12-2022

A Hydropower Facility as an Energy Water Signal Processor

Asha Shibu
University of Tennessee, Knoxville, ashibu@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Asha Shibu entitled "A Hydropower Facility as an Energy Water Signal Processor." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Srijib K Mukherjee, Major Professor

We have read this dissertation and recommend its acceptance:

Fangxing Li, Russell Zaretzki, Brennan T. Smith

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Hydropower Facility as an Energy Water Signal Processor

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Asha Shibu

December 2022

Dedication

To my father, Shibu, for believing in me even when I was nothing near to believing in myself, and to my mother, Bindu, for always being my comfort when I needed it the most.

All that I am, I owe to both of you!

Acknowledgements

"It takes a village to write a dissertation"

I'm deeply indebted and forever grateful to the following people who have helped me undertake this research. First and foremost, I'd like to express my profound gratitude to Dr. Brennan Smith, whose expertise and insight into the subject steered me through this research. Without your help and guidance, this dissertation would not have been possible. I want to thank my advisor, Dr. Srijib Mukherjee, for motivating me and allowing me to grow as a researcher. Your advice on both research and my career has been invaluable. Also, I must thank my committee members, Dr. Fran Li, and Dr. Russell Zaretzki, for their valuable feedback and interest in my work. I cannot forget friends who supported me during difficult times, encouraged me, and occasionally just listened: Sruthi Ashraf, Kaustubh Vibhandik, Kaustubh Mungale, Purva Mallya, Sandhya Mohanakumar, Vaikrant Pandav, and Thomas Savarino. Nobody has been more important to me in this journey than my family and I want to thank my parents, Shibu George and Bindu T.G., for their unwavering support, love, and guidance. I'm so thankful to my grandparents, T.N. Gopalakrishnan Nair and Indira Bai, for always bringing a smile to my face. Finally, I thank with love Stalin Sebastian, who helped me get through trying times in the most positive way by keeping things in perspective, even when I felt like I couldn't go on.

Abstract

In recent times, various efforts have been made to address the challenge of adequately representing hydropower systems in modeling frameworks, accounting for the lack of data to represent the multiple constraints in hydropower operation. This research is a pilot data-driven methodology for characterizing, classifying, and comparing the water-to-energy and energy-to-water signal transformations that hydropower facilities as signal processors accomplish. In this study, a Box Jenkins transfer function/noise model is used to identify the relationship between reservoir inflows and outflows. For examining the feasibility of this methodology, 5-minute fleet data for five storage and five run-of-river facilities was provided by the Tennessee Valley Authority (TVA) and transfer function models are developed. The influence of past inflow and outflow values on the current outflow decisions was investigated and summarized by examining the results of Box Jenkins methodology. Finally, dominance analysis was introduced to add value to the Box Jenkins model results and provide different stakeholders with a set of concepts to convey the functionality of hydropower.

Table of Contents

Chapter 1: Introduction 1

1.1 Objectives and Problem Definition 1

1.2 Overview of Hydropower Generation 2

1.2.1 Classification of Hydropower facilities 2

1.2.2 Hydropower operation and the constraints involved 4

1.3 Relevance of the research 8

1.3.1 Role of Hydropower in future power systems 8

1.3.2 Water system perspective of hydropower operation 8

1.3.3 Power system perspective of hydropower operation 9

1.3.4 Using data to validate representation 10

1.4 Research Objectives 11

1.5 Thesis Outline 11

Chapter 2: Literature Review 13

2.1 Background 13

2.2 Hydropower modeling and fidelity 13

2.2.1 Watershed Models 14

2.2.2 Dispatch Models 15

2.2.3 Production Cost Models 16

2.3 Hydropower representation: RiverWare vs. PLEXOS 17

2.3.1 Representation of hydropower in RiverWare 17

2.3.2 Reservoir modeling in RiverWare 18

2.3.3 Representation of Hydropower in PLEXOS 20

2.3.4 Reservoir Modeling in PLEXOS 20

2.3.5 Drawbacks in the representation 21

2.4 Summary 22

2.5 Significance of the Study 23
2.5.1 Characterization of hydropower – need for a common rubric/nomenclature 23
2.5.2 Modeling water and energy systems 23
2.5.3 Relevance of time series analysis 24
2.6 Statistical Modeling Techniques 25
2.6.1 Supervised Learning 26
2.6.2 Unsupervised Learning 26
2.6.3 Selecting the required model 26
2.6.4 Objectives of regression models 27
2.6.5 Linear regression models 27
2.6.6 Dynamic Regression models 28
2.7 Box-Jenkins ARIMA Modeling 28
2.7.1 Stationarity 29
2.7.2 Differencing 29
2.7.3 Correlation 30
2.7.4 Stationary Models 30
2.7.5 Backshift Notation 32
2.7.6 Differencing Operator 32
2.8 The Box-Jenkins Methodology 33
2.8.1 Information based criteria for model identification 37
2.9 Application of Transfer Function Models 37
Chapter 3: Methodology *40*
3.1 Research Hypothesis 40
3.2 Transfer function modeling 42
3.2.1 Identification Stage 44
3.3 Identification of the transfer function model 47
3.3.1 Identification of a parsimonious noise model 49
3.3.2 Estimation Stage 49

3.4 Diagnostic Checking **50**

Chapter 4: Data and Analysis *51*

4.1 Data Availability **51**

4.1.1 Fleet Data **51**

4.2 Data Preprocessing **51**

4.2.1 Description of the study area- Norris hydropower facility **51**

4.2.2 Inflow computation **53**

4.2.3 Choosing the appropriate time step **62**

4.3 List of software used **62**

4.4 Analysis **64**

4.5 Application of the Box- Jenkins Approach to data from storage facilities **65**

4.5.1 Transfer function Modeling **65**

4.6 Application of the Box- Jenkins approach to data from run-of-river facilities **70**

4.6.1 Transfer function modeling **70**

4.7 Checking the fitted models **75**

Chapter 5: Results and Discussion *76*

5.1 General results of transfer function modeling **76**

5.1.1 Storage Facilities **76**

5.1.2 Run-of-River Facilities **78**

5.2 Discussion – Interpretation of Box Jenkins model results **78**

5.2.1 The intuition behind the transfer function coefficients: inflow and outflow dynamics **78**

5.2.2 Comparing transfer function coefficients – Dominance Analysis **87**

5.2.3 Practical application of this research **89**

5.2.4 Limitations and scope for future studies **93**

5.3 Summary **94**

Chapter 6: Conclusions and Recommendations *95*

6.1 Outcomes from Box Jenkins representation 95

6.2 Research Contributions 96

6.2.1 Recommendations for further research 96

References 98

Appendix A: Time series plots 106

Storage Facilities 106

Run-of-river facilities 111

Appendix B: Residual analysis plots 116

Storage Facilities 116

Run-of-river facilities 121

Vita 126

List of Figures

Figure 1: Components of a hydropower facility 3
Figure 2: The three major domains in hydropower operation 7
Figure 3: ACF plot for a white noise series 35
Figure 4: Box Jenkins methodology 36
Figure 5: Proposed transfer function model 41
Figure 6: Schematic representation of a linear system 43
Figure 7:Procedure for estimating impulse response function 46
Figure 8: Examples of impulse response functions of commonly adopted transfer functions 48
Figure 9:Norris Dam 52
Figure 10:Operating guide of Norris dam 54
Figure 11: Synopsis of operation data 54
Figure 12:Curve-fitting- Reservoir volume as a function of elevation 55
Figure 13:Computed reservoir inflow 57
Figure 14: Computed reservoir Inflow – data 57
Figure 15:Recorded elevation levels 58
Figure 16:Computed net inflows- daily and weekly 58
Figure 17: Elevation levels after smoothing 59
Figure 18: Replacing negative inflow values -different stages 61
Figure 19: Inflow across different time intervals 63
Figure 20: Map showing location of storage facilities 66
Figure 21: CCF Plots of storage facilities 69
Figure 22: Map showing locations of run-of-river facilities 71
Figure 23: CCF plots of run-of-river facilities 73
Figure 24: Inflow-Outflow dependance 81
Figure 25: The TVA Water Control System 82
Figure 26: Results of Dominance Analysis- Douglas Dam 88
Figure 27: Factors significant for model interpretation 92

List of Tables

Table 1: List of Ancillary services and their functions 5

Table 2: Workspace objects in RiverWare related to hydropower 19

Table 3:ACF and PACF patterns for ARMA 35

Table 4: Hydropower projects studied 52

Table 5: Results of correlation analysis 61

Table 6: Storage facilities studied 66

Table 7: Pre-whitening of Inflow time series- storage facilities 67

Table 8: Identified Transfer- Function Models: storage 71

Table 9: Run-of-river facilities studied 72

Table 10: Pre- whitening of Inflow time series- run-of-river facilities 72

Table 11: Identified Transfer- Function Model: run-of-river 74

Table 12: Diagnostic checks on parameter estimates- storage facilities 77

Table 13: Inflow-Outflow relationship of storage facilities 77

Table 14: Diagnostic checks on parameter estimates- run-of-river facilities 79

Table 15: Inflow-Outflow relationship of run-of-river facilities 79

Table 16: Effect of coefficients and lags 81

Table 17: Relative importance: run-of-river facilities 90

Table 18: Relative importance: storage facilities 90

Chapter 1: Introduction

"*All models are wrong, but some are useful*."

George E. P. Box.

1.1 Objectives and Problem Definition

Due to evolving electrical grid conditions, hydropower operations have undergone substantial changes in the past decade. The electrical power grid and hydropower basins are correlated complex systems with competing objectives and multiple constraints; therefore, proper characterization of hydropower operation is crucial in a model of an energy system. Additionally, suppose hydropower is to be employed as a default driver for flexibility requirements of the electric grid. In that case, the representation of hydropower generation in energy system models with respect to water dynamics (inflows and outflows) should be considered for better representation within electrical grid models. A driver of this is also the ever-evolving climate change conditions, which impact river patterns with droughts and floods.

There are multiple optimizations and rule-based water management models available to assist stakeholders in decision-making. Many initiatives have been made to address the issue of proper representation of hydrological and energy systems in modeling frameworks. However, it is evident that there are fundamental disparities in the ways hydropower is represented in the existing watershed, dispatch, and production cost models((Stoll, Andrade, Cohen, Brinkman, & Brancucci Martinez-Anido, 2017),(Voisin, Bain, Macknick, & O'Neil, 2020)). This research is a pilot data-driven methodology for characterizing, classifying, and comparing the water-to-energy and energy-to-water signal transformations that hydropower facilities as signal processors accomplish. Success in this effort is being reviewed by multiple industries and research partners, and advisors and includes:

1. The ability of the methodology to distinguish between intuitively different facilities (e.g., storage versus run-of-river),
2. The feasibility and efficacy of the methodology to be applied to facilities in different regions and contexts,
3. The feasibility (including data availability) of automating and scaling the methodology for application to the entire North American hydropower fleet, and

4. The extent to which the resulting "hydropower signal processor parameters" are intuitively and quantitatively linkable to conventional methods such as production cost modeling (e.g., modes of operation for hydropower facilities) and water balance modeling, routing, and scheduling.

The proposed method uses a time series modeling approach to derive a transfer function that models the water and energy transformations that hydropower plants are expected to accomplish as signal processors. The goal of this research is not to develop a new model for how hydropower interacts with power systems; instead, it will provide practitioners such as system modelers, grid operators, and other stakeholders with well-informed concepts to help them understand and improve the functionality and value of hydropower in their existing efforts.

1.2 Overview of Hydropower Generation

Hydropower generates electricity by utilizing water stored in a reservoir to spin a turbine. A typical hydropower facility includes the civil structures, hydraulic conveyance facilities (head race, headworks, penstock, gate, valves, and tailrace), the turbine-generator unit, electrical components (transformer, instrumentation, and controls, switchgear), and transmission lines (Gulliver & Arndt, 1991). The different components of a hydropower facility are shown in Fig. 1. Water released through the dam spins the turbine and converts the potential to mechanical energy, turning a generator connected to the turbine. There are different ways to classify hydropower facilities. Common types are discussed below.

1.2.1 Classification of Hydropower facilities

On a local scale, hydropower plants serve multiple objectives:

- flood control,
- water supply (industrial, public, and drinking water),
- electricity generation and irrigation.

However, on a national scale, a further classification can depend on the hydrological cycle. Nations with extreme volatile climatic conditions support low head hydropower generation while cascade or diversion type facilities may be favored in countries significant annual precipitation rates. Some common types of classifications include:

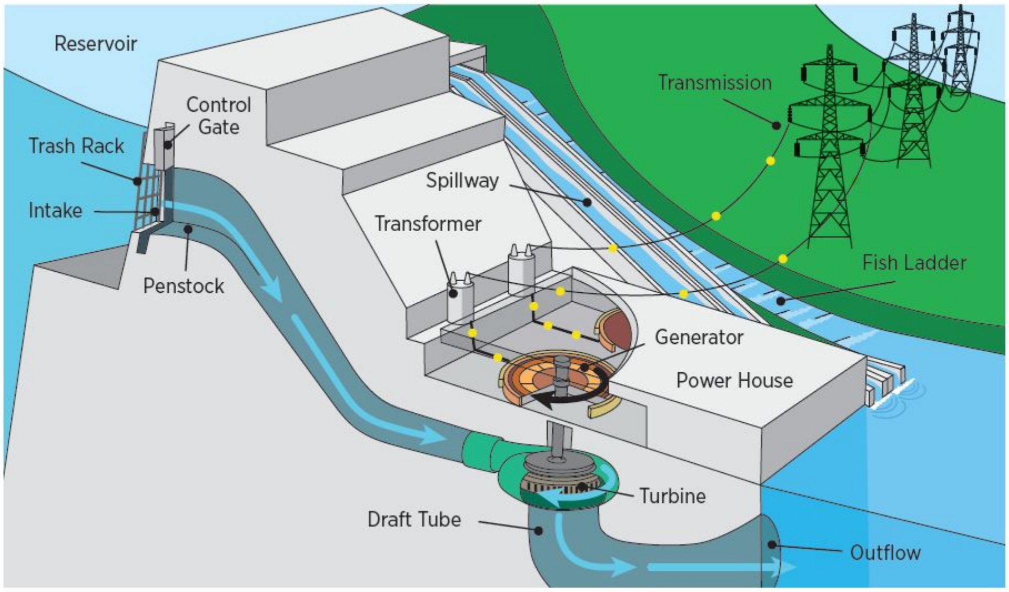


Figure 1: Components of a hydropower facility

(O'connor et al., 2016)

1. Classification based on head: One of the key elements in determining the amount of electricity generated by a hydropower facility is the hydraulic head, which is the elevation at which water is maintained in a reservoir. Hydropower facilities can be classified into high head, medium head, and low head. There is no standard on the exact range of values that fall in these categories, and it varies in different countries. According to Majumder et.al, low head facilities utilize heads of less than 30 m while medium- head facilities have head values falling under 30-300 m (Majumder & Ghosh, 2013).
2. Classification based on capacity: Depending on the power generation capacity, hydropower plants can be divided in micro, small and large hydropower facilities. Like head, there is no universally accepted definitions for capacity size within the U.S. Micro hydropower systems generate up to 100 kW of electricity while small while hydropower up to 10 MW can be classified as small. Plants with capacities larger than 10 MW are defined as large hydropower plants (Johnson, Hadjerioua, & Martinez, 2015).
3. Classification based on availability of water flow: Based on the quantity of water available, hydroelectric facilities are categorized as run-of-river facilities and storage facilities. Run-of-river facilities are unable to store water and must release water as it comes. In storage facilities, water is stored behind the dam and is available for generation as required. Storage facilities have greater operation range. They also provide a wide range of energy services such as base load and peak load and can dispatched to provide energy production when it is most valuable to the power system. Run-of-river facilities typically function as base-load power plants, with the generation varying according to water availability.

1.2.2 Hydropower operation and the constraints involved

Hydropower generation depends on water availability, which may vary by time and reservoir type. Many water control projects provide services beyond electricity generation, including flood and drought management, irrigation, navigation, recreational services, and water supply. Only 25 percent of reservoirs globally have hydropower generation as their primary purpose. (Uria-Martinez et al., 2021) Hydropower is one of the most flexible sources of electricity generation in the power grid. It has the ideal properties to provide flexible operation (as shown in Table 1) that can support the integration of renewable energy sources in the grid.

Table 1: List of Ancillary services and their functions

Ancillary Service	**Function**
Spinning Reserve	Reserve capacity that is online and synchronized to the grid. Capable of meeting system demand within 10 minutes of a dispatch instruction
Non-Spinning Reserve	Offline generation capacity can be ramped to capacity and synchronized to the grid within 10 minutes of a dispatch instruction
Voltage Support	Ability to produce or absorb reactive power
Regulation and Frequency Response	Capability to provide continuous balancing of the generation with load and to maintain the system frequency
Black Start Capability	The capability of a generating unit to go from shutdown to operating condition and generate without assistance from a power system.

According to a recent hydropower value assessment undertaken by the U.S. Department of Energy (DOE), even though hydropower accounts for just around 10% of total US generating capacity, hydropower turbines account for 40% of units assessed as capable of providing black start services to restore power systems operations (Ingram, 2019). In a storage hydropower project, a reservoir is constructed behind a dam to store water for the purpose of generating electricity as well as to provide ancillary grid services. Therefore, it also contributes to the grid's overall stability and reliability. Because they participate in both water systems and power systems, hydropower plants are subject to a wide range of constraints on their operation. As a reservoir, hydropower facilities encounter ecological and regulatory constraints such as availability of water rights, reservoir level restrictions, spillage limitation, and water quality concerns. The effects of water releases and their timing on the participants downstream, including aquatic species, their food chains, and other water users, must be considered while considering the operation of a hydropower project. Hydropower plants thus have the potential to adversely affect both the water availability and water quality in the specific habitats where aquatic organisms can develop and thrive.

