studies. Correct stochastic representation of forced outages is necessary because the net worth of energy produced by the system is impacted by the frequency of outages (failure rates) and by the duration of outages (repair rates).

A database on unit operating state and records should be built and maintained to gain insights about unit performance and outage behaviour. The database classifies outages in various categories. For better representation of the system behavior, outage data can be reclassified into deterministic sources of plant unavailability and stochastic sources of plant unavailability. For energy modeling studies executed with a 3-5-year planning horizon, maintenance outages should be treated as a stochastic source of unit unavailability. It was also shown that the statistical properties of the duration of maintenance outage were similar to those of forced outages, thereby

indicating that the underlying processes that impact forced outages also impact maintenance outages. Hence, maintenance outages and forced outages were merged together and classified as stochastic sources of unavailability and referred to as forced outages.

The concepts of reliability analysis were found to be helpful in probabilistic analysis of outage events. Frequency and duration was chosen as most suitable reliability index in comparison to other indices like Loss of Load Probability to analyse outage events. A two-state unit representation with one state being available and other being forced out, is commonly used for outage modeling of base load generating units. Using the two-state model, Time to Fail (TTF) and Time to Repair (TTR) were defined. The cumulative probability distributions of TTF and TTR help in quantifying and assessing the frequency and duration of forced outages in a two-state representation of generating units.

## **5.2 Conclusions from statistical analysis of data**

The statistical analysis carried out in this thesis helped in analyzing properties like trend, seasonality and independence of the historic outage data. While various statistical methods can be used to analyse the outage dataset, the current research shows the importance of asking the right questions and consulting with energy planners and modelers to verify modeling assumptions. The main conclusions of the statistical analysis of outage data are listed below:

- 1) Generating units are homogenous in terms of occurrence and frequency of outage events. The 10 units of GMS were found to have similar cumulative distributions of TTR and TTF values. Therefore, all TTF and TTR datasets for each of the 10 units could be combined in a single dataset for the analysis of TTR and TTF.

- 2) No significant trend in the last 20 years of outage record was observed. Most of the units are over 35 years old and some units are over 40 years old. The units have crossed their infant mortality period and are in their useful life period where assumptions of constant hazard rate are applicable. Individual units may have varying rate of failure, some experiencing too many failures due to wear and tear, and some experiencing fewer failures due to certain upgrades. Periodic maintenance of different units is usually scheduled in such a way that the random failure process exhibited stationarity for the entire GMS plant. Hence, assumption of stationarity is valid in frequency and duration of outages for power plants undergoing regular upgrades and preventive repairs.
- 3) Comparing empirical distributions (ECDFs) of TTR and TTF for different seasons was found to be a very pragmatic way of checking seasonality. Results showed that there was no seasonality in TTR dataset. The TTF values were found to be slightly longer in winter months, but the winter distribution was within 95% interval of the total annual distribution. Hence, for practical purposes, it was assumed that there was no seasonality for TTF distribution.
- 4) There was no correlation between a TTF value and the preceding TTR duration. Similarly, there was no impact of TTR on successive TTF. The analysis showed that both TTR and TTF are independent of each other. A small TTF can be followed by a large TTR and vice versa.
- 5) Impact of planned outages on TTR and TTF of forced outages had not been discussed much in literature. The database developed using a simple two state representation is not suitable to analyse the impacts of planned outages. A modified approach to quantify the impacts of planned outage on forced outage is presented in this thesis. It was found that planned outages have no effect on TTR distributions, but TTF distributions are impacted. It was found that TTF

values are higher if a forced outage is preceded by a planned outage (TTF-PF). This is in comparison to TTF values where forced outages are preceded by forced outages (TTF-FF).

The analyses to identify the most suitable probability distributions were assessed using the AIC/BIC criteria. It was found that the Gamma distribution was the best fit for TTF values, with the Weibull distribution being a close approximation. For TTR, the Generalised Pareto distribution was found to be the best fit with lognormal distribution being a close approximation. This analysis showed that the assumptions of exponential distribution, widely used in the theoretical analysis power systems, leads to incorrect and erroneous representation of both TTF and TTR values. Hence, the distribution of TTR and TTF should always be assessed using the historic dataset on generating units operating status.