The primary operational constraints for hydropower are connected to the equipment's capabilities, such as the minimum and maximum power that can be produced and the frequency of maintenance. Some turbine types have a minimum amount of power required to operate the turbine and a maximum power capacity based on the turbine's rating. Finally in deregulated markets hydropower projects may have power purchase agreements in place. These contracts call for a facility to supply a specific amount of energy at a set cost. Due to contractual obligations to supply a certain amount of power regardless of economic and market conditions, such agreements place further restrictions on how hydroelectric operations can be conducted(Stoll et al., 2017). The operational decision-making thus relies upon several uncertainties. In a multi-purpose reservoir, the tradeoffs that make decisions beneficial for one purpose and detrimental for another are often identified only be discerned through analysis of historical data in conjunction with reservoir operating studies. Over the past few decades, significant changes in hydropower operations have called for its better representation in the power system models. Consistent hydropower representation across the three domains as shown in Fig. 2 (environmental, operational, and power system services) is imperative to avoid inaccurate estimates of the ability of the hydropower fleet to provide flexible operation (Voisin et al., 2020)

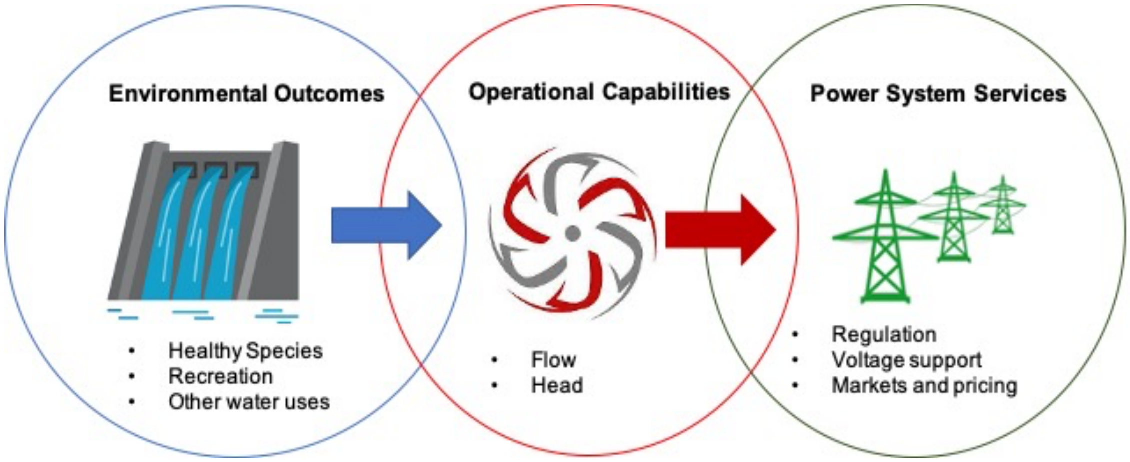


Figure 2: The three major domains in hydropower operation

1.3 Relevance of the research

1.3.1 Role of Hydropower in future power systems

Hydropower is the world's most significant renewable energy source, accounting for 17% of total electricity production (Moran, Lopez, Moore, Müller, & Hyndman, 2018).Hydropower contributes to the decarbonization of the power grid in two ways: first, it generates clean, renewable electricity; second, it acts as a grid stabilizer and enabler, allowing for higher penetration of variable renewable energy sources by helping to stabilize demand and supply fluctuations. Humanity's need for clean and affordable energy, as well as the scarcity, variability, and unpredictability of water resources, will become more pressing concerns in the coming years.

The electricity in a hydropower facility is produced by the movement of water. Rain and melted snowfall from the hills and mountains form streams and rivers that finally flow into the sea. A conventional hydro plant is thus composed of three parts: a river or reservoir that supplies the water, a dam or canal that controls water flow, and a power plant that generates. As a result, hydropower is a complex system made up of water and power systems with distinct views and goals. The bulk of hydroelectric plants in the United States are governed by complex agreements that were created to accommodate a variety of social objectives and to function within specific operational constraints.

1.3.2 Water system perspective of hydropower operation

Water systems are vast networks that include a wide range of participants, including dams and reservoirs, river basins, animals, and downstream agricultural users. Most water system activities are governed by the following categories of water use: water supply, flood control, navigation, water quality, recreation, fish and wildlife, and hydropower. Planning and managing a river basin thus consider a wide range of factors, including economic development, environmental protection concerns as well as water-related issues. The production and storage of a specific hydroelectric project are frequently dictated by regulations and agreements on water use and as a result, the water regime in which a hydroelectric plant is located has a significant impact on the constraints in its operation. The operational policies of a reservoir are significantly influenced by a large number of different public agencies, project beneficiaries, and interest groups in addition to the organizations that own and operate the reservoir system. The objectives of each project decide which entities are in charge of planning and managing reservoir projects within this complicated institutional framework (Wurbs, 2005). Therefore, it is crucial to plan for and manage water resources over the long term. A water system operator's role is to use a hydroelectric facility to operate these complex systems in a way that meets the multiple objectives of water resources. Water managers have no control over the volume of the incoming water, which is determined mainly by the weather and geography. Stream inflow is typically underestimated when modeling river basins. Groundwater flow models are frequently incorrect due to the inability of current monitoring systems to accurately monitor groundwater flow. Furthermore, evaporation from reservoirs cannot be measured directly. For modeling purposes, stochastically varying inputs are required to analyze the uncertainties associated with the water entering and leaving the reservoir (Stoll et al., 2017). Therefore, the many goals of water system management and operation, as well as the numerous limitations and regulations that govern these operations, are exceedingly complex and often ill-defined.

1.3.3 Power system perspective of hydropower operation

To maintain frequency stability in electric power systems, the consumption and production of electricity must always be in balance. The system operator must be able to balance supply and demand for electricity at all times in order to provide a dependable electrical system. Demand exceeding supply will cause the system frequency of the electrical grid to drop below 60 Hertz. If the system frequency drifts slightly from 60 Hz the spinning generators will naturally apply greater force to one another to restore the frequency back to 60 Hz. If the deviation is really large, the grid will collapse on its own. An imbalance between supply and demand also causes voltage instability which occurs when the reactive power provided by the power system is insufficient to fulfill demand. Therefore, it is crucial for the power system to have flexible resources to ensure that users can get electricity when they need it. Along with being a source of cheap, abundant renewable energy on a bulk scale, hydropower also provides large-scale flexibility to the power grid.

The nature of the system and the market under which they operate has a significant impact on the contribution of hydropower in the power system. The objective of power system operators is to provide reliable electricity supply at the lowest possible cost. Because demand varies over time (from seconds to decades), the resource mix has evolved to the point where different types of resources supply the power system with different types of services and energy. The role of power system operators is to select the various services provided by generation resources as efficiently as possible. Generation from hydropower has a significant contribution to the national electric grid by providing essential generation and ancillary grid services, such as energy for baseload and peak load, load following, black start, reactive power control, spinning, and non-spinning reserves, regulation, and frequency response (Hirth, 2016). The widespread deployment of variable renewable energy sources has pushed the need to provide ancillary services to manage the increased variability and uncertainty of the power grid. Consequently, hydropower facilities that were operated consistently in the past were called upon to provide these services owing to their capability to meet the immediate demands of the power system. Hydropower operators often must adhere to a range of operational and environmental constraints in order to maximize revenues from grid services. As a result, power systems, like water systems, present a challenging system operating problem for hydropower operations.

In regions with vertically integrated utilities, a single company responsible for the generation, transmission, and distribution of electricity to their consumers. Power system operators in such regions attempt to schedule generators in such a way that system load is met reliably while costs are reduced and then passed on to end-users. This is challenging because the system operator must forecast both short- term and long-term electric power demands, as well as estimate generation, transmission, and operating costs. In a deregulated electricity market, the utilities that cater to retail customers are only accountable for distribution of electricity to the consumers; the electricity is produced by other entities. Through competitive power markets such as Independent System Operators (ISO) and Regional Transmission Organizations (RTO), these organizations sell the electricity generated. It is difficult to generalize about hydropower generation and subsequent involvement in the power system now and in the future due to the diversity of operational and market organization structures.

1.3.4 Using data to validate representation

Power and water systems are vital in hydropower, responsible for supplying a wide range of services, many of which are interconnected. Hydropower facilities, unlike other generating sources, are planned and run to serve multiple objectives; water-related objectives are often given higher priority in hydropower generation operating policies than power-related purposes. Furthermore, it is clear from the preceding discussion that there is a substantial difference in the representation of hydropower in water and energy systems. While the primary objective of a water system domain is to ensure and maintain a healthy river system, the quantity (maximum energy produced) and the reliability of the electricity generated by the facility are more critical from a power system domain. The same could be said about models employed in these two domains. For hydropower modeling, a variety of water and power system models are currently available, and the model used depends on the desired output.

- Watershed models simulate the availability of water availability and environmental impacts and the operational decisions of a hydropower plant.
- Dispatch models and production cost models are primarily concerned with representing the operational capabilities and the power system constraints.

Modeling tools in the hydropower sector are thus extremely diverse, and while the existing models are clear about the questions they can help address, there appears to be a lack of clarity about which model is best for answering facility-specific questions, which can lead to incompatibility in decision-making.

Data-driven modeling is based on examining the data that characterizes the system under consideration. With only a few assumptions about the physical behavior of the system, a model is constructed based on the relationship between the different state variables (input, controls, and output) of the system. Given the complexities and multi-objective operations of the water and power systems, using data to highlight the bi-directional transformation between the two systems can aid in conveying the functionality and value of hydropower from the context of the facility.

1.4 Research Objectives

As previously stated, this research aims to construct a transfer function model that characterizes hydropower operation. The research objectives are to

1. Explore the relationship between the time series of inflow to the reservoir and time series of downstream flow
2. Develop a Box Jenkins transfer function model based on the relationship identified
3. Examine the methodology's ability to distinguish among various types of facilities (e.g., storage versus Run-Of-River)

1.5 Thesis Outline

This thesis is organized into six chapters.

- Chapter 1 lays the background along with the definition of the problem.

- Chapter 2 summarizes the literature review on transfer function modeling in non-hydro power-related sectors and the methodology characteristics of existing energy system models. Various research gaps are also identified.
- Chapter 3 describes the methodology which includes various steps involved in Box Jenkins models
- Chapter 4 outlines the many procedures necessary to preprocess the fleet data for applying Box Jenkins methodology and the subsequent analysis
- Chapter 5 provides results of transfer function modeling explores potential explanations of the model outputs obtained.
- The conclusions and recommendations for future work are summarized in Chapter 6

Chapter 2: Literature Review

2.1 Background

Historically, efforts at statistically relating the system's input to its output started with regression analysis. However, a regression model only considers the simultaneous response between the input and output variables. Additionally, regression analysis would only be successful if the system is in stable equilibrium and is inappropriate in circumstances where there is a time-lagged relationship and noise in the system(Pankratz, 2012).

Transfer functions demonstrate the causal relationship between the input and output of a process. In 1976, George Box and Gwilym Jenkins introduced a statistical method to model the relationship between input and output of a system by using transfer functions. The Box-Jenkins transfer function methodology presents a set of procedures for identifying, fitting, and checking autoregressive integrates moving average (ARIMA) models with the time series data (G. E. P. Box, Jenkins, Reinsel, & Ljung, 2015). This chapter presents the results of the literature review, which includes:

a) representation of hydropower systems in existing models and the significance of transfer function modeling.
b) A summary of different statistical techniques and description of the Box-Jenkins method
c) Application of Box Jenkins methodology in non-hydropower-related research

2.2 Hydropower modeling and fidelity

Decision-making in a hydropower facility falls into three domains: Environmental outcomes, operational capabilities, and power system services.

- <u>Environmental Outcomes</u>: Despite the advantages of hydropower as a relatively clean fuel, the development of hydropower facilities have been linked to severe and irreversible alterations in the natural hydrologic river regimes affecting the quality of habitat and fish species. Hydropower generation has negative impacts on water quality, habitat, landscape, and biodiversity. Consequently, the interaction of hydropower facilities with upstream and downstream results in significant physical, chemical, and biological transformation of the local ecosystem.

- Operational capabilities: Water inflow into the reservoir significantly impacts the operation of the hydropower facility and the reservoir storage. The flow into the reservoir determines the operational bounds of hydropower generation and creates restrictions on the energy and ancillary services provided by the facility.
- Power System Services: Hydropower facilities can participate in both the power system's energy and ancillary services markets. The power system services offered by hydropower include voltage support, regulation and frequency response, load following, spinning, and non-spinning reserve.

Decision-making, therefore, involves multiple stakeholders with conflicting perspectives, values, and proposed solutions. Existing hydropower representation in energy system models mainly falls into three categories: watershed, dispatch, and production cost.

2.2.1 Watershed Models

Watershed models focus on the water systems and aim to evaluate the impacts of different operational regimes on reservoir storage and releases. Watershed management models concentrate more on Best Management Practices for water uses, and the most used models are:

- **RiverWare**: RiverWare is a river and reservoir modeling tool developed by CADSWES (Center for Advanced Decision Support for Water and Environmental Systems) at the University of Colorado Boulder with a wide range of applications, including operational scheduling and forecasting, policy evaluation, planning, and other decision processes. The tool models the entire water system, including the reservoir and associated environmental outcomes, and it is thus used by many agencies, including the Tennessee Valley Authority and the U.S. Bureau of Reclamation, and the U.S. Army Corps of Engineers(Cotter, Hydraulic Engineer, District, & Zagona).

  The RiverWare model can be run in three modes: pure simulation, rule-based simulation, and optimization. In pure simulation mode, variables like reservoir storage, pool elevation, and turbine discharge are used to begin the simulation. This mode solves a problem, which is completely specified, and the object-oriented approach makes it easier to identify whether the model may be over-or underdetermined. In rule-based simulation mode, multiple unknown values are allowed to be inputs, and additional information is provided by prioritized rules which are user-specified. These "if-then-else" operating policy statements examine the system's state and then drive the simulation by setting slot values on the variables depending on that state. The optimization mode works through a linear programming approach for prioritized policy objects and constraints. The reservoir outflow is optimized for a prioritized set of user-specified objectives such as navigation, water supply, hydropower production, recreation and flood control, and fish and wildlife habitat ("RiverWare,").

- **MODSIM**: MODSIM is a river basin management decision support system (DSS) developed by the Colorado State University. It utilizes a network flow optimization algorithm to simulate a priority-based water allocation mechanism in a river system. The most recent version of the tool is developed under the Microsoft .NET Framework and provides the users with the ability to customize it for any specific input, operating regime, and output. MODSIM is based on the hypothesis that any complex river basin can be represented in a network formulation comprised of nodes and links connecting the nodes. Therefore, in addition to simulation of reservoir allocation and operations, MODSIM could also perform complex water rights accounting without writing scripts or rules (Labadie, 2006).

2.2.2 Dispatch Models

Dispatch models are used to optimize the revenue in a power plant and are typically utilized for short-term applications (up to 14 days). As the generation must match the load, a set of network constraints in addition to security and stability constraints needs to be accommodated to ensure the safe operation of the system.

- **SHOP** (Short-term Hydro Operation Planning): Developed by SINTEF, a research organization in Norway, in collaboration with Norwegian University of Science and Technology (NTNU), SHOP is a hydropower scheduling tool to maximize profit. Components of SHOP include reservoirs, hydropower units, discharge gates, and junctions. Successive Linear Programming and Mixed Integer Programming are utilized in the software, and the market process and inflows are assumed for the entire horizon. Unit commitment and dispatch plans could be determined, and depending on the planning task prepared, SHOP models can be run in different modes. Examples of operational constraints included in the software consist of time-dependent ones such as minimum and maximum production values, reservoirs, and gates, and the results are given as times series(Skjelbred, 2020).

- **GTMax:** GTMax (Generation and Transmission Maximization Model) was created by Argonne National Laboratory in 1995. This model utilizes a network representation of the power system, which is constructed from objects representing demand, supply, and transport systems. Data is entered at various time periods, including annual, monthly, weekly, daily, and hourly. The power system operations and energy transactions are optimized and solved using linear and mixed-integer programming. Hydropower units are one of the six power supply resources in the GTMax model. A hydro node consists of three options; Run-Of-River, storage, and pumped storage. Hydropower dispatch is constrained by reservoir-specific limitations, and GTMax computes the marginal value of water by considering those operational restrictions(T. D. Veselka, 2009).

2.2.3 Production Cost Models

The main objective of production cost models is minimizing the production costs while adhering to the operating constraints. These models calculate hourly production costs and market clearing prices which are location specific.

- **PLEXOS**: Developed and commercialized by Energy Exemplar, PLEXOS models unit commitment and dispatch of generators in the power system. The model uses a deterministic mixed-integer linear program to minimize the overall cost of operation. Depending on the data available, the modeler can choose between three hydro model settings: Energy, Level, and Volume. The software can assume either a fixed or economic dispatch for hydropower generation. For fixed dispatch, the software read in the file specifying the electricity generated for every hour of every day for the entire year. While in an economic dispatch, hydropower is dispatched when it is most beneficial for the system operation while accommodating operation constraints(Bain & Acker, 2018).
- **PROMOD**: PROMOD is a production cost model developed and marketed by Ventyx. It provides an extensive representation of the topology of the power system and is used for a variety of applications, including locational marginal price (LMP), asset valuations, financial transmission right (FTR) validation, and forecasting. In PROMOD, the hourly electricity generation is optimized based on the type of energy source (thermoelectric,

hydropower, solar, wind) and asset characteristics (capacity, cost, contract types, etc.) to satisfy the hourly loads in each zone for the lowest cost. In this model, hydro units and scheduled before the thermoelectric units, and for energy scheduling, hydropower units are defined as either run of river or peak shave. In the Run Of River option, the units are scheduled to uniformly transmit the provided energy limit, while in the peak shave option, the units are allocated to meet the upper-most load(Nekooie, 2018).

- **GridView**: GridView is an analytical tool developed by Hitachi ABB Power Grids Inc. for market simulation and asset performance evaluation. Given the unit characteristics and chronological load, the software dispatches generators to minimize production costs. Under normal as well as contingency conditions, GridView performs dispatch to ensure that the transmission line restrictions are not exceeded. Additionally, the shadow prices on lines and spot rates on buses are also estimated. If using the load following schedule option, the software adjusts each hydro generator's weekly schedule using the weekly energy budget, minimum and maximum generation according to the weekly k factor at the start of each week in simulation. The k factor is the value that characterizes the plant's ability to respond to the load by combining hydraulic and environmental constraints into a single number (Nathalie Voisin, 2021)

2.3 Hydropower representation: RiverWare vs. PLEXOS

2.3.1 Representation of hydropower in RiverWare

RiverWare is a watershed modeling tool developed as a collaborative effort by the TVA, the U.S Bureau of Reclamation (USBR) and the University of Colorado Center for Advanced Decision Support for Water and Environmental Systems (CADSWES). The features of the river basin are represented by *objects* which are represented by icons on the graphical workspace. The object in turn has different *slots* which corresponds to the data structure for a variable or parameter used in the physical process equations for that feature. The required data are entered through direct manual entry or through importing database. The objects also contain various *methods* to model the different processes. There are two method types, namely dispatch and user selectable. In dispatch method, the user specifies the input/output configuration to solve the process using conventional algorithms while in user selectable methods, the basin is modeled in accordance with the algorithm/model which the user selects(Zagona & Magee, 1999). The main objects associated with hydropower modeling are listed in Table 2. Although inline power is shown as the object that represents run of river production, it cannot be generalized. Strictly, RoR schema means that the river is not dammed and thus do not have any water storage capability. However, some run-of-river facilities do use a small weir or dam to make sure that adequate water reaches the penstock and have a little pondage to store water for immediate use. As they cannot store water for future use, these facilities cannot be categorized as storage plants as well. Therefore, in RiverWare such facilities are represented as Slope Power Reservoirs. Contrary to the general definition of a storage reservoir, RiverWare's version of the object doesn't have any power-generating capability and the only process performed is the storage of water. Level Power Reservoir represents the object in which water is stored behind a reservoir and utilized for energy production (Singh & Frevert, 2010).

2.3.2 Reservoir modeling in RiverWare

RiverWare offers three different kinds of solution techniques: simple simulation, rule-based simulation, and optimization. In a simple simulation, the user provides the inputs that drive the solution, which is based on an object-oriented modeling paradigm where each object waits to solve until it has enough information. In the other two techniques, operational policies drive the solution. In a rule-based simulation, the user-specified priority policy rules add additional information on the objects to solve the system and then modify the slot values on the objects based on the system state. In optimization, linear programming is set to optimize each of the prioritized goals input by the used. Optimization offers a universal solution across all objects and all the time steps taken, in contrast to the other two procedures that solve each item individually, one time step at a time. This enables the optimization solution to trade off objectives both spatially and over time. Modeling inflow, storage, and outflow in the power reservoir (facilities with power generating capability) objects is accomplished using mass balance approach: The common equations for reservoir mass balance in RiverWare are:

$$Storage_t = Storage_{t-1} + \sum(Inflows \times \Delta t) - \sum(Outflows \times \Delta t) + Gains - Losses$$

$$Outflow = Release(s) + Spill(s)$$

$$Total\ Inflow = Inflow + Hydrologic\ Inflow$$

Table 2: Workspace objects in RiverWare related to hydropower

Objects	Functions
Reach	A section of the river that routes water using the many user-selectable routing algorithms.
Inline power	A hydropower plant on reach with no storage and simulates run-of-river production.
Level Power Reservoir	Reservoir with hydropower plant and outlets. The power and energy are computed via user-selected methods and solves mass balance equation.
Slope Power Reservoir	Similar to level power reservoir but with the capability to model the backwater storage effects of a sloped water surface.
Pumped Storage	Reservoir which can be linked to an in-line reservoir with the capability to store energy as well as generate power.
Storage Reservoir	A reservoir with outlets and spillways but with no hydropower facilities.

Hydrological inflows are inflows into a reservoir that are not a part of the main stream and/or ungauged (Zagona & Magee, 1999). RiverWare provides the capacity to simulate reservoir hydrology and hydrological processes, hydropower generation and energy use, as well as water ownership and rights.

2.3.3 Representation of Hydropower in PLEXOS

The PLEXOS system consist of PLEXOS GUI which includes the input and output interface and the PLEXOS engine. In the input interface, the user enters or imports the energy system data (description of power system, analysis specification) which in turn is read by the PLEXOS Engine. After the data has been read, a solver interprets it to produce results that may be seen in the output interface. The object model that forms the foundation of PLEXOS is based on three levels of hierarchy: objects, memberships, and properties. Entities to be modeled in a system are called objects, while memberships refer to the relationships between the objects. Properties are the characteristics of objects which is used to store the data associated with object. Hydropower systems are modeled in PLEXOS using four main classes: Generator, Waterways, Storages, and constraint. The different classes are briefly described below(Papadopoulos, Johnson, Valdebenito, & Exemplar, 2014):

1. Generator: The generator class includes the properties of the hydro generators such as energy, capacity factor, load, and units.
2. Storage: Reservoirs with any given capacity and short, medium, or long-term storages are represented by the storage class. They can also be used to represent simple river junctions, as well as the head and tail ponds of generators. There are generally three different kinds of storage: pumped storage reservoirs, short-term storages that cycle every few hours, days, or weeks, and long-term storages, whose "water value" is estimated exogenously or decided over an extended period of time.
3. Waterways: The waterway class is used to model the canals and spillways. By assigning a membership, waterways can either join the storages together or let the water ‘spill to the sea”.
4. Constraints: Constraint objects can be used to define the custom constraints define the individual or combination of elements in the hydro system. The different constraints can be set for generators, waterways, and storages in any combination.

2.3.4 Reservoir Modeling in PLEXOS

The hydro model in PLEXOS has three types of settings: energy, level, and volume. In an energy model, storage volumes are expressed in terms of potential energy which is determined by the generating efficiency of all the power plants located downstream of the storage. Simple "linear" cascaded systems and models of closed-circuit pumped storage are ideal applications for this kind of model. The level model uses elevations and reference areas are used and storages are modeled as trapezoidal, indicating that their surface area increases as they fill up. The storage volumes are measures in thousands of cubic meters. In contrast to energy model, generator efficiency must be stated and are measured in $\frac{MW}{m^3/s}$. In the volume model, the unit of storage is a volume of water, instead of levels of potential energy and the storage volume is represented in cubic meter days (CMD). The generating efficiency must be defined and are measured in $\frac{MW}{m^3/s}$, similar to the level model. The hydro dispatch approach employed in PLEXOS maximizes the utilization of hydropower while being constrained by monthly maximum and minimum power outputs as well as monthly energy constraints for the dispatchable units(Exemplar, 2022).

2.3.5 Drawbacks in the representation

1. **Ambiguity in hydropower classification**: The common forms of classification of hydropower facilities are storage, run-of-river, and pumped storage. However, this classification does not apply to either of the models mentioned above. The term "storage reservoir" in RiverWare refers to an object without hydropower capabilities, whereas "inline power" is used to describe facilities that produce power without storage. The run-of-river facilities are mostly depicted as slope power reservoirs while storage facilities are depicted as level power reservoirs in the RiverWare workspace that represents the TVA system's reservoirs.(Biddle, 2001) While in PLEXOS, all the reservoirs come under the Storage Class which is divided in to pumped storage, short term storages which operates under a RoR schema and long term storage for representing traditional modeling facilities.
2. **Modeling Fidelity**: Hydropower facilities generates electricity by discharging water from the reservoir into the penstocks and through turbines. Although the different operational aspects of hydropower operation are well understood and easier to model, the accompanying hydrological parameters and water management choices are complex and

site-specific. Thus, the fidelity of hydropower modeling performed by a dam operator, or a power provider sets it apart from external modeling(Turner & Voisin, 2022). On analyzing hydropower representation in both these models, it clear that they are based on general guidelines for water release that are derived from the prior knowledge of inflows, reservoir elevations, flow and capacity constraints and demands. Both these models however cannot account for the reservoir operation events that are the result of manual decisions made by the reservoir operators, or that fall outside of the standard operating procedures of the reservoirs.

3. **Incorporating non-stationarity:** A key assumption in reservoir design and operation is hydrologic stationarity. Stationarity indicates that recorded observations have a probability distribution function that does not change with time and whose properties can be inferred from the past. Although some hydrological processes are stationary, others may change over time due to variables including changes in regional resource management and hydroclimatic change(Milly et al., 2008). The assumption of stationarity is made in both the models outlined above including in both simulation and optimization, by either directly utilizing historical inflows as an input or by employing synthetic inflow based on historical streamflow statistics.

2.4 Summary

Hydropower development is frequently subjected to environmental and regulatory constraints and representing these constraints inside the existing energy models model is a complex process. The literature research also reveals that there is diversity in the models and their representation of hydropower. Most of the models are focused on either water or power systems, and because both are complex systems with competing objectives and multiple constraints, it has not been possible to capture these intricacies in a single model yet. Both water and power system operators require the right tools to examine the impact of renewable energy integration on water system operations, as well as quantify the flexibility of hydropower plants to address power system concerns. Characterization of hydropower is thus a major weakness of energy systems models because:

a) Because the majority of models are concerned with linear programming optimization, non-linear dynamic aspects of a hydropower plant, such as evaporation losses and hydraulic head effects, must be simplified in order to be incorporated into such models.

b) Reservoirs and pumped storage systems are being employed to provide energy storage and ancillary services to the power grid, and standard hydropower modeling frameworks may not adequately capture the flexible properties of hydropower generation.

2.5 Significance of the Study

2.5.1 Characterization of hydropower – need for a common rubric/nomenclature

The majority of the models available (watershed, dispatch, and production cost) are utilized in a variety of applications and are adopted as a standard by many utilities. However, there is a need for improvement in how they represent hydropower. Because water has its own constraints as a fuel source, it is challenging to represent hydropower the available models. Additionally, as observed from previous discussions, both water models and power system models place distinct constraints and values on hydropower and many potential opportunities exist for better representation. In models and communications that support electric power and water management decision-making, it is challenging to clearly identify, classify, and express the key consequences of hydropower facility operations within the scheduling of power systems and water systems. While practitioners in the water and power domains have an intuitive understanding of the multiple assets they analyze and schedule on a daily basis, only a handful of them have gained relevant insights. Fewer still have collected data for entire fleets of hydropower assets spread across multiple owners, watersheds, balancing authorities, and interconnections. What is required is a quantitative, data-driven rubric and nomenclature to compare and contrast the essential functionality of numerous diverse hydropower facilities that translate the dynamics of water and electric power across various operational and planning time scales.

2.5.2 Modeling water and energy systems

In the past, various efforts were made to accurately model water and power systems to achieve proper hydrological and power systems independent of one another. They usually fall into two categories: merging existing models or constructing a new model for a specific research region. Linking energy systems models with watershed models can provide additional information for both these systems, allowing their individual capabilities to assist each other. Large-scale renewable energy integration studies have simulated power systems at varying penetration levels, although frequently using simplified representations of hydropower operations. Combined optimization of both electrical and water system models has proven to be reliable in this scenario(Ibanez et al., 2014). The adequacy of integrating a hydrological model with a production cost model (PROMOD) to estimate the susceptibility of the US western electric grid to climatic conditions was investigated in another study (Voisin et al., 2016). Cardenal et al. devised a methodology for introducing power markets into hydro-economic models to analyze the economic tradeoffs between hydropower, and other water uses in the Iberian Peninsula(Pereira-Cardenal, Mo, Riegels, Arnbjerg-Nielsen, & Bauer-Gottwein, 2015).

Due to the high computational cost, a model that optimizes water resource management and power system planning simultaneously has yet to be adopted. The Science and Technology Directorate of the United States Department of Homeland Security (Petri, 2009) stated a "serious unmet need" in comprehending the interdependence of electrical and water system infrastructure to aid in the recovery of the national power grid after regional or national-scale incidents. In March 2019, the Pacific Northwest National Laboratory (PNNL) and the National Renewable Energy Laboratory (NREL) conducted a workshop supported by the U.S. Department of Energy (DOE) to better understand the need for research to improve hydropower representation in grid models(Voisin et al., 2020). The current research looks to narrow the knowledge gap in several of the research themes outlined in the workshop report released in November 2020. In particular, it will particularly assist with theme two, "Data Availability as a Barrier to Modeling," which examines how a lack of publicly available hydropower-specific data impedes various hydropower modeling tasks from a power system perspective.

2.5.3 Relevance of time series analysis

Time series modeling has been explored in the machine learning and statistics communities for decades. In general, there are two types of applications for time analysis: creating predictions of future values and learning representation, which entails determining the nature of the phenomenon represented by the observations. For decades, the former has been a prominent research topic, and time-series models are used for forecasting if three requirements are met (Makridakis, Wheelwright, & Hyndman, 2008) :

- Historical recordings of the data are available
- Required information is quantified as numerical/categorical data

- It can be assumed that at least some portion of the past pattern will be repeated in the future (assumption of continuity)

Many components of the hydrologic cycle are described using time series; common ones include precipitation, flowrate, discharge levels, and streamflow(Survey, 2016). Hydrological time series data has been widely used for forecasting studies in a variety of domains, including hydropower generation, drought mitigation, and water resource management (Adamowski, 2008; Alemu, Palmer, Polebitski, & Meaker, 2011; Pozzi et al., 2013; Waage, Baldwin, Steger, & Bray, 2001). However, employing time series analysis for diagnostic learning to determine system behavior based on historical data has not been explored adequately in the hydropower sector. This is significant both in terms of renewable energy integration and climate change.

2.5.3.1 Data-Driven Modeling

Based on the review of existing models, they can be classified as simulation, optimization, or combinations. A simulation model represents a system that is used to forecast its behavior under a given set of conditions. In contrast, optimization models consist of objectives, variables, and constraints used to generate an "optimum" result. Both approaches rely on prior knowledge of the system in question and are usually based on the first-order principles from the physics of a phenomenon or system.

Data-driven modeling is based on examining the data that characterizes the system in question and relies on the measurements taken from real-world systems. The most common methods used are statistical methods, artificial neural networks, and fuzzy rule-based systems(Solomatine & Ostfeld, 2008). With only a few assumptions about the system's physical behavior, a model can be evaluated by analyzing concurrent input and output time series.

2.6 Statistical Modeling Techniques

Statistical modeling is the process of applying statistical analysis to a set of data to find the mathematical relationship between the variables and draw inferences about its characteristics. Using statistical modeling to examine raw data allows scientists to adopt a strategic approach to data analysis by creating representations that uncover the relationship between variables and make informed decisions. Statistical modeling methodologies falls into two categories: supervised and unsupervised learning (Sathya & Abraham, 2013).

2.6.1 Supervised Learning

Supervised learning is a method of training a model on a labeled historical dataset such that it can predict the outcome. Supervised learning can be further classified into two types: classification models and regression models.

- Classification Models: In classification models, the learning algorithm learns a function to translate inputs to output where the output value is in discrete class label. The test data is assigned to specific groups by using this process. Common types of classification algorithms include Random forests, Naïve Bayes, K-Nearest Neighbors, Decision Trees, and SVM.
- Regression Models: In regression models, the algorithm is used to identify the relationship between the dependent and independent variables and the output is continuous real number. Linear regression, logistic regression, and polynomial regression models are all common types of regression models.

2.6.2 Unsupervised Learning

Unsupervised learning is a technique to build models from unlabeled data without human intervention. These models are used for the following tasks:

- Clustering: In clustering the data is grouped according to their similarities (or differences). An example is K-means clustering, where K represents the size and granularity with which the algorithm groups data.
- Dimensionality reduction: When the features involved in the dataset is too high, this technique is applied. Dimensionality reduction attempts to maintain data integrity while lowering the number data inputs to a manageable level.
- Association: In this unsupervised learning method, different rules are applied to determine the association (dependency, relationship) between group of objects in a large data set.

2.6.3 Selecting the required model

Selection of the appropriate model generally depends on the questions posed, time restraints and data available (Love, 2002). This research tries to address the following questions:

1. What changes are required in the existing models to effectively address the complexities of water and power management decisions?

2. When assessing the functionality of a hydropower facility, what influence does water dynamics have on electric power dynamics, and vice versa?
3. What specific information pertaining to the above questions could be harnessed from hydropower fleet data?

Unsupervised learning methods are given little consideration as data from the hydropower fleet is labeled and is in the form of historical time series. Furthermore, the appropriate time to apply unsupervised learning is when there is no available data on desired outcomes. However, because we are aiming to better understand the relationship between two systems, it may not be practical to apply such techniques in this study. As fleet data is in continuous time series format, regression models are believed to the best choice among the types of models available in supervised learning.

2.6.4 Objectives of regression models

Regression analysis is one of the most widely used statistical approaches in practice, with applications in a variety of scientific domains such as economics, engineering, biology, agriculture, medicine, geology, and others. Regression models are used to define the relationship between a dependent (or response) variable y and the independent (predictor) variables x_1, x_1, $x_3 \dots . x_n$. Regression analysis has several objectives (Yan & Su, 2009):

a) Based on a set of values of x_1, x_1, $x_3 \dots . x_n$, predict the value of y
b) Establish a causal relationship between the response variable and the predictor variables
c) Examining the independent variables x_1, x_1, $x_3 \dots . x_n$ to see which one is more essential than the others in explaining the dependent variable y and determining the relationship appropriately.

The different regression models include linear regression, logistic regression, and polynomial regression, with linear regression being the most fundamental and extensively employed.

2.6.5 Linear regression models

A simple linear regression model is used to define the relationship between a dependent (or response) variable y and the independent (predictor) variable x. The linear regression model is written as the following form (Pankratz, 2012):

$$y_t = \beta_0 + \beta_1 x_t + n_t$$

β_0 = y intercept, β_1 = gradient/slope of the regression line, n_t= stochastic disturbance term

In simple linear regression, it is assumed that the disturbance n_t is normally distributed with $E(n_t) = 0$ and a constant variance $Var(n_t) = \sigma^2$.

2.6.5.1 Drawbacks in estimating with linear regression

a) Relationships with time lag: If y_t is related to x_t with a time lag, there are chances that it may be related to previous lag terms including x_{t-1}, x_{t-2} .. and so forth. If we adhere to solely linear regression in that scenario, the impacts of prior time lags may be missed and as a result, the estimate of the random error term will be higher than necessary.

b) Self-correlation of the noise series: The disturbance term n_t ma be related to its own past values n_{t-1}, n_{t-2} and if we ignore this in estimating the regression model, the predicted residuals will be much larger than they need to be. In addition, failing to account for autocorrelation in the disturbance results in inefficient coefficient estimates and erroneous statistical tests(Pankratz, 2012).

2.6.6 Dynamic Regression models

Alan Pankratz introduced the term "Dynamic Regression" to improve the shortcomings of simple linear regression models in his book Forecasting using Dynamic Regression Models. A dynamic regression model states how the output y_t is linearly related to the current and past values of one or more inputs(Pankratz, 2012). It is crucial to note that the inputs are not affected by the output and the noise term n_t has an autocorrelation structure in this model. The dynamic regression model is written as the following form:

$$y_t = C + \frac{\omega(B)}{\delta(B)} x_{t-b} + +n_t$$

b = delay terms

$\omega(B)$ = describes the magnitude of immediate effects of the x_t on y_t

$\delta(B)$= describes the duration and pattern of the decay

Pankratz introduced dynamic regression in 1991 to describe what Box and Jenkins termed as "transfer function models" in 1976.