### **5.3 Discussions on Scenario Generation Methods**

This research work analysed different Markov based and Semi-Markov based methods to model forced outages. Three different algorithms were developed to model the statistical properties of forced outages and to generate forced outage scenarios that could be used as inputs for energy studies.

Markov Chain - Monte Carlo Simulation (MC-MCS) was the first algorithm to be tested in the case study. The only inputs for this algorithm were the transition probability of generating unit to enter from one state to other. The output results required a post-processing step to convert the output to TTR and TTF values. The results confirmed that the assumptions of constant transition probabilities for events would give exponentially distributed TTR and TTF output values, which

are distinctly different from the historic input distribution. The most probable reason for this discrepancy was the assumption of constant transition probabilities. Although there is no distinct yearly trend or seasonality in the occurrence and duration of outage events, there are fluctuations within a year due to many factors like planned maintenance, load, start/stops, availability of manpower and material, transmission line outages, etc. All these factors make it difficult to model the outage patterns using a constant failure rates, and this makes Markov processes unsuitable to generate scenarios of outages for the GMS hydropower plant. It was also concluded that a homogenous Markov Process is also unsuitable to model forced outage events if the state of units is represented by a simple two-state model.

The Semi-Markov process provides the flexibility to sample duration of a state from any distribution. This process required the probability distribution of TTR and TTF as inputs. Two other algorithms were developed based on Semi-markov process.

The second algorithm investigated in this research was based on a Semi-Markov based FO-FO model. It did not consider the impact of planned outages on TTF values. Trials using parametric CDF and ECDF showed that it was better to use ECDF for sampling TTR and parametric CDF for sampling TTF. This model was able to correctly reproduce the input TTR/TTF distribution. However, application of a heuristic approach to prevent overlaps between generated forced outages and existing planned outages impacted the final distribution of TTF. Preventing overlap of planned and forced outage required the addition of a conditional rule in the sampling process. Hence, the sampling output of the TTF distribution deviated from the input TTF distribution. It was found that the FO-FO method with planned outage heuristics leads to under-prediction of the number of

outages per annum and the annual outage duration in the generated data as compared to the historic data.

The third model developed was the Semi-Markov PO-FO model to overcome the problems encountered in FO-FO model by basically introducing more complex heuristics to comprehensively account for the impacts of planned outages on TTF. Parametric CDFs were used for the two TTFs and ECDF was used for TTR as input. The results showed a systemic difference between the input and output distributions of both TTFs. However, this difference was not the same as the deviation in TTF in FO-FO model. In the FO-FO model, the difference between input and output was diverging as TTF values increased, whereas in PO-FO model exhibited a very systemic and apparently constant bias. This bias was easily corrected for by modifying the parameters of the input TTFs such that the historic distribution coincided with the output distribution. In the output TTR, there was a very slight deviation from the input distribution towards the end of the upper tail of the distribution. This difference resulted from the curtailment of TTR values to avoid overlap with planned outages but for practical purposes it was deemed to be insignificant. The results of PO-FO model showed improvement from FO-FO model in terms of annual number of outages and annual duration of outage hours.

#### **5.4 Recommendations:**

From the research conducted and the three scenario generation algorithms tested, the following recommendations are made for modeling forced outages:

- 1) Historic outage data should be used for development of the TTR and TTF database.

- 2) Statistical analysis should be carried on the developed database to develop and select a suitable probability distribution and select appropriate modeling algorithm.
- 3) Homogenous Markov chains should not be used to generate forced outage scenarios unless the input TTR and TTF exhibits the properties of exponential distributions.
- 4) The only disadvantage of using the FO-FO model is the under-prediction of frequency and duration of forced outages due to overlap of generated forced outages with existing planned outages. The PO-FO model was developed to generate scenarios that are superior to the FO-FO model and to comprehensively account for the statistical properties of forced outages.
- 5) Consultations with the system planning modelers, who will be utilising the generated scenarios, are invaluable at every stage of model development.