2.7 Box-Jenkins ARIMA Modeling

Some common terminologies in time series analysis are summarized and the elements of Box-Jenkins ARIMA methodology described in this section.

2.7.1 Stationarity

A times series is defined as observations recorded sequentially in time. By convention, time series can be classified as discrete or continuous. In a discrete series, the observations are taken at discrete points, while for the latter, the measurements are made continuously through time.

A stationary time series is one whose statistical properties (mean, variance, autocorrelation) remain constant over time. Stationarity plays a relevant role in enhancing our ability to analyze a time series and its various application. A stationary model assumes that the stochastic process remains in *statistical equilibrium* as the probabilistic laws that govern the behavior of the process remain constant over time. (Hyndman & Athanasopoulos, 2018)

2.7.2 Differencing

Differencing is a method used to transform a non-stationary time series to achieve stationarity and helps to stabilize the mean by removing existing trends in a time series.

First-order difference: First-order difference or first differencing is the change between consecutive observations in a time series. Let $\mathrm{Y_t}$ denote the value of time series Y at a given time t, then first differencing is calculated by:

$$Y_t{}' = Y_t - Y_{t-1}$$

Random walk model: If the first differencing results are entirely random (where the current observation is the same as the prior one with a random step up or down), then the time series Y is said to follow a random walk. The random walk can be represented as:

$$Y_t = Y_{t-1} + \epsilon_t$$

Where ϵ_t is the random shock, the value of error term at time t.

Second-order difference: If the first differencing does not produce stationarity, it may be necessary to produce a second-order difference.

$$Y_t{}'' = Y_t{}' - Y_{t-1}{}'$$

Seasonal difference: Seasonality in a time series is a characteristic in which the data observe a regular pattern of changes that repeats over a specific time frame (weekly, daily, quarterly, etc.). The seasonal difference is thus the difference between an observation and the previous observation from the same season, rather than in consecutive periods. If a given series has *m* seasons, seasonal differencing is calculated by: (Hyndman & Athanasopoulos, 2018)

$$Y_t{}' = Y_t - Y_{t-m}$$

2.7.3 Correlation

In statistics, correlation measures the local strength and direction of a linear relationship between two random variables. Analyzing the correlation between a series and its lags (shifts in time) is relevant in time series analysis as past lags may contain patterns or properties that might influence subsequent periods' values.

Correlation between two random variables x and y is expressed as:

$$r_{x,y} = \frac{cov(x,y)}{\sigma_x \times \sigma_y}\ , where -1 \leq r_{x,y} \leq 1$$

$cov(x,y)$ = Covariance between x and y

$= E[(x - E(x)) - (y - E(y))]$

σ_x = Standard deviation of x

σ_y = Standard deviation of y

Autocorrelation function: In time series, the autocorrelation function (ACF) is used to quantify the correlation between two adjacent values (Krispin, 2019). For example, for a lag k, the sample autocorrelation function is:

$$r_k = \sum_{i=1}^{n-k} \frac{((y_i - \bar{y}))((y_{i+k} - \bar{y}))}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

n = Number of observations

$\bar{y}$= Mean of the series

Partial Autocorrelation Function: The partial autocorrelation function (PACF) at lag k is the autocorrelation between y_t and y_{t-k} which are not considered for by lags 1 through k-1. Thus, the PACF directly correlates a series at lag k after removing the interventions of the shorter time lags $(y_{t-1}, y_{t-2}, \dots, y_{t-k-1})$.

2.7.4 Stationary Models

1. Autoregressive Model: In an Autoregressive (AR) model, the current value of the seriesY_t can be predicted using previous outputs $Y_{t-1}, Y_{t-2}, Y_{t-3}$ and a random shock value

a_t. The term autoregression indicates the regression of a variable on itself. The order of the model is the immediately preceding values in the series that are used to predict the current observation. Mathematically, an AR model of order p can be expressed as (Burges, 1998)

$$Y_t(p) = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + a_t$$
$$= c + \sum_{i=1}^{p} \Phi_i Y_{t-i} + a_t$$

a_t = random error at time t with mean 0 and constant variance, independent of Y_t

c = constant

Φ_p= coefficients of the autoregressive process

2. Moving Average Model: In a moving average (MA) model, the past forecast errors are used as the explanatory variables while predicting the current observation. A moving average model of order q is represented as:

$$Y_t(q) = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$
$$= \mu - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t$$

a_t = random error at time t with mean 0 and constant variance, independent of Y_t

μ = mean of the time series

θ_q= coefficients of the moving average process

3. Mixed Models: Mixed models attempts to capture the properties of both Autoregressive and Moving Average processes. These models usually have fewer parameters than an AR (p) or MA (p) model by themselves. There are two common types of mixed models: Autoregressive Moving Average (ARMA) model and Autoregressive Integrated Moving Average (ARIMA) process.

An ARMA (p, q) model is defined by the equation:

$$Y_t(p, q) = \sum_{i=1}^{p} \Phi_i Y_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{j-i}$$

The ARMA model assumes that the time series is stationary, however, in practice, many datasets have trends and seasonality. When the time series exhibit-non stationarity, an initial differencing is included in the model to achieve stationarity. Combining differencing of an ARMA process produces Autoregressive Integrated Moving Average (ARIMA) models.(Hyndman & Athanasopoulos, 2018)

An ARIMA (p, d, q) model is represented as:

$$Y_t{}'(p, d, q) = c + \sum_{i=1}^{p} \Phi_i Y'{}_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t$$

$Y_t{}'$= differenced time series (which may have been differenced more than once)

d = degree of differencing involved.

2.7.5 Backshift Notation

ARIMA models are often represented in backshift notation B and is useful when working with time series lags. Using notation B before a series indicates that element should be moved back one time,

$$B^k Y_t = Y_{t-k}$$

2.7.6 Differencing Operator

The backshift operator is also convenient for describing differencing process. As explained above, first order differencing is represented as:

$$\begin{aligned} Y_t{}' &= Y_t - Y_{t-1} \\ &= Y_t - BY_t \\ &= (1 - B)Y_t \\ &= \nabla Y_t \end{aligned}$$

The differencing operator ∇ takes the difference between an observation and the previous observation in a time series. Seasonal differencing operator ∇_mtakes the difference between two points in the same season. These are expressed as:

$$\nabla Y_t = Y_t - Y_{t-1}$$

$$\nabla_m Y_t = Y_t - Y_{t-m}$$

Using these definitions, the general ARIMA (p, d, q) thus can be represented as (Pankratz, 2012):

$$\Phi(\mathrm{B})\nabla^d Y_t = c + \theta(\mathrm{B})a_t$$

$$\nabla^d = (1-B)^d \qquad \text{(the d} - \text{order differencing operator)}$$

$$\Phi(\mathrm{B}) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p \quad \text{(the p} - \text{order AR operator)}$$

$$\theta(\mathrm{B}) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \quad \text{(the q} - \text{order MA operator)}$$

2.8 The Box-Jenkins Methodology

George Box and Gwilym Jenkins developed a systematic methodology for selecting an appropriate model for estimating and forecasting a univariate time series (G. E. Box & Jenkins, 1976). It is a three-step iterative strategy for identifying, estimating, and forecasting ARIMA models. This approach relies on the principle of parsimony which refers to modeling a time series with as few parameters possible for adequately representing the process. Parsimonious models, according to Box and Jenkins, yield better forecasts than models with numerous parameters(G. E. P. Box et al., 2015). The following are the three steps in the Box-Jenkins methodology:

1. **Model Identification:** In the identification stage, values of d and then p and q in the ARIMA (p, d, q) model is selected by visually examining the ACF and PACF plots of the time series. Because the Box-Jenkins approach can only be applied to stationary series, the first step in the identification process is to assess if the provided time series is stationary. Plotting each observation in a sequence against time yields helpful information about outliers, missing values, and periodic trends. While differencing is the common transformation technique used to achieve stationarity, logarithmic transformations can also be used to stabilize the variance of a time series. However, they can only be applied to positive-valued series.

   <u>Unit root tests</u>

   A unit root process is a stochastic trend in a time series, and the presence or absence of unit roots can assist in identifying some of the underlying properties of a time series. If a series does not have any unit-roots, it is better characterized as stationary with a constant mean.

   Unit root tests are statistical hypothesis tests that can be used to determine the stationarity of a time series when visual detection fails to detect trends and/or seasonal components. The two common unit root tests used in time series analysis are the Augmented Dickey-Fuller (ADF) Test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test. After achieving stationarity and determining the value of d, the next step is to identify the orders p and q of the AR and MA process. A comparison of the sample ACF and PACF plots may reveal patterns that provide additional information on the data and its characteristics. Box and Jenkins (G. E. P. Box et al., 2015) summarized the behavior of ACF and PACF in three types of stationary ARIMA models, which are listed in Table 3.

2. **Model Estimation:** Following the estimation of the orders of $p, d,$ and q, the next step is to obtain precise estimates of the coefficients ARIMA model (Φ_p, θ_q) chosen during the identification stage. For this, an efficient nonlinear least-squares algorithm is used, the common one being the Maximum Likelihood Estimation (MLE). The residuals, which are the difference between the observed and "fitted" values of the time series, are also obtained. The sum of squared residuals is minimized with the least-squared estimates of Φ_p, θ_q. In addition to the above, parameter estimates should meet certain conditions for the suggested model to be considered acceptable. The generated AR and MA parameters must be statistically significant, and this is verified by T-tests for the estimates parameters must produce a fitted time series that is stationary (for AR) and invertible (for MA). Finally, it is critical to ensure that the parameter estimates are not overly correlated, as strong correlations suggest low-quality estimations. In most circumstances, a 0.9 correlation is used as a rule-of-thumb cutoff level. If the correlation coefficient calculated in greater than 0.9, then the parameters estimated would have been dependent on each other (Pankratz, 2009).
3. **Diagnostic Checking:** The diagnostic checking stage verifies that the residuals calculated during the estimation stage constitute a white noise process, meaning that no more improvement in the residual variance can be gained by adding another parameter. To do so, the ACF and PACF plots of the residuals are inspected to detect any unaccounted pattern. At any lag order, the ACF plots must show no significant correlation, and the PACF plots must show no significant spikes (Fig.3). The Ljung-Box Chi-Square test is another statistical test used to ensure that the residuals are random. The null hypothesis in this test indicates that the residual autocorrelation represents a white noise series (Pankratz, 2009). The different steps involved in the Box-Jenkins methodology are summarized in Fig. 4.

Table 3:ACF and PACF patterns for ARMA

	AR (p) process	**MA (q) process**	**ARMA (p, q) process**
ACF	Infinite (exponential decay and/or damped sine waves)	Finite	Infinite (exponential decay and/or damped sine waves after first $q - p$ lags)
	Tails off towards 0	Cuts off to 0 after lag q	Tails off towards 0
PACF	Finite	Infinite (dominated by exponential decay and/or damped sine waves)	Infinite (dominated by exponential decay and/or damped sine waves after first $p - q$ lags)
	Cuts off to 0 after lag p	Tails off towards 0	Tails off towards 0

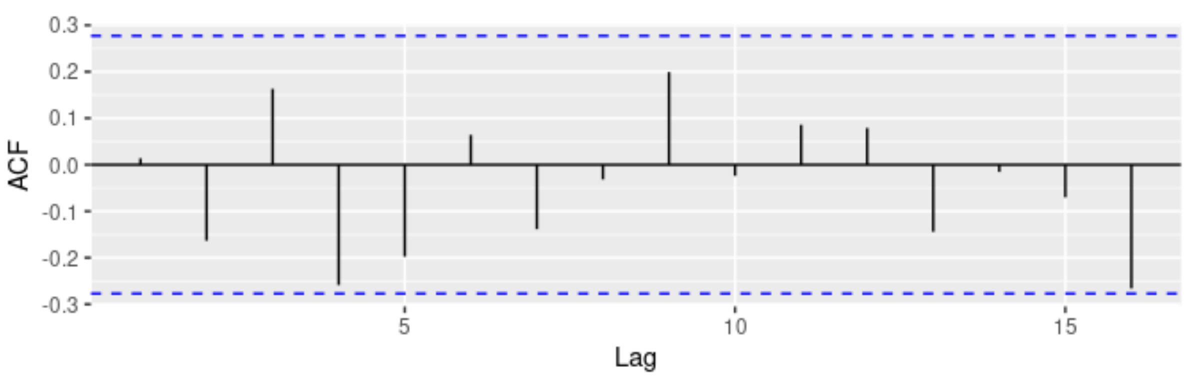


Figure 3: ACF plot for a white noise series

(Hyndman & Athanasopoulos, 2018)

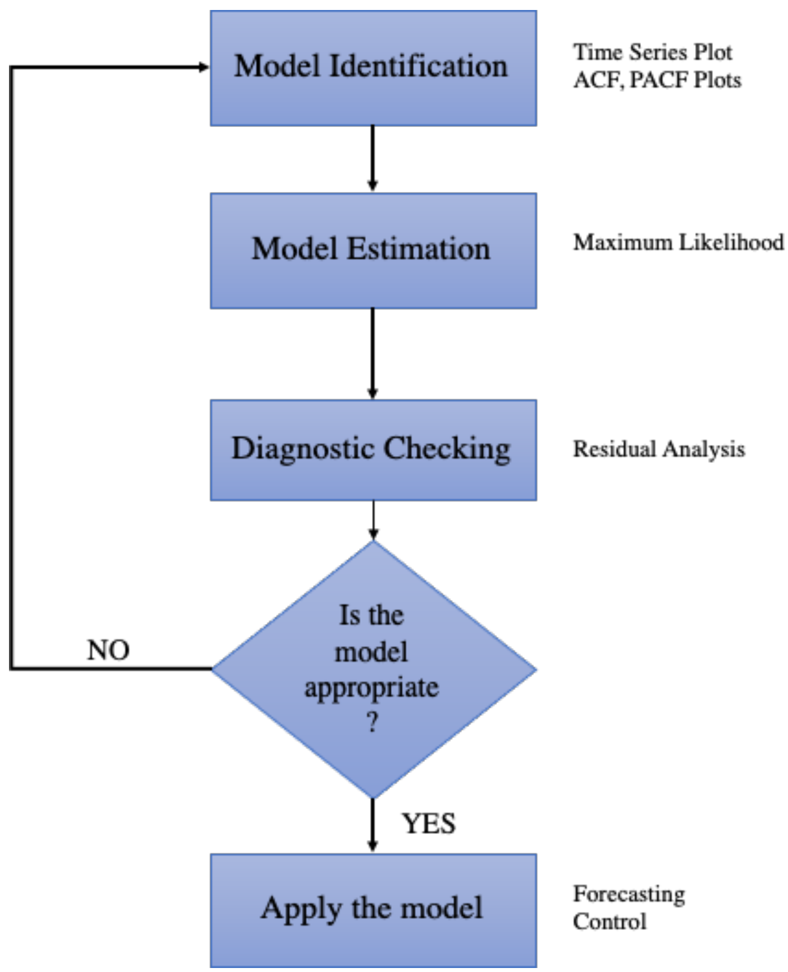


Figure 4: Box Jenkins methodology

2.8.1 Information based criteria for model identification

Model selection by analyzing sample plots can be time-consuming and error-prone, because real-life time series rarely reveal simple patterns. As a result, information-based criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) have been proposed to help choose the best mode(Akaike, 1973; Schwarz, 1978).

1. Akaike Information Criterion (AIC)

AIC is a commonly used statistical model measure. It essentially combines the model's goodness of fit and its parsimony into a single metric. AIC is computed as:

$$\begin{aligned} \text{AIC} &= -2\ \ln(\textit{maximized likelihood}) + 2r \\ &\approx n \ln\left(\hat{\sigma}_a{}^2\right) + 2r \end{aligned}$$

n = number of observations in the series

r = number of parameters estimated in the model

$\hat{\sigma}_a{}^2$= maximum likelihood estimation of the residual variance $\sigma_a{}^2$

2. Bayesian Information Criterion (BIC)

BIC is another statistical tool for model selection which attempts to correct the tendency of AIC to overfit. BIC is computed as

$$\begin{aligned} \text{BIC} &= -2\ \ln(\textit{maximized likelihood}) + r \ln(n) \\ &\approx n \ln\left(\hat{\sigma}_a{}^2\right) + r \ln(n) \end{aligned}$$

For both these tools, the preferred model is the one with the minimum value for the criterion chosen.

2.9 Application of Transfer Function Models

A transfer function relates two variables: the cause (forcing function or input variable) and the effect (response or output variable). There are many published examples of application of Box Jenkins transfer function models and over the years transfer function modeling have been applied to several fields including physical science, biology, transportation, economics, and engineering. Forecasting commodity prices, population response, and unemployment prediction are all examples of applications of transfer function models in the economic, social, and behavioral sciences. Liu studied the relationship between gasoline and crude oil prices in the U.S by employing transfer function models (Liu, 1991). By examining data from the PJM Interconnection to anticipate day-ahead electricity prices, transfer function models were constructed relating the electricity demand with prices(Nogales & Conejo, 2006). A similar study was conducted employed an ARIMA model approach to assess and forecast the day-ahead prices for the German electricity market(Jakaša, Andročec, & Sprčić, 2011). A transfer function analysis was used to investigate the impact of economic decisions made by the local county on the number residential, commercial, and industrial permits issued (McGinnis, 1994).

In biology, transfer function models are used to study the effects of different medications, the structure–function relationship of proteins, and the assessment of various cell populations. Monnet et al studied the relationship between antimicrobial use and bacterial resistance by developing a linear transfer function between the two parameters (Monnet et al., 2001). Aldeyab et.al. did a similar study over a 5-year period, using a multivariate ARIMA model to link antibiotic use to extended-spectrum beta (ESB) generating bacteria incidence rates and resistance patterns(Aldeyab et al., 2011). Parsons and Colbourne presented another application of the transfer function model by forecasting annual capture rates in a shrimp fishing location off the coast of Labrador using the annual winter ice cover as an input variable (Parsons & Colbourne, 2000).

One of the first industries to use the Box Jenkins technique for transfer function modeling was the transportation sector, notably in traffic volume forecasting and transportation planning studies. Nihan and Homesland analyzed monthly traffic flow from freeways from 1968 to 1976 to anticipate traffic levels in 1977, using the Box Jenkins technique (Nihan & Holmesland, 1980). Cools et al. used a seasonal ARIMA model with explanatory variables to analyze seasonality in daily traffic data and evaluated the influence of holidays on incoming traffic to two different sites. Holidays have a noticeable impact on commuter highways when compared to those used for leisure travel, according to ARIMA models developed. (Cools, Moons, & Wets, 2009). Kumar and Vanajakshi used seasonal ARIMA models to predict traffic patterns in the short run with limited input data(Kumar & Vanajakshi, 2015). Using historical values, transfer function modeling was also utilized to predict missing observations in traffic data. (Zhong & Sharma, 2006),(Harvey & Pierse, 1984).