## **5.5 Future Work**

The PO-FO model developed in this work can be improved upon as a part of future research in this domain. The following tasks can be carried out:

- 1) In this research work, a 1000-year scenario of forced outages was developed because the purpose was to focus on statistical significance of scenario generation algorithm. If scenarios are generated for smaller periods like 5 years, then hundreds of scenarios would have to be generated to account for sampling variability. There is a lot of scope in this area to find the minimum number of scenarios that would represent the entire range of variability witnessed in the historic data. There is a trade-off between number of scenarios required for stochastic representation and number of scenarios that can be practically modeled in energy studies.

- 2) In this work, a manual trial and error approach was adopted to correct the systemic bias in the PO-FO model to obtain the final results. This step can be automated by using an optimization technique that would find the adjusted input parameters for CDFs of TTF\_FF and TTF\_PF, and that minimizes the difference between historic TTF values with generated TTF values.
- 3) This scenario generation algorithm can be applied to generate outage scenarios for other power plants in the BC Hydro system to verify and generalize the conclusions drawn from this case study.
- 4) And finally, the different scenarios algorithm developed in this work can be applied in various simulation and optimization models for energy studies, operations planning and for long-term planning studies of the BC Hydro system.

Finally, it can be concluded that the statistical investigation and development of reliability-based scenario generation methods described in this thesis achieved all the research goals listed in Chapter 1. The relevant literature on the reliability analysis of power system was carried out with an emphasis on the research work done conducted on modeling forced outages of generating units. Then historic unit operating records were used to develop the outage database. The statistical properties of outage events were explored, and suitable probabilistic models were developed. Three scenario generation algorithms were developed, and the PO-FO model was accepted as the most comprehensive approach to generate scenarios of forced outages for BC Hydro energy studies.

## Bibliography

- Agrawal, Abhishek, Quentin Desreumaux, Ziad Shawwash, and Mehretab Tadesse. 2017. "Modeling Forced Outages of Hydropower Generator Units for Reliable Dam Operation." In *Canadian Dam Association (CDA) Conference*, 2017.
- Alkuhayli, Abdulaziz A., Srinath Raghavan, and Badrul H. Chowdhury. 2012. "Reliability Evaluation of Distribution Systems Containing Renewable Distributed Generations." *North American Power Symposium(NAPS)*, 2012 (1): 1–6.
- Anderson, C. Lindsay, and Matt Davison. 2005. "An Aggregate Weibull Approach for Modeling Short-Term System Generating Capacity." *IEEE Transactions on Power Systems* 20(4): 1783–89.
- Armstrong, M., Tang, Z.D. and Kusuma, G. (2016). Accessed January, 2018  
<https://www.hydroworld.com/articles/hr/print/volume-35/issue-8/cover-story/expanding-into-the-future-gm-shrum-station-service-rehab.html>
- Barroso, L.a., and A.J. Conejo. 2006. 2006 IEEE Power Engineering Society General Meeting *Decision Making under Uncertainty in Electricity Markets*.  
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1709323>.
- BC Hydro. 2016. *British Columbia Hydro and Power Authority - Annual Service Plan Report 2016/17*.
- Begovic, Miroslav, Joshua Perkel, and Rick Hartlein. 2006. "Equipment Failure Forecasting Based on Past Failure Performance and Development of Replacement Strategies." *Transactions on Electrical and Electronic Materials* 7(5): 217–23.  
[http://kieeme.or.kr/storage/journal/T/7\\_5/10607/articlefile/article.pdf](http://kieeme.or.kr/storage/journal/T/7_5/10607/articlefile/article.pdf).
- Billinton, R., R. Karki, and A. K. Verma. 2013. *Reliability and Risk Evaluation of Wind*