The broadest application of transfer function modeling is in engineering and physical sciences. In engineering, however, transfer functions are often represented using a frequency-based technique rather than a time series analysis. Electrical engineers frequently perform linear analysis using transfer functions, which are then used to compare different designs. In their research, Raghavan and Satish computed a transfer function of an electric transformer by analyzing its equivalent circuit model (Ragavan & Satish, 2007). Transfer function modeling also has numerous applications in control systems. Qiang and Kui, in their study, analyzed the transfer function model of an open-loop Buck converter in a continuous conduction mode (CCM) and, from the models derived, found the various elements that influence its operation(Wang Fa-Qiang, 2013).

Transfer function modeling could be used in ways that are not traditionally associated with science and engineering. Huang and Wu conducted an empirical study in which they used an ARIMA model to examine and forecast a student's academic performance based on previous test results(Huang & Wu, 2014). By studying student enrollment data, ARIMA modeling was also utilized to investigate gender parity in accessing higher education in Taiwan, which can help support a higher education expansion program in the country(Chang, 2018).

Time series analysis in the hydropower sector necessitates the availability of characteristics such as reservoir inflow, pool elevation, total flow, and power generated, and the longer the series, the better. Instead of using a pre-defined simulation/optimization model, the time series of the outflow can be extracted and compared to the inflow in the upstream to examine the dependencies among them, and exploratory approaches like this uncover information regarding reservoir operation that might otherwise go unnoticed in a priori models. The use of the Box-Jenkins transfer function framework in this research is beneficial since trends in these two series can be recognized for better understanding and decision making, and can eventually be linked to water systems modeling, dispatch, and hydropower production cost modeling. This research and proposed methodology are not intended to establish a new model; instead, they should be able to supplement existing models by providing robust data management and modeling tools for water managers, policymakers, and other stakeholders.

Chapter 3: Methodology

Hydropower plants are highly non-linear and complex systems, and over the years, much research effort has been put into the modeling of facilities with different levels of detail. The nonlinear dynamic characteristics of hydro plant rely upon several uncertainties, and in a multi-purpose reservoir, the tradeoffs that make decisions beneficial for one purpose and harmful for other are often identified only by recording the effects of water released and reservoir elevations (T. Veselka, Ploussard, & Christian, 2020). Facility specific parameters like total flow, power generated, headwater, and tailwater is usually represented in a time series, and several insights could be gained by examining how they affect they functionality of a hydropower project. The proposed transfer function model is shown in Fig.5. Environmental outcomes, operational capabilities, and power system services are all inter-linked through hydropower operator decisions. Therefore, to consider a hydropower facility as a system, it is crucial to identify the different parameters entering and leaving each of these domains, the time scales, and the various external factors involved which may influence these signals.

3.1 Research Hypothesis

The main hypothesis of this thesis can be stated as follows:

Given information about the operating environment, a transfer function can represent the relationship between the inflow and outflow of the reservoir. For hydropower projects with the same classification (Run-Of-River, impoundment), these derived transfer functions will provide features that can be used to represent and model these facilities more consistently in the existing power and river system models.

The justification of this hypothesis begins with the analysis of time-series data of inflow and outflow, which are then used to develop a multivariate transfer function model.

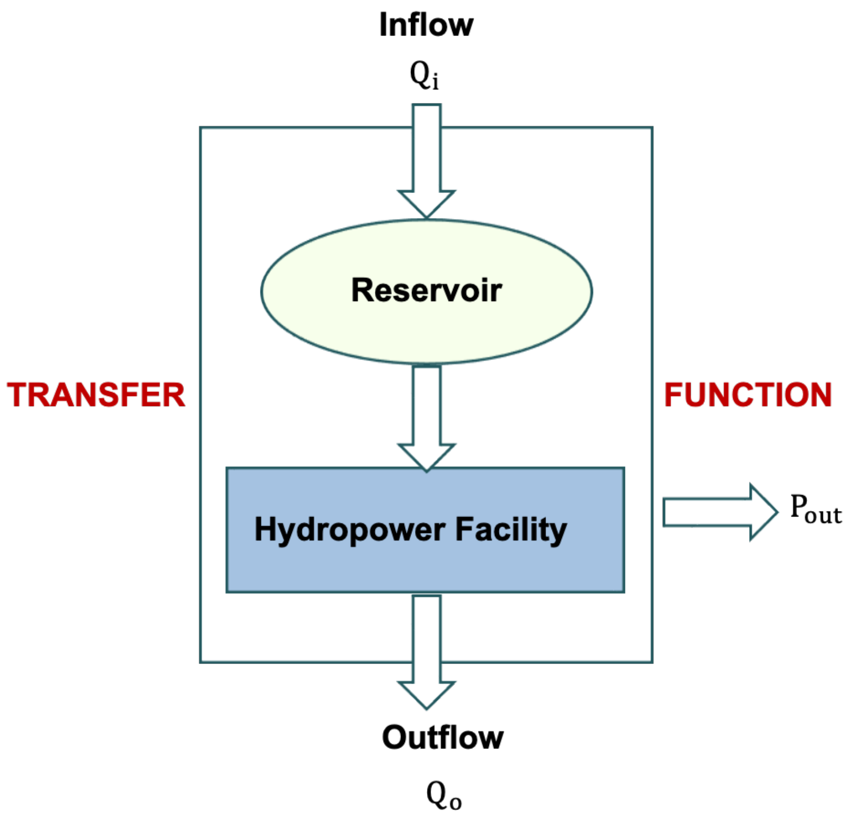


Figure 5: Proposed transfer function model

3.2 Transfer function modeling

ARIMA models are univariate time series models and as only one variable is analyzed, no relationships can be determined from this model. The purpose behind transfer function modeling is to evaluate the relationship between a target/output series and one (or more) explanatory/input series. If a series Y_t is influenced by another series X_t,a transfer function model can be deduced with Y_t as output and X_t as input of a dynamic linear system. This particular approach was developed principally by Box and the different procedures involved are discussed here.(G. E. P. Box et al., 2015)

Fig. 6 gives a schematic diagram of a linear system where the input x_t and output y_t is assumed to be stationary. According to the Box-Jenkins approach, a transfer function model with one input variable $\mathrm{x_t}$, may be split into two components,

$$y_t = u_t + n_t \tag{3.1}$$

Where $\mathrm{y_t}$ is the dependent variable which is transformed to achieve stationarity and $\mathrm{u_t}$ is the portion of $\mathrm{y_t}$ which can be described in terms of the input variable $\mathrm{x_t}$ and $\mathrm{n_t}$ is the error term which represents the sum of the effects of all variables other than the input.

The linear dynamic relationship between $\mathrm{x_t}$ and $\mathrm{u_t}$ can be represented as

$$\mathrm{u_t - \delta_1 u_{t-1} - \cdots - \delta_r u_{t-r} = \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \cdots - \omega_s x_{t-b-s}}$$

i.e.,

$$\mathrm{u_t} = \frac{\omega_0 - \omega_1 \mathrm{B} - \cdots - \omega_s \mathrm{B}^s}{1 - \delta_1 \mathrm{B} - \cdots - \delta_r \mathrm{B}^r} \mathrm{x_{t-b}}$$

$$= \frac{\omega(\mathrm{B})}{\delta(B)} \mathrm{x_{t-b}}$$

$$= \mathrm{v(B)x_t}$$

where $\mathrm{v(B)} = \frac{\omega(\mathrm{B})}{\delta(\mathrm{B})} \mathrm{B^b}$. The polynomial operator v(B), is also referred to as transfer function filter according to Box and Jenkins(G. E. Box & Jenkins, 1976), reflects the output-to-input transfer function and highlights the dynamic structure of the effect transferred from the input to the output sequence. The coefficients of the transfer function model are the impulse response weights $\mathrm{v_0}$,$\mathrm{v_1}$, $\mathrm{v_2}$ etc. of the polynomial operator v(B).

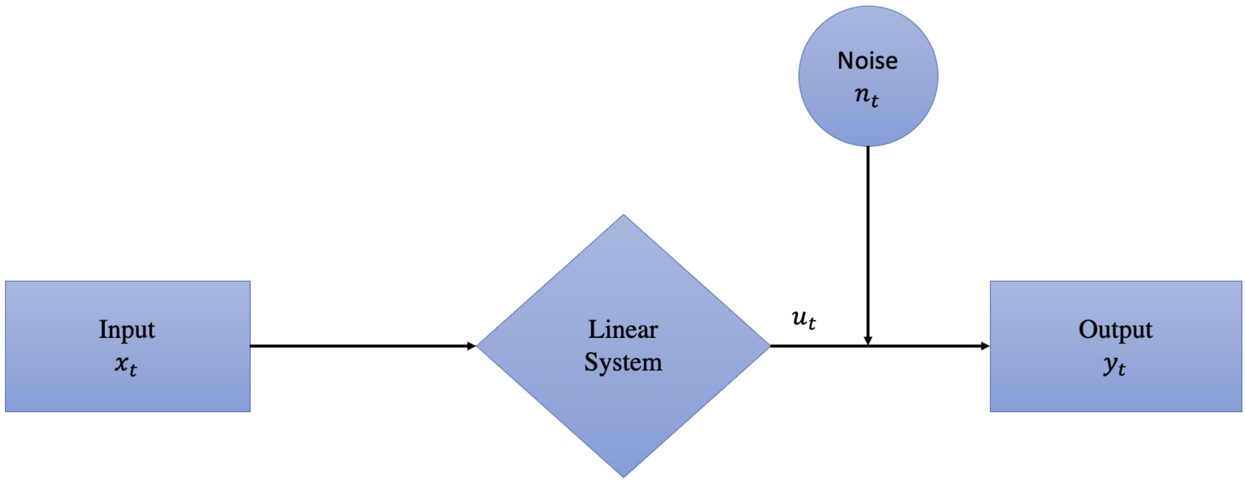


Figure 6: Schematic representation of a linear system

The noise Term n_t may be replaced by an ARMA (p, q) model of the form

$$n_t = c + \frac{\theta(B)}{\Phi(B)} a_t$$

where $\theta(B)$ and $\Phi(B)$ are AR and MA polynomials of order p and q respectively and a_t , a white noise series.

Equation (3.1) can be re-written as

$$y_t = c + \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\Phi(B)} a_t \tag{3.2}$$

In the transfer function model $\delta_1, \delta_2 \ldots \ldots \delta_r, \omega_0, \omega_1, \ldots \ldots \omega_s$ are the parameters, and (r, s, b) are integers greater than or equal to zero. The polynomial operator $\omega(B)$ described the magnitude of the immediate effects of input whereas $\delta(B)$ describes the duration and pattern of their decay. Hence, the order of model is defined by:

r=the number of lagged terms on output y_t

s =the number of lagged terms on x_t

b = delay time of response representing the number of periods before any visible effects

Transfer function modeling follows the same steps of ARIMA modeling, identification, estimation, and diagnostic checking and are outlined below.

3.2.1 Identification Stage

Box and Jenkins (G. E. Box & Jenkins, 1976) suggest the following procedures for the identification of the transfer function model:

1. Derive an estimate of the transfer function weights $\hat{v}_j$
2. Use the estimates of $\hat{v}_j$ to make approximate orders of r, s and b
3. Substitute the estimates of $\hat{v}_j$ in the following equations

$$v_j = 0 \qquad ; j < b$$

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots \delta_r v_{j-r} + \omega_0 \qquad ; j = b$$

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots \delta_r v_{j-r} - \omega_{j-b\mathrm{Com}} \qquad ; j = b+1, b+2, .. b+s$$

$$v_j = \delta_1 v_{j-1} + \delta_2 v_{j-2} + \cdots \delta_r v_{j-r} \qquad ; j > b + s.$$

If the actual $\hat{v}_j$ values were known, (r, s, b) can be estimated from the general information governing the impulse response weights. Impulse response weights consist of:

a) b zero values $v_0, v_1 ...., v_{b-1}$
b) A further s-r+1 values $v_b, v_{b+1} ......, v_{b+s-r}$ following no fixed pattern (only if $s \geq r$)
c) Values v_j with $j \geq b + s - r + 1$ following the pattern of a difference equation of order r, which has r starting values $v_{b+s}, ... v_{b+s-r+1}$

The estimation of the impulse response function and the identification of the transfer function noise model are described in the following two sections.

3.2.1.1 Estimation of impulse response function

Box and Jenkins present three strategies for determining transfer function weights in their book. Regression and the pre-whitening cross-correlation approach are the time domain methodologies, while cross spectral analysis approach is a frequency domain method. The pre-whitening cross correlation approach was preferred over the regression method in their analysis.

The transfer function weights $\hat{v}_j$ can be estimated from the sample cross correlation between the pre-whitened input and the transformed output. The steps involved in pre-whitening and subsequent impulse response weights are outlined in Fig. 7.

3.2.1.2 Pre-whitening of the input

If the input series x_t is autocorrelated, the effect of any changes in the input will take some time to manifest their effect. Therefore, we may observe non-causal effects, or changes in the output that appear to have occurred before changes in the input (Bisgaard & Kulahci, 2011). The process of removing the autocorrelation from an input series by identifying and fitting an ARMA model is known as pre-whitening. Removing all systematic and predicable components converts the input to a white noise process.

From the literature review, the general ARMA(p, q) model is represented by:

$$\Phi\,(B)x_t = \theta(B)\alpha_t \tag{3.4}$$

Given the fitted model the residuals a_t is computed by Equation (3.5)

$$\Phi\,(B)\theta^{-1}(B)x_t = \alpha_t \tag{3.5}$$

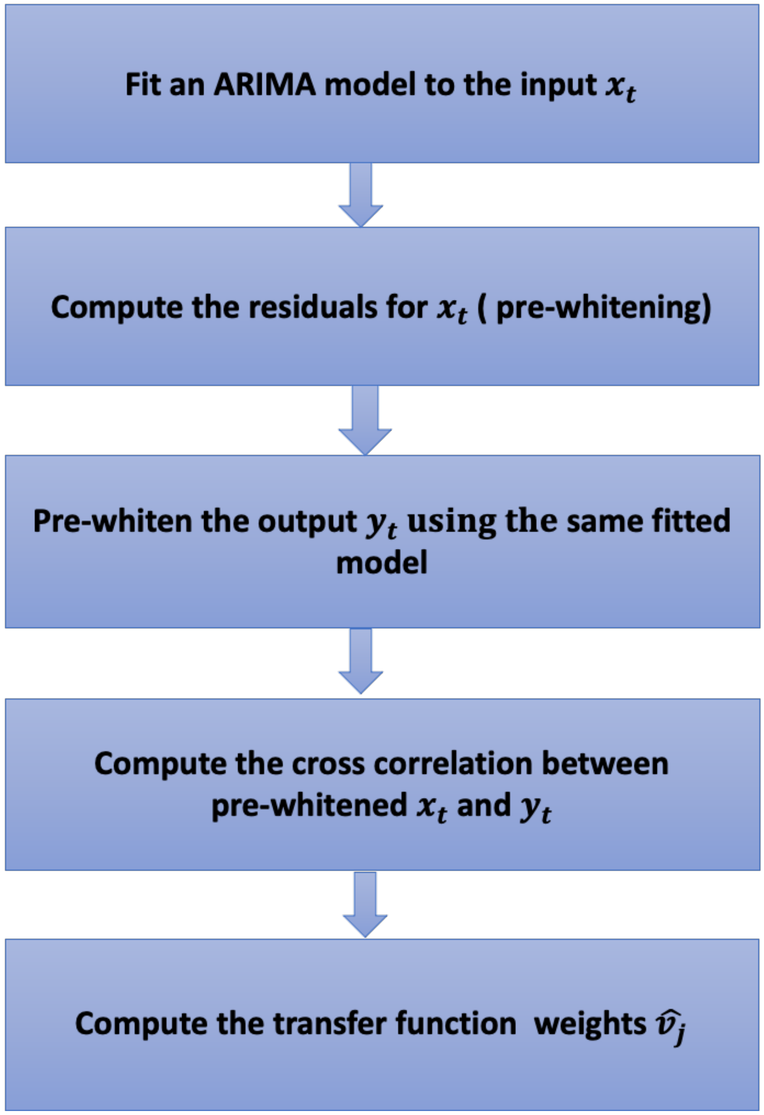


Figure 7:Procedure for estimating impulse response function

3.2.1.3 Transformation of the Output

Once the residuals α_t are computed, the output data is filtered through the same model. Thus, applying the same model to the output y_t gives

$$\Phi\,(B)\theta^{-1}(B)y_t = \beta_t \tag{3.6}$$

3.2.1.4 Computing the sample cross-correlation function and transfer function weights

From the flow chart in Fig. 6, the next step is to compute the cross-correlation $r_{\alpha,\beta}(k)$between pre-whitened input α_t and output β_t. The cross correlations at lag k are directly proportional to the impulse response function $\widehat{v}_j$ and therefore, the sample cross correlation function provides estimates of the transfer function weights (G. E. P. Box et al., 2015). Box and Jenkins showed that the rough estimates of $\widehat{v}_j$ can be computed as

$$\widehat{v}_j = \frac{s_\beta}{s_\alpha} r_{\alpha,\beta}(k),\ \ j = 0, \dots, k \tag{3.7}$$

Where s_β and s_α are the estimated standard deviations of the pre-whitened output and input. A reasonable approximation of the standard error of the cross correlation for n observation is provided by $\frac{1}{\sqrt{n}}$,and the significance of a given $\widehat{v}_j$ can be determined.

3.3 Identification of the transfer function model

Identification of transfer function model involves finding the appropriate order of (r, s, b) as shown in Eq. (3.2). Once the transfer function weights are computed, the characteristic decay pattern it is plotted. In most cases, only a few parameters controlling $\omega(B)$ and $\delta(B)$ are enough to represent the lags found in the input. Some examples of impulse response functions for specific transfer function models (with n_t assumed to be zero) are shown in Fig. 8.

The task of visually identifying an appropriate model Is inherently subjective and becomes more complicated when the noise term n_t grows more significant relative to the input x_t.

From Eq. (3.2)

$$v(B) = \frac{\omega(B)}{\delta(B)} B^b$$

$$\delta(B)v(B) = \omega(B)B^b$$

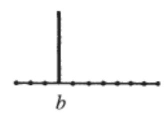
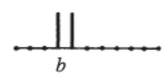
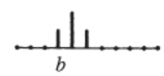
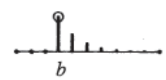
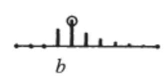
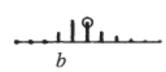
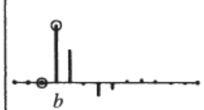
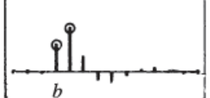
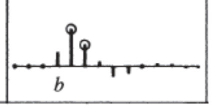
r, s, b	∇ Form	B Form	Impulse Response v_j
003	$Y_t = X_{t-3}$	$Y_t = B^3 X_t$	
013	$Y_t = (1 - 0.5\nabla)\, X_{t-3}$	$Y_t = (0.5 + 0.5B)\, B^3 X_t$	
023	$Y_t = (1 - \nabla + 0.25\nabla^2)\, X_{t-3}$	$Y_t = (0.25 + 0.50B + 0.25B^2)\, B^3 X_t$	
103	$(1+\nabla)\, Y_t = X_{t-3}$	$(1 - 0.5B)\, Y_t = 0.5B^3 X_t$	
113	$(1+\nabla)\, Y_t = (1 - 0.5\nabla)\, X_{t-3}$	$(1 - 0.5B)\, Y_t = (0.25 + 0.25B)\, B^3 X_t$	
123	$(1+\nabla)\, Y_t = (1 - \nabla + 0.25\nabla^2)\, X_{t-3}$	$(1 - 0.5B)\, Y_t = (0.125 + 0.25B + 0.125B^2)\, B^3 X_t$	
203	$(1 - 0.25\nabla + 0.5\nabla^2)\, Y_t = X_{t-3}$	$(1 - 0.6B + 0.4B^2)\, Y_t = 0.8B^3 X_t$	
213	$(1 - 0.25\nabla + 0.5\nabla^2)\, Y_t = (1 - 0.5\nabla)\, X_{t-3}$	$(1 - 0.6B + 0.4B^2)\, Y_t = (0.4 + 0.4B)\, B^3 X$	
223	$(1 - 0.25\nabla + 0.5\nabla^2)\, Y_t = (1 - \nabla + 0.25\nabla^2)\, X_{t-3}$	$(1 - 0.6B + 0.4B^2)\, Y_t = (0.2 + 0.4B + 0.2B^2)\, B^3 X$	

Figure 8: Examples of impulse response functions of commonly adopted transfer functions

Adapted from G. E. P. Box, Jenkins, Reinsel, & Ljung (2015, p. 410)

$$(1-\delta_1 B-\cdots..-\delta_r B^r)(v_0+v_1 B+\cdots)=(\omega_0-\omega_1 B-\cdots-\omega_s B^s)B^b$$

This gives.

$$v_j-\delta_1 v_{j-1}-\delta_2 v_{j-2}-\cdots-\delta_r v_{j-r}=\begin{cases}-\omega_{j-b} & j=b+1,\dots,b+s\\ 0 & j>b+s\end{cases} \tag{3.8}$$

Once the values of r, s and b are identified, it is then substituted in Eq.7 to get the polynomial parameters $\omega(B)$ and $\delta(B)$.

3.3.1 Identification of a parsimonious noise model

Once the initial transfer function is identified and estimated, the next step is to identify the noise model.

$$Y_t=v(B)x_t+n_t$$

$$n_t=y_t-v(B)x_t$$

As preliminary estimates of transfer function weights $\hat{v}_j$ are estimated the noise series is given by:

$$n_t=y_t-v_0 x_t-v_1 x_{t-1}-v_2 x_{t-2}-\cdots$$

Alternatively, the noise term can also be calculated by replacing $v(B)$ with the initial estimates of the parameters $\omega(B)$ and $\delta(B)$ computed in Eq. (3.8)

$$n_t=y_t-\frac{\omega_0(B)}{\delta_0(B)}x_{t-b}$$

Then, the ACF and PACF of the estimated noise series can be obtained and used to identify the ARIMA (p, q) noise model represented as

$$n_t=c+\frac{\theta(B)}{\Phi(B)}a_t$$

3.3.2 Estimation Stage

As outlined in the literature review, the maximum likelihood estimates for the parameters in $\omega(B)$ and $\delta(B)$ for each potential transfer function model and in $\theta(B)$ and $\Phi(B)$ for the corresponding noise model are calculated in this stage. The estimation procedure is calculation-intensive and iterative, leading to long execution times. Finally, the model parameters with the best least square fit are chosen.

3.4 Diagnostic Checking

The final step is to put the chosen transfer function model to diagnostic tests. The residuals a_t of the tentative model are examined to see if they correlate with the unsystematic changes in the input x_t. Statistical methods, such as the Ljung-Box Chi-Square test, can be used to establish that the residuals are random.

Chapter 4: Data and Analysis

4.1 Data Availability

Approximately half of the hydroelectric generating capacity is owned and operated by the federal government in the U.S. (Bracmort, Stern, & Vann, 2013) The major federal entities include the U.S Army Corps of Engineers (Corps), the Bureau of Reclamation, and the Tennessee Valley Authority (TVA). Fleet owners collect the operations and maintenance data for their records and mandate reporting to the North American Electric Reliability Corporation (NERC).

4.1.1 Fleet Data

For this research, the required time series data were obtained from the TVA staff for the hydropower projects as shown in Table 4. The data was recorded on a 5-minute time step between January 2004 and December 2016, and these include the date, time, total power, gross head, headwater, tailwater, water temperature, spill, and total flow. The TVA also provided the data for plotting the elevation-storage curve for the reservoir. Because the TVA system's reservoirs operate as a network, operating policies relating to a single reservoir can vary depending on external events throughout the system.

4.2 Data Preprocessing

This study uses the inflow to the reservoir and the downstream discharge to compute a transfer function model of the ten facilities. The outflow, which is the amount of water that departs the reservoir every second, is recorded by the TVA. The pre-processing techniques required to construct a transfer function model are presented using the Norris hydropower plant as an example, and identical steps are followed for the other facilities studies. Relevant equations are also explained, and save for clarity, are not repeated in the other facilities evaluated.

4.2.1 Description of the study area- Norris hydropower facility

In East Tennessee, Norris Reservoir stretches 73 miles up the Clinch River and 56 miles up the Powell River from the Norris Dam. It is the first dam built by the TVA and its largest tributary storage impoundment (Fig. 9). A multipurpose project, the reservoir's primary purpose includes flood control and hydroelectric power generation while also supporting secondary uses such as water supply, recreation, and providing habitats for aquatic life(Authority).

Table 4: Hydropower projects studied

Storage Facilities	Run-of-river facilities
Norris Dam	Watts Bar Dam
Cherokee Dam	Chickamauga Dam
Fontana Dam	Guntersville Dam
Douglas Dam	Nickajack Dam
Blue Ridge Dam	Fort Loudon Dam

Figure 9:Norris Dam

Located in the Clinch River basin, the hydroelectric power plant consists of two generating units with a summer net dependable capacity of 126 MW, about 3 percent of the total hydropower capacity of the TVA system. The reservoir also has a flood storage capacity of 1,113,000 acre-ft, and the flood detention capacity varies seasonally. (Authority, 2022). The reservoir's annual operating cycle includes releasing water through summer and fall to generate hydropower during peak demand periods and drawing the reservoir to its flood control level at the beginning of the year to store the runoff from heavy rains during the winter months. The operating guide is shown in Fig. 10 and a snapshot of the data is shown in Fig. 11. (Authority).

4.2.2 Inflow computation

Efficient water management strategies and resource planning are required to maximize the value of water resources. Thus, the knowledge of historical water inflow data is very relevant as it defines the input into the reservoirs and, therefore, the eventual plant output. Unfortunately, there are no direct methods to measure water inflows as it typically comprises the stream runoffs and the surrounding tributaries. Even though the water discharge data of the mainstream is readily available, corresponding observations from the tributaries are not easily quantifiable. Reservoir inflow is conventionally estimated using the water balance method, which involves the reservoir release and the change in storage during the period considered. According to the water balance method (Chow, Maidment, & Mays, 1988),

$$\mathrm{Q_i} = \mathrm{Q_r} + \frac{\mathrm{V_{i+1}} - \mathrm{V_i}}{\Delta\mathrm{T}} + \mathrm{Q_l}$$

$\mathrm{Q_i}$=Reservoir inflow, $\mathrm{Q_r}$=Reservoir release $\mathrm{V_i}$=Initial storage volume, $\mathrm{V_{i+1}}$=Final storage volume $\Delta\mathrm{T}$= Time period under consideration $\mathrm{Q_l}$= Water losses (including evaporation and seepage losses)

The reservoir storage $\mathrm{V_I}$ is estimated by the reservoir stage-elevation relationship

$$\mathrm{V_t} = f(HW)$$

Here, HW is the observed water level. It should be noted that the water losses are not usually quantified so for this calculation, it is assumed to be negligible. The data for computing the relationship between the reservoir storage and elevation were provided by the TVA through direct correspondence, which is plotted to obtain the function as shown in Fig 12. A second-degree polynomial relationship is identified between the elevation level and reservoir volume, expressed as:

$$V_t = (5.7 \times 10^5)\, HW^2 - (1.04 \times 10^{10}) HW + 4.69 \times 10^{12} \tag{4.1}$$

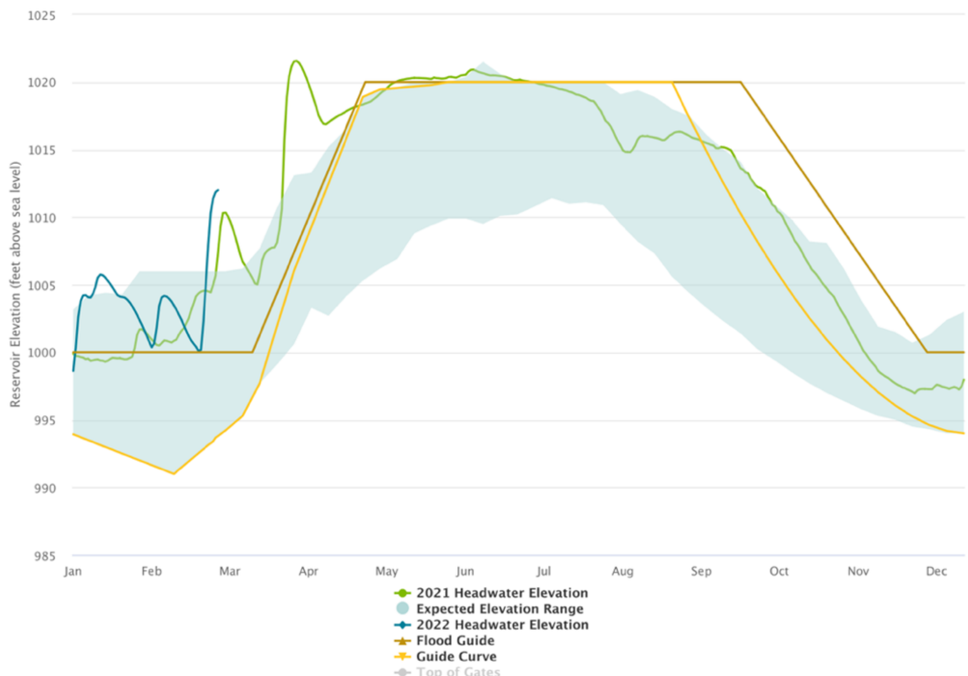


Figure 10:Operating guide of Norris dam

(Authority, 2021)

Category	2021 Headwater Elevation	Expected Elevation Range		2022 Headwater Elevation	Flood Guide	Guide Curve
		low	high			
01/02/2021	999.89	993.93	1003.17	998.63	1000	993.93
01/03/2021	999.73	993.86	1003.34	1000.32	1000	993.86
01/04/2021	999.71	993.79	1003.51	1002.63	1000	993.79
01/05/2021	999.65	993.71	1003.69	1003.79	1000	993.71
01/06/2021	999.61	993.64	1003.86	1004.21	1000	993.64
01/07/2021	999.49	993.57	1004.03	1004.23	1000	993.57
01/08/2021	999.53	993.5	1004.2	1004.08	1000	993.5
01/09/2021	999.37	993.43	1004.23	1004.05	1000	993.43
01/10/2021	999.41	993.36	1004.26	1004.33	1000	993.36
01/11/2021	999.45	993.29	1004.29	1004.81	1000	993.29
01/12/2021	999.47	993.21	1004.31	1005.53	1000	993.21
01/13/2021	999.39	993.14	1004.34	1005.75	1000	993.14
01/14/2021	999.38	993.07	1004.37	1005.71	1000	993.07
01/15/2021	999.31	993	1004.4	1005.55	1000	993

Figure 11: Synopsis of operation data

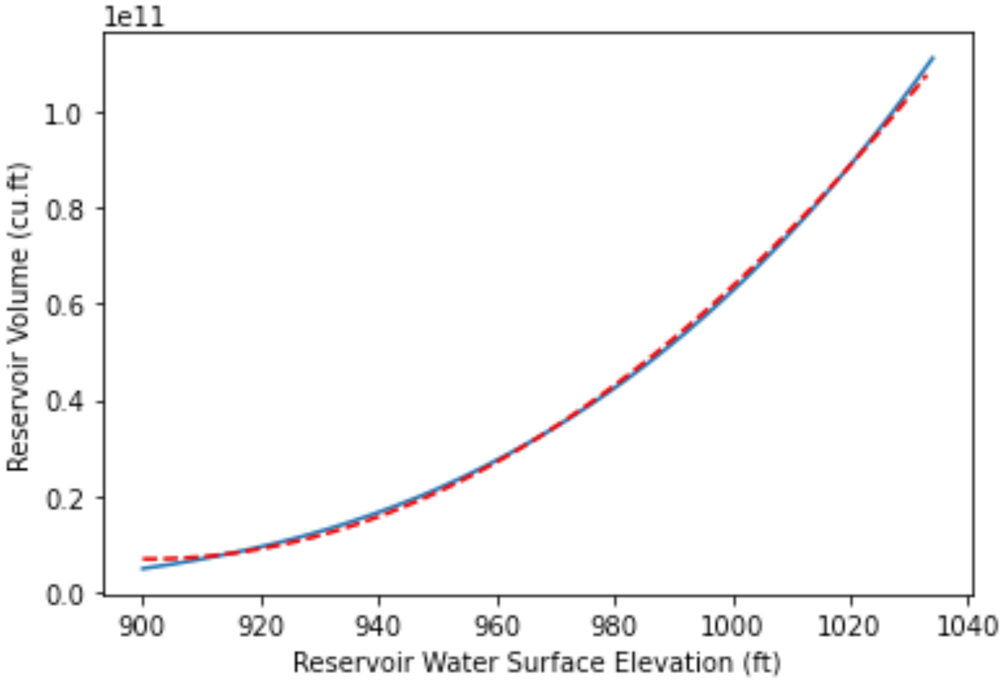


Figure 12:Curve-fitting- Reservoir volume as a function of elevation

4.2.2.1 Negative inflow values: sources of errors

The initial inflow estimation produced inflow values with significant negative terms as shown in Fig.13 and 14. When computing the reservoir volume, each coefficient in Eq. (4.1) has associated errors, and the following reasons can cause estimated negative inflows.

a) A small inaccuracy in the reservoir elevation observations might cause a considerable variation in the reservoir volume estimation (Fig 15).
b) Due to the lack of recent data for computing the reservoir volume- elevation curve, the reservoir capacity at specific elevations can be miscalculated.
c) The estimated uncertainties in computed outflows for lakes having continuous-record gaging stations at or near their outlets range from 5 to 10% (Winter, 1981)

However, when daily and weekly inflows are calculated, the magnitude of negative values is reduced, and the effects of headwater fluctuations are minimized, as illustrated in Fig.16.

4.2.2.2 Replacing computed negative inflow values

The primary input for transfer function modeling is reservoir inflow and fitting an ARIMA model to the input for obtaining the impulse response function is the first step in the Box Jenkins identification stage. Because the input is so critical in the design of a transfer function model, it's crucial to modify the negative inflows so that the inflow time series can be used for further analysis. The leading cause of errors in reservoir storage estimation and subsequent inflow computation is uncertain elevation level change and Fig. 15 depicts the fluctuations in the headwater levels of Norris Dam. Waves and seiches can also potentially produce inaccuracies in elevation level readings, in addition to human errors during measurement. In a USGS report that assessed the daily inflows and outflows of eight regulated lakes in the Oswego River basin, Lumia and Moore devised a reservoir level hydrograph smoothing technique to deal with changes in lake levels (Lumia & Moore, 1983). They used interpolation to replace null headwater data points, and fluctuations in the headwater observations are smoothed using an analog-based low pass Butterworth filter, as illustrated in Fig.17. Even though the resulting inflows have been much reduced, the time series still contains negative inflow values.

To replace the remaining negative values, a methodology proposed by Goel et al. is applied. In this process, the sum of negative values for a water year is adjusted in the positive values in proportion to their magnitude. The negative inflow term is then changed to zero, resulting in water balance for the hydrological year (Goel, Jain, Rani, & Chalisgaonkar, 2018).