- Integrated Power Systems*. <https://books.google.com/books?id=yflHAAAAQBAJ&pgis=1>.
- Billinton, Roy, and Ronald Norman Allan. 1992. *Reliability Evaluation of Engineering Systems*.
- Billinton, Roy, and Jingdong Ge. 2004. "A Comparison of Four-State Generating Unit Reliability Models for Peaking Units." *IEEE Transactions on Power Systems* 19(2): 763–68.
- Billinton, and W. Li. 1994. *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*. <https://books.google.com/books?hl=en&lr=&id=qakACAAAQBAJ&pgis=1>.
- Binder, Gerald E et al. 1991. "Predicting Unit Availability: Top-Down Analyses for Predicting Electric Generating Unit Availability."
- Boomsma, Kristoffersen, and Trine Krogh. 2008. *1666 WP3 Prototype Development for Operational Planning Tool*.
- Bornak, Lars Pauli. 2013. "Modelling Energy Systems - A Study on Power System Reserves in South Africa."
- Braglia, Marcello, Gionata Carmignani, Marco Frosolini, and Francesco Zammori. 2012. "Data Classification and MTBF Prediction with a Multivariate Analysis Approach." *Reliability Engineering and System Safety* 97(1): 27–35. <http://dx.doi.org/10.1016/j.ress.2011.09.010>.
- Canadian Electricity Association. 2016. *Generation Status Manual*.
- Van Casteren, Jasper F L, Math H J Bollen, and Martin E. Schmiegel. 2000. "Reliability Assessment in Electrical Power Systems: The Weibull-Markov Stochastic Model." *IEEE Transactions on Industry Applications* 36(3): 911–15.
- Cepin, Marko. 2011. *Assessment of Power System Reliability*.
- Curley, G Michael. 2013. "Reliability Analysis of Power Plant Unit Outage Problems."
- Dalal, Gal, Elad Gilboa, Shie Mannor, and Louis Wehenkel. 2018. "Chance-Constrained Outage

- Scheduling Using a Machine Learning Proxy.” : 1–11. <http://arxiv.org/abs/1801.00500>.
- Das, D., and B.F. Wollenberg. 2012. “Minimizing Forced Outage Risk in Generator Bidding.” *EUROPEAN TRANSACTIONS ON ELECTRICAL POWER* 36(E1): 51–66.
- David, Trindade. 1996. “Confirming Trends in Repairable System Reliability.” *ASA Joint Statistical Meetings*.
- Ellis, K.E., and G.J. Gibson. 1991. “Trend Analysis of Repair Times.” *Annual Reliability and Maintainability Symposium. 1991 Proceedings*: 85–92.  
<http://ieeexplore.ieee.org/document/154419/>.
- Endrenyi, J. 1979. *Reliability Modeling in Electric Power Systems*. Wiley.  
<https://books.google.ca/books?id=AIGRQgAACAAJ>.
- ERIS-CEA. 2012. *Generation Equipment Status Annual Report 2007*.
- Finger, Susan. 1979. “Electric Power System Production Costing And Reliability Analysis Including Hydro- Electric, Storage, And Time Dependent Power Plants.” *MIT Energy Laboratory Technical Report #MIT-EL-79-006* (February).
- Hall, J D, and R J Ringlee. 1968. “System Reliability Calculations : I-Generation System Model.” (9): 1787–96.
- Karson, Marvin. 1968. “Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967, \$9.00.” *Journal of the American Statistical Association* 63(323): 1047–49.  
<https://doi.org/10.1080/01621459.1968.11009335>.
- Konstantopoulos, Takis. 2009. “Markov Chain and Random Walks.” *Markov Chain and Random Walks*: 1–122. <http://www2.math.uu.se/~takis/L/McRw/mcrw.pdf>.