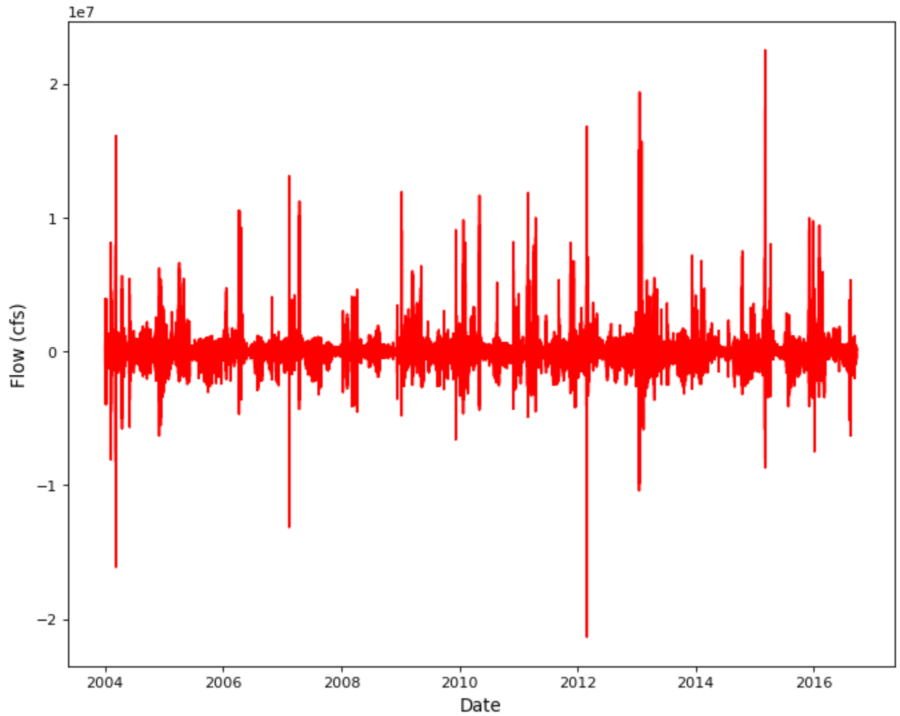


Figure 13:Computed reservoir inflow

Date	Time	GrossHead	HeadWater	PlantTotalFlow	PlantTotalPower	TailWater	Reservoir Volume	Change in Storage	Inflow
2004-01-01	07:45:00	163.69	991.01	6584.0	83.9	827.32	4.645778e+14	8.318392e+07	283863.737708
2004-01-01	07:50:00	163.70	991.00	6735.0	86.0	827.30	4.645778e+14	-1.040203e+07	-27938.436458
2004-01-01	07:55:00	163.73	991.00	7413.0	94.2	827.27	4.645778e+14	0.000000e+00	7413.000000
2004-01-01	08:00:00	163.73	990.99	7979.0	98.9	827.26	4.645778e+14	-1.040088e+07	-26690.588333
2004-01-01	08:05:00	163.75	991.00	8107.0	100.1	827.25	4.645778e+14	1.040088e+07	42776.588333
...	...	...	...	...	...	...	...	...	...
2016-09-21	07:50:00	181.38	1002.45	0.0	-2.3	821.07	4.645904e+14	0.000000e+00	0.000000
2016-09-21	07:45:00	181.36	1002.45	0.0	-2.5	821.09	4.645904e+14	0.000000e+00	0.000000
2016-09-21	07:40:00	181.34	1002.45	0.0	-2.5	821.11	4.645904e+14	0.000000e+00	0.000000
2016-09-21	12:05:00	178.83	1002.43	3440.0	45.5	823.60	4.645904e+14	-2.344445e+07	-74708.169792
2016-09-21	06:10:00	181.02	1002.45	0.0	-2.3	821.43	4.645904e+14	2.344445e+07	78148.169792

Figure 14: Computed reservoir Inflow – data

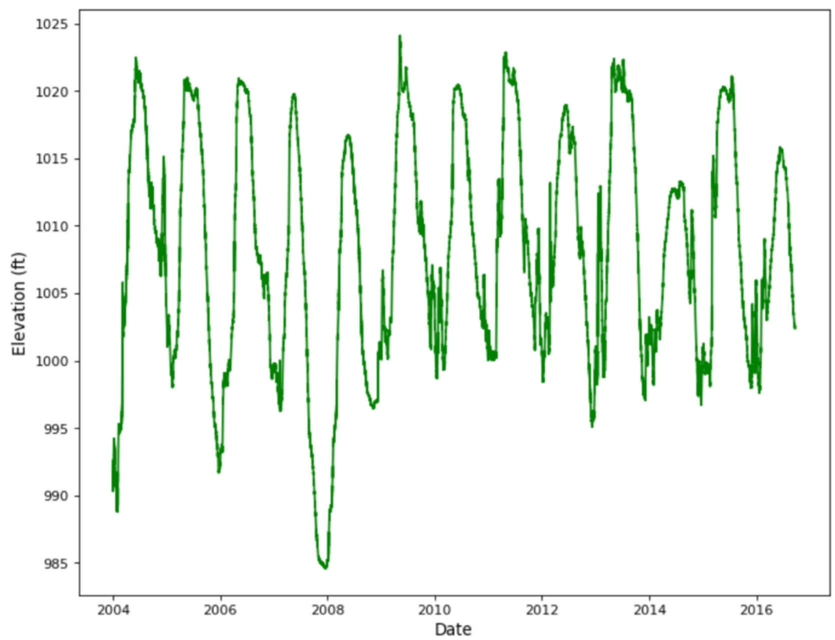


Figure 15:Recorded elevation levels

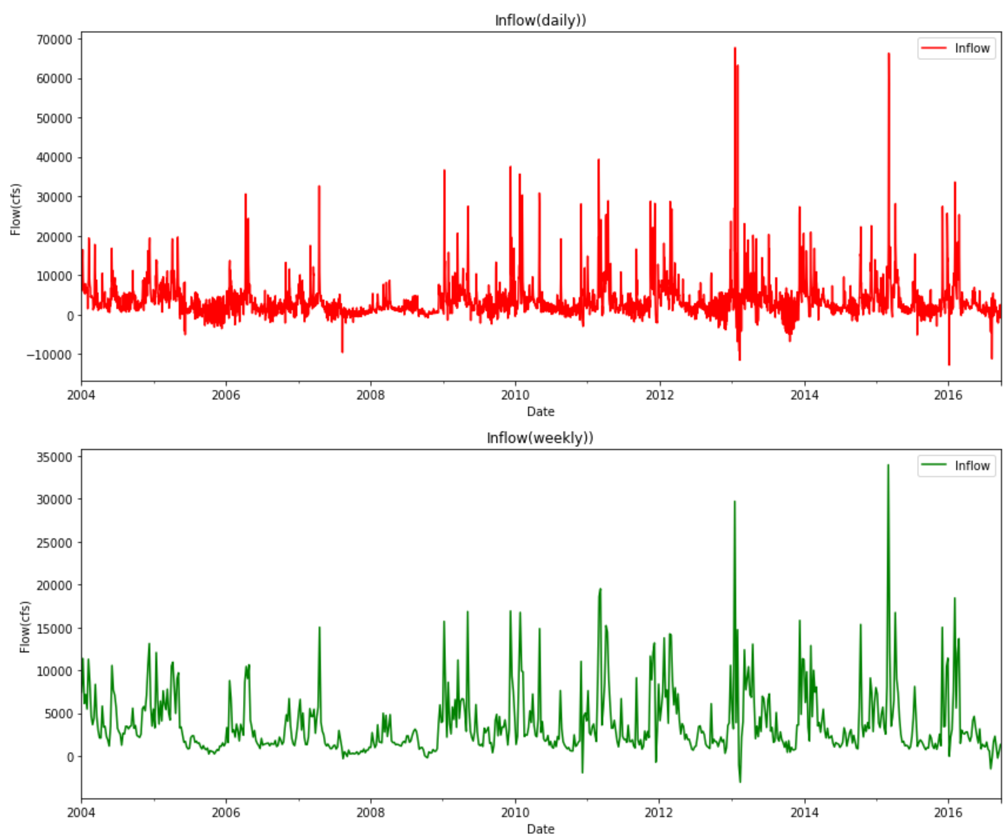


Figure 16:Computed net inflows- daily and weekly

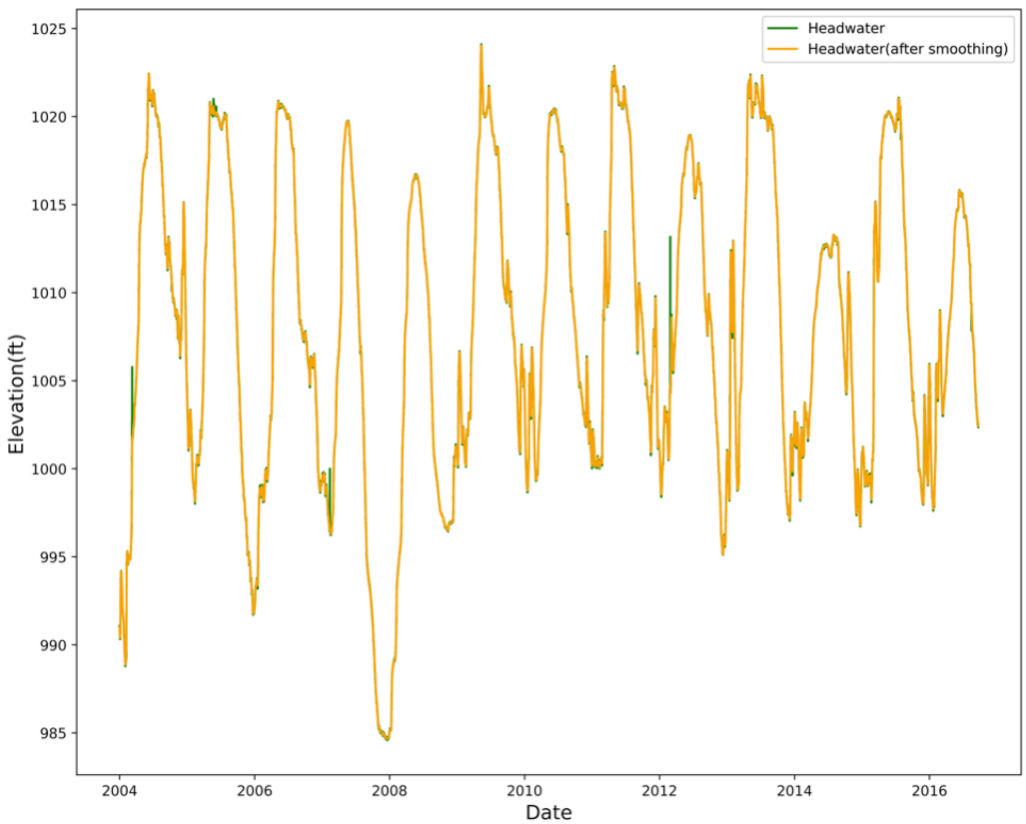


Figure 17: Elevation levels after smoothing

According to USGS, a water year is defined as a 12-month period from October 1 through September 30, for any given year. The updated inflow, $\mathrm{I_u}$ is calculated as

$$\mathrm{I_u} = \begin{cases} \mathrm{I_C} - \left(\frac{\mathrm{I_C}}{\mathrm{S}}\right) \times \mathrm{N} & \mathrm{I_c} \geq 0 \\ 0 & \mathrm{I_c} < 0 \end{cases}$$

S = Magnitude of the sum of total positive inflows in a water year

N = Magnitude of the sum of total negative inflows in a water year

$\mathrm{I_C}$ = Computed inflow from the water balance equation

4.2.2.3 Evaluation of the methodology

Direct statistical evaluation of the proposed methodology is impossible since there are no direct methods for computing the inflow into a reservoir. To evaluate the proposed technique, the Pearson correlation coefficient between computed inflows and the streamflow at the nearby unregulated streams is calculated. The Pearson correlation coefficient, also known as Pearson's r, is a measure of the linear dependence between two variables. The value of r lies between $[-1,1]$, with -1 representing complete negative correlation and 1, a positive correlation (Boslaugh, 2012). For the Norris Dam, about 70 percent of the total inflow comes from the Clinch River basin, and the Powell River provides the remaining 30 percent (Authority). The daily historical measurement from these gauges were obtained from the public USGS database (USGS NWIS, 2010) and correlated to the concurrent inflow computed using water balance approach and the inflow computed using the proposed methodology. The results of the correlation are presented in Table 5. The r values in the inflows calculated using the water balance method show a moderate correlation with the streamflow values. However, applying the proposed methodology shows a significant improvement in the correlation with the results indicating a strong positive correlation between the updated inflow values and the corresponding values from the gages. A comparison of the resulting inflow time series after applying smoothing techniques and replacing the negative values with the updated inflow is illustrated in Fig. 18.

Table 5: Results of correlation analysis

Gages used	Type of inflow data	Value of r	Improvement in correlation
Clinch River above Tazewell Powell River near Arthur	Inflow (Water Balance Method)	0.58	0.16
	Inflow (Proposed Methodology)	0.74	

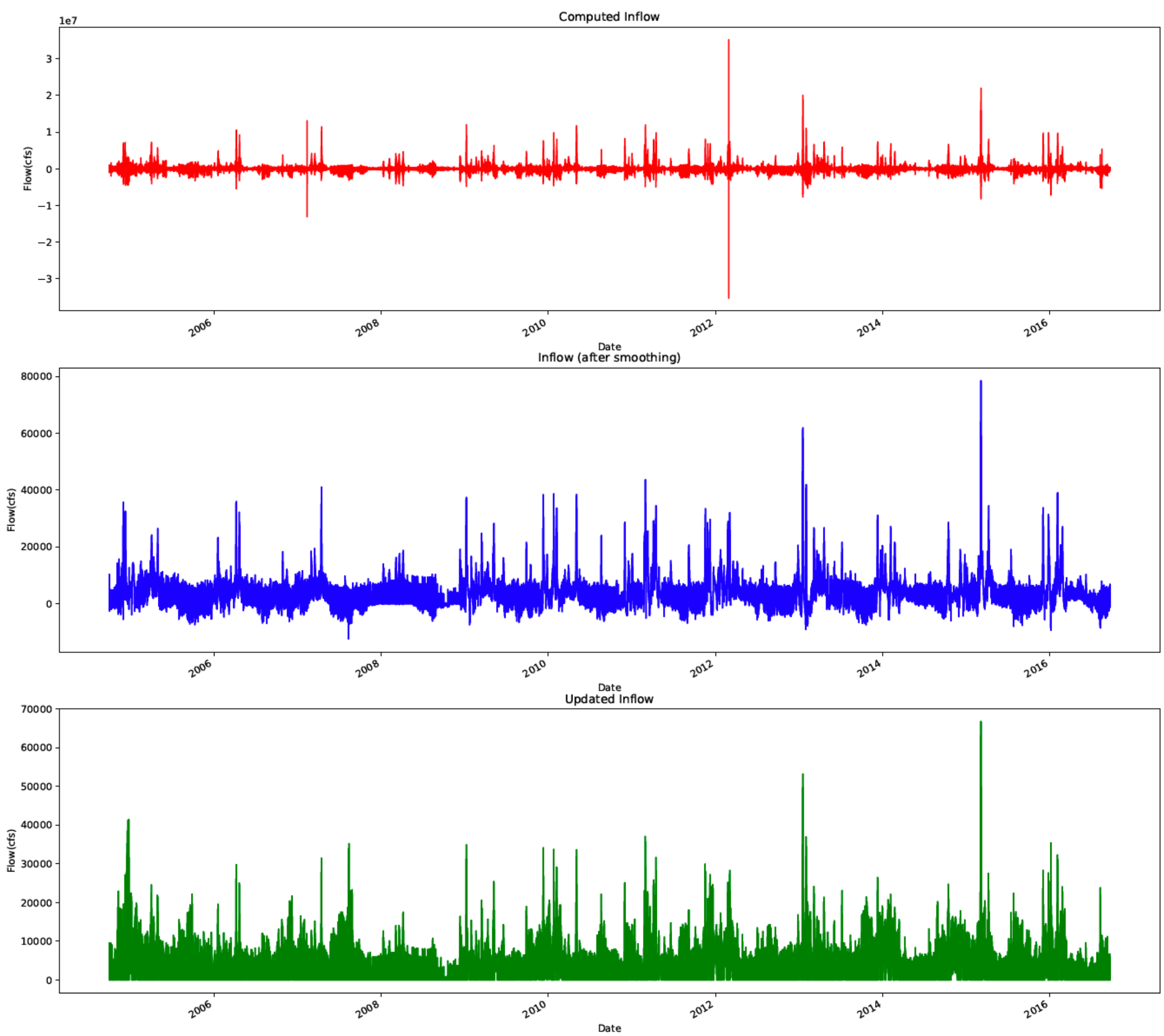


Figure 18: Replacing negative inflow values -different stages

4.2.3 Choosing the appropriate time step

The time-series data provided by the TVA is recorded at the 5-minute interval, and for developing a transfer function model, there are four options available (Fig.19):

a) Using the dataset as is, with 5-minute frequency resulting in 277882 input values. This may necessitate much processing, and understanding a pattern may be challenging
b) In hours, resample the data, yielding 104929 values. While some patterns are emerging, most of the data produced are sparse and there is still a significant number of data points to examine.
c) Resample the data for obtaining the daily inflow value. Even if the number of terms has decreased dramatically (4373), the trend remains the same as in the hourly data.
d) Resample the data every week, for a total of 626-time steps. Even if the data is less granular and follows the same daily data trends, performing diagnostic tests could be an issue later.

The changes in inflow time series across different time scales are illustrated in Fig.19. It should also be observed that the monthly data shows a pattern, but this is of less importance in terms of modeling because most of the existing models focus on simulations with short time periods.
While there is no "correct" time step for transfer function modeling, hourly and daily values are used for model development in this research.

4.3 List of software used

The following software and packages were used in this thesis for developing the Box Jenkins model:

1. Python: The Pandas library is used to perform the initial data preparation, which includes steps like combining the 5-minute fleet, imputing missing values and incorrect data points, data resampling and determining appropriate range for the different variables. Machine learning libraries such as NumPy, SciPy and scikit-learn were utilized for designing Butterworth filter and corresponding inflow computations.
2. JMP Pro: After obtaining the daily inflows and outflows, JMP software's ARIMA modeling capabilities was used to compare various ARIMA models and choose the one with the lowest AIC values.