- Koval, D.O., and A.A. Chowdhury. 1994. "Generating Peaking Unit Operating Characteristics." *IEEE Transactions on Industry Applications* 30(5): 1309–16.  
<http://ieeexplore.ieee.org/document/315244/> (December 4, 2017).
- Kuha, Jouni. 2004. "AIC and BIC: Comparisons of Assumptions and Performance." *Sociological Methods and Research* 33(2): 188–229.
- Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root." *Journal of Econometrics* 54(1–3): 159–78.  
<http://www.sciencedirect.com/science/article/pii/030440769290104Y>.
- Li, Yan-fu, and Enrico Zio. 2012. "A Multi-State Model for the Reliability Assessment of a Distributed Generation System via Universal Generating Function." *Reliability Engineering and System Safety* 106: 28–36. <http://dx.doi.org/10.1016/j.ress.2012.04.008>.
- Li, Yan, Lirong Cui, and Cong Lin. 2017. "Modeling and Analysis for Multi-State Systems with Discrete-Time Markov." *Reliability Engineering and System Safety* 166(March): 41–49.  
<http://dx.doi.org/10.1016/j.ress.2017.03.024>.
- Li, Z, and Y Wang. 2017. "Modeling of Time-Varying Forced Outage Rate for Power Transformers Considering Online Monitoring Information." *Dianli Xitong Zidonghua/Automation of Electric Power Systems* 41: 63–68.
- Lindsay, J. K. 2001. "The Statistical Analysis of Stochastic Processes in Time."
- Lisnianski, Anatoly, David Elmakias, David Laredo, and Hanoch Ben Haim. 2012. "A Multi-State Markov Model for a Short-Term Reliability Analysis of a Power Generating Unit." *Reliability Engineering and System Safety* 98(1): 1–6.  
<http://dx.doi.org/10.1016/j.ress.2011.10.008>.

- M. Milligan et al. 2010. “Operating Reserves and Wind Power Integration: An International Comparison.” *Proc. of the 9<sup>th</sup> Int’l Workshop on Large-Scale Integration of Wind Power into Power Sys.* (October 2010).
- Moatti, Marie. 1988. “Discretization and Markov Modeling of a State Variable in Dynamic Programming.” In *System Modelling and Optimization*, eds. Masao Iri and Keiji Yajima. Berlin, Heidelberg: Springer Berlin Heidelberg, 530–38.
- NERC. 1992. *Publications-Generating-Unit-Availability-Following-Planned-Outages.pdf*.  
 ———. 2015. *Generating Availability Data System: Data Reporting Instructions*.
- Oak Ridge National Lab. 2014. *2014 Hydropower Market Report*.
- Parrish, Mark. 2015. “Quantifying the Cost of Unit Outages Across the USACE Hydropower Facilities.” (January): 1–7.
- Perrica, Giuseppe, Gabriele Goldoni, and Federico Raimondi. 2009. “Time to Failure and Time to Repair Profiles Identification.” : 1–6.
- Prada, José Fernando. 1999. “The Value of Reliability in Power Systems: Pricing Operating Reserves.” *Mit El 99-005 Wp* (June): 79p.
- R. Billinton, R. Ghajar. 2001. “THE CANADIAN ELECTRICAL ASSOCIATION APPROACH TO TRANSMISSION AND DISTRIBUTION EQUIPMENT RELIABILITY ASSESSMENT.” (406): 76–85.
- Raychaudhuri, S. 2008. “Introduction to Monte Carlo Simulation.” *Simulation Conference, 2008. WSC 2008.*: 91–100. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4736059](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4736059).
- Rondla, Preethi. 2012. “Monte Carlo and Analytical Methods for Forced Outage Rates Calculation of Peaking Units.”
- Ryan, S. M., and M. Mazumdar. 1990. “Effect of Frequency and Duration of Generating Unit

- Outages on Distribution of System Production Costs.” *IEEE Transactions on Power Systems* 5(1): 191–97.
- Ryan, S., and M. Mazumdar. 1990. “Effect of Frequency and Duration of Generating Unit Outages on Distribution of System Production Costs.” *IEEE Transactions on Power Systems* 5(1): 191–97.
- Scully, A. et al. 1992. “Using a Semi-Guided Monte Carlo Method for Faster Simulation of Forced Outages of Generating Units.” *IEEE Transactions on Power Systems* 7(3): 1313–21.
- Simonoff, Jeffrey S et al. 2005. *Electricity Case : Statistical Analysis of Electric Power Outages CREATE Report New York University-Wagner Graduate School , Institute for Civil Infrastructure Systems.*
- Unit Status Recording Primary Reference, 2009, BC Hydro Intranet, Accessed August, 2017.
- Wang, L, N Ramani, and T C Davies. 2004. “FOR PEAKING AND CYCLING OPERATIONS.” (7): 2004–11.
- Zhou, Jiyi (University of British Columbia, Vancouver). 2015. “Reliability-Based Hydro Reservoir Operation Modeling.” University of British Columbia.