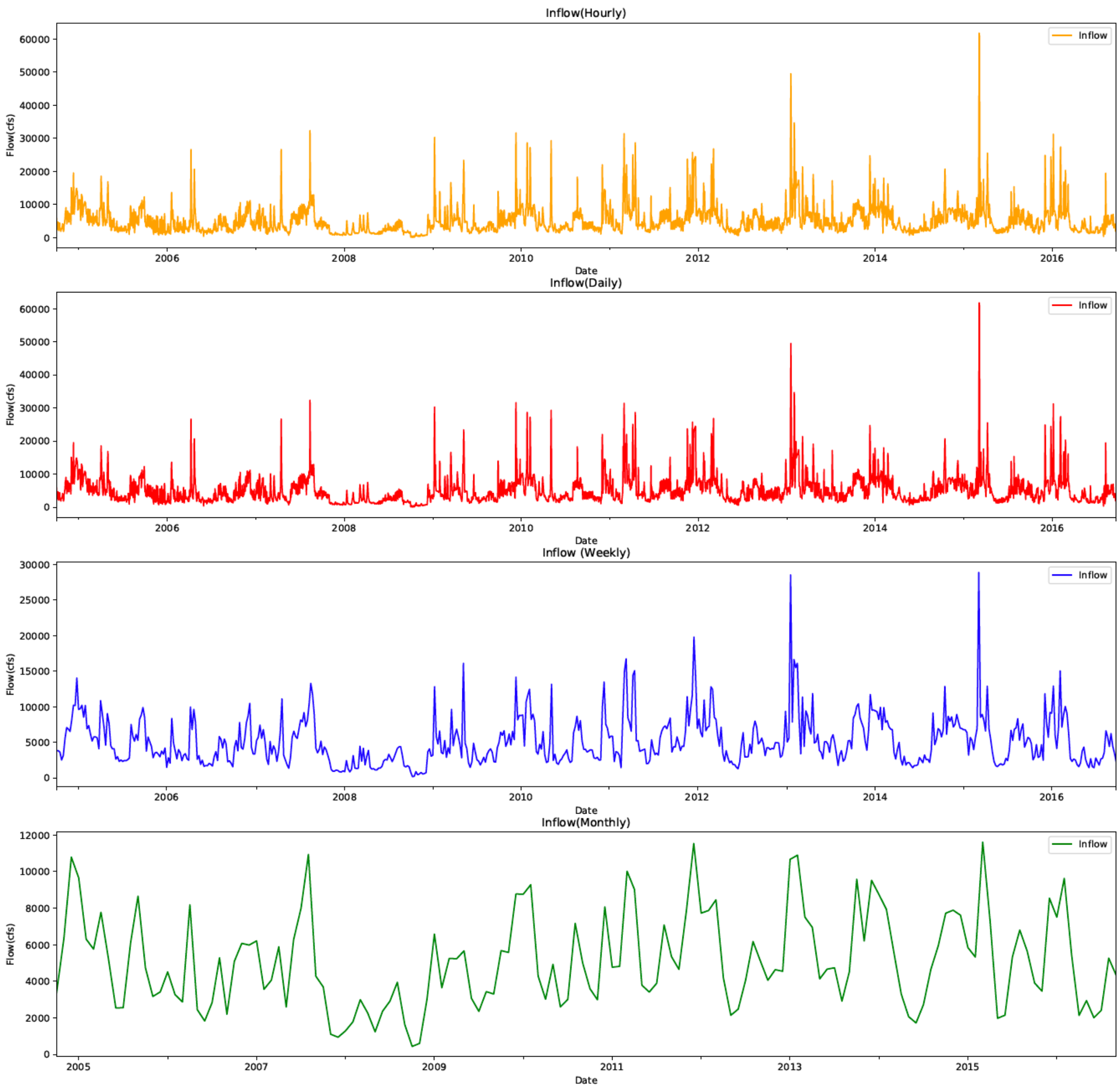


Figure 19: Inflow across different time intervals

3. SAS Studio: Using PROC ARIMA in SAS Studio, the results for the various stages in the Box Jenkins approach, including pre-whitening, CCF computation, and transfer function model coefficients are obtained.

4.4 Analysis

After the data has been pre-processed and the inflow has been calculated, the Box- Jenkins approach discussed in Chapter 3 has been applied to data from five storage facilities and five runoff river facilities. The Box-Jenkins approach is based on the premise that the time series being studied is stationary. As a result, the data structure must be examined as because it could lead to the selection of the inappropriate ARIMA model and transfer function.

Decomposing a time series involves considering a combination of level, trend, seasonality, and noise. Level is the value that goes on average with time, while the trend is the data's progressive upward or downward movement over time. The seasonal component explains the patters that repeat at regular intervals and when you separate seasonality and trend from the time series, you are left with the noise or random component. One can obtain more insight and understanding into the nature of a time series by decomposing it based on such elements and plotting the result. There are two types of decomposition: additive and multiplicative. If S_t, T_t and R_t represents the seasonality, trend, and residuals respectively, for a datax_t, additive decomposition is written as (Hyndman & Athanasopoulos, 2018),

$$x_t = S_t + T_t + R_t$$

A multiplicative decomposition is represented as

$$x_t = S_t \times T_t \times R_t$$

If the seasonal variations are constant and periodic, additive decomposition is recommended. Seasonal fluctuations are constant for hydropower flow data, except for years with extreme weather conditions, and so additive decomposition is used. R software has various time series analysis packages, including decomposition, forecasting, ARIMA modeling, and more functions

4.5 Application of the Box- Jenkins Approach to data from storage facilities

The location of the storage sites selected are shown on the map in Fig. 20, and additional information on the elevation levels, number of units and net dependable capacity is given in Table 6. The daily data from 2004-2016 were used for analysis and the development of transfer function model. The plots in Appendix A show the inflow and outflow of the five facilities studied over the 13 years. The ADF test is conducted utilizing the *statsmodels* package in Python and the results of suggest that the inflow time series is stationary, but the seasonal decomposition shows a seasonal pattern, so a seasonal differencing is necessary to remove the cyclic trend from the data.

4.5.1 Transfer function Modeling

The Box Jenkins transfer function modeling methodology described in Chapter 3 are applied to the inflow and outflow time series obtained from these facilities. The relevant equations were presented in Chapter 3 and are only mentioned here for clarity. Mathematical symbols are used if equations are utilized, and they follow the notations used in Chapter 3.

4.5.1.1 Pre-whitening inflow and outflow

The first step in Box Jenkins methodology is to pre-whiten the inflow time series. In this case the ACF and PACF plots of the inflow series are insufficient to anticipate the p and q parameters of the ARIMA model. Therefore, an automated algorithm, the ARIMA model group of jmp software ("ARIMA Modelling,") was used to identify the best fit with the lowest AIC value (Section 2.7.1). These include setting the maximum value of p and q to 4 and seasonal orders P and Q to 1. Due to the presence of seasonality, the value D is also set to 1. The "best fit" seasonal ARMA model and the pre-whitening filter for the five storage facilities are listed in Table 7. The coefficients of the pre-whitening filter is obtained using SAS Studio software, and in particular SAS proc ARIMA (SAS & ETS, 2014). The outflow time series was then transformed using the pre-whitening filter that has been fit into the inflow.

4.5.1.2 Interpreting the CCF plots

The outflow from a reservoir may be related to the prior lags of inflow and the sample cross correlation is useful for finding the inflow lags that might be useful to establish that relationship. A negative lag value represents a correlation between outflow at time t and the inflow at a time before t and a positive lag value indicates a correlation between outflow at time t and the inflow at a time after t.

Figure 20: Map showing location of storage facilities

Table 6: Storage facilities studied

Facility	Minimum Elevation (ft)	Maximum Elevation (ft)	No. of units	Net dependable capacity (MW)
Norris Dam	960	1031	2	126
Cherokee Dam	1020	1071	4	122
Fontana Dam	1580	1710	3	304
Douglas Dam	940	994	4	182
Blue Ridge Dam	1616	1691	1	16

Table 7: Pre-whitening of Inflow time series- storage facilities

Facility	SARIMA fit	Pre-whitening filter for Inflow
Norris	(2,0,4) (1,1,1) [7]	$\left[\frac{(1+1.7\mathrm{B}+0.93\mathrm{B}^2+0.23\mathrm{B}^3+0.001\mathrm{B}^4)(1-0.88\mathrm{B}^7)}{(1+0.44\mathrm{B}-0.55\mathrm{B}^2)(1-0.06\mathrm{B}^7)}\right]\mathrm{Inflow}_t$
Cherokee	(3,0,1) (1,1,1) [7]	$\left[\frac{(1+\mathrm{B})(1-0.95\mathrm{B}^7)}{(1+0.17\mathrm{B}-0.78\mathrm{B}^2+0.05\mathrm{B}^3)(1-0.12\mathrm{B}^7)}\right]\mathrm{Inflow}_t$
Fontana	(4,0,4) (0,1,1) [7]	$\left[\frac{(1-2.1\mathrm{B}+1.6\mathrm{B}^2-0.13\mathrm{B}^3-0.25\mathrm{B}^4)(1-0.98\mathrm{B}^7)}{(1-2.9\mathrm{B}+3.5\mathrm{B}^2-1.9\mathrm{B}^3+0.4\mathrm{B}^4)}\right]\mathrm{Inflow}_t$
Douglas	(3,0,2) (1,1,1) [7]	$\left[\frac{(1-0.78\mathrm{B}-0.14\mathrm{B}^2)(1-0.95\mathrm{B}^7)}{(1-1.7\mathrm{B}+0.84\mathrm{B}^2-0.09\mathrm{B}^3)(1-0.05\mathrm{B}^7)}\right]\mathrm{Inflow}_t$
Blue Ridge	(4,0,4) (1,1,1) [7]	$\left[\frac{(1-0.88\mathrm{B}-0.1\mathrm{B}^2+0.13\mathrm{B}^3-0.11\mathrm{B}^4)(1-0.99\mathrm{B}^7)}{(1-1.38\mathrm{B}+0.51\mathrm{B}^2-0.28\mathrm{B}^3+0.16\mathrm{B}^4)(1-0.003\mathrm{B}^7)}\right]\mathrm{Inflow}_t$

Additionally, significant positive spikes in the CCF plot indicate that changes in the inflow causes changes in outflow to follow suit, while negative spikes suggest that the input variable may be providing feedback to the output variable. Sample cross-correlations between the pre-whitened inflow and outflow time series are plotted in Fig. 21.

Although the CCF plots of the storage facilities differed widely, they shared some crucial characteristics. First, the lag-zero weight was the largest across all the sites, indicating that the inflow at current day was more significant on the outflow response. Second, in four of these sites except for Blue Ridge, the lag-1 and lag+1 weights were larger relative to other lags. Third, considerable inflow input for Norris and Douglas dam was suggested at higher order lags than t-1 and t+1. The disparity between these two plots and the plot for other facilities leads one to believe that the quantity of inflows from the previous day that influenced the present outflow may range significantly from site to site.

4.5.1.3 Identification of transfer function models

The impulse response function $\mathrm{v}(B)$ is given by

$$\mathrm{v}(B) = \frac{\omega(B)}{\delta(B)} \mathrm{B}^{\mathrm{b}}$$

where

$$\omega(B) = \omega_0 - \omega_1 \mathrm{B} - \cdots - \omega_s \mathrm{B}^{\mathrm{s}}$$
$$\delta(B) = 1 - \delta_1 \mathrm{B} - \cdots - \delta_r \mathrm{B}^{\mathrm{r}}$$

The impulse-response weight pattern can be visualized from the CCF plots. The numerator $\omega(B)$ and denominator $\delta(B)$ play distinct roles in representing the impulse response patterns (Pankratz, 2012) :

1. Up until the lag, which establishes the delay time b between the input and the output, the impulse response weights are 0.
2. The spikes in the CCF plot that are not a part of the decay pattern are captured by the numerator factor $\omega(B)$. The lag after which the transfer function weights exhibit a decay can be thought of as the value s.
3. The denominator $\delta(B)$ dictates the pattern of decay. In case of simple exponential decay, $r = 1$ and in compound exponential or sinusoidal decay patters, $r = 2$

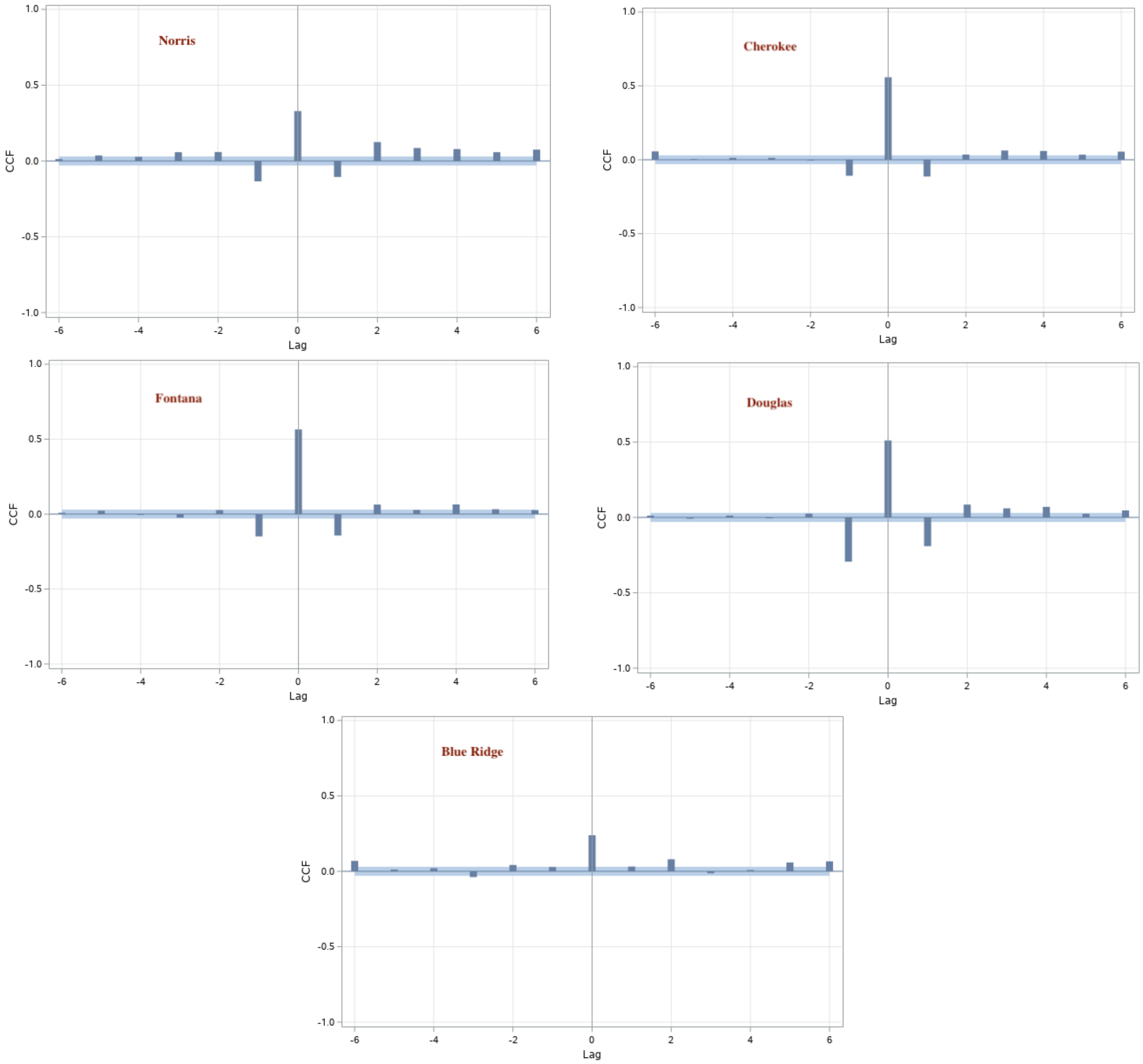


Figure 21: CCF Plots of storage facilities

In all the CCF plots above, as there are no initial zero-valued weights, the delay time $b = 0$. The values of r and s are determine by examining the characteristics of the decay pattern. To fit the transfer function, IDENTIFY and ESTIMATE functions in the proc ARIMA functionality of SAS was utilized, and the residuals were examined. The plots of residual correlation and normality diagnostics are listed in Appendix B. From the residual correlation analysis, the order of the noise term $\mathrm{n_t}$ was also identified. The outcomes of the final transfer function model are listed in Table 8 where Y_t represents outflow and X_t represents inflow.

4.6 Application of the Box- Jenkins approach to data from run-of-river facilities

The location of the storage sites selected are shown on the map in Fig. 22, and additional information on the elevation levels, number of units and net dependable capacity is given in Table 9. Like storage facilities, daily data from 2004-2016 were used for analysis and the development of transfer function model. The plots in Appendix A shows the inflow and outflow of the five facilities studied over the 13 years.

4.6.1 Transfer function modeling

The inflow and outflow time series from these facilities are modeled using the Box Jenkins modeling approach.

4.6.1.1 <u>Pre-whitening inflow and outflow</u>

The details seasonal ARIMA model and the pre-whitening filter obtained using the SAS Studio software are listed in Table 10. The outflow is subsequently transformed using the same filter.

4.6.1.2 <u>Interpreting the CCF Plots</u>

Sample cross-correlations between the pre-whitened inflow and outflow time series are plotted in Fig. 23. The CCF plots of the five run-of-river facilities displayed different characteristics. The lag-zero weight was the largest across all the sites, indicating that the inflow at current day has higher effect on the outflow. Additionally, in three of these sites, Watts bar, Chickamauga and Guntersville, the lag-2 and lag+2 weights were significant than other lags. Lastly, in both Fort Loudon and Nickajack facilities, the only significant correlation is at lag 0.

4.6.1.3 <u>Identification of transfer function model</u>

The outcomes of the final transfer function model developed for these five run-of-river facilities studied are listed in Table 11.

Table 8: Identified Transfer- Function Models: storage

Facility	$(\boldsymbol{r}, \boldsymbol{s}\ \boldsymbol{b})$	**TFN Model**
Norris	(1,1,0)	$Y_t = -0.98 + 0.08\left[\frac{1 + 0.03\mathrm{B}}{1 + 0.64\mathrm{B}}\right] X_t + \left[\frac{1}{(1 - 0.7\mathrm{B})}\right] \mathrm{a_t}$
Cherokee	(2,1,0)	$Y_t = -4.6 + 0.81\left[\frac{1 - 0.3\mathrm{B}}{1 - 0.07\mathrm{B}^2}\right] X_t + \left[\frac{(1 + 0.37\mathrm{B})}{(1 - 0.56\mathrm{B})}\right] \mathrm{a_t}$
Fontana	(2,1,0)	$Y_t = -0.37 + 0.59\left[\frac{1 - 0.36\mathrm{B}}{1 - 0.07\mathrm{B}^2}\right] X_t + \left[\frac{(1 + 0.42\mathrm{B})}{(1 - 0.53\mathrm{B})}\right] \mathrm{a_t}$
Douglas	(1,1,0)	$Y_t = -7.78 + 0.6\left[\frac{1 - 0.24\mathrm{B}}{1 + 0.28\mathrm{B}}\right] X_t + \left[\frac{(1 + 0.47\mathrm{B})}{(1 - 0.55\mathrm{B})}\right] \mathrm{a_t}$
Blue Ridge	(1,0,0)	$Y_t = -7.78 + 0.08\left[\frac{1}{1 + 0.38\mathrm{B}}\right] X_t + \left[\frac{(1 + 0.22\mathrm{B})}{(1 - 0.48\mathrm{B})}\right] \mathrm{a_t}$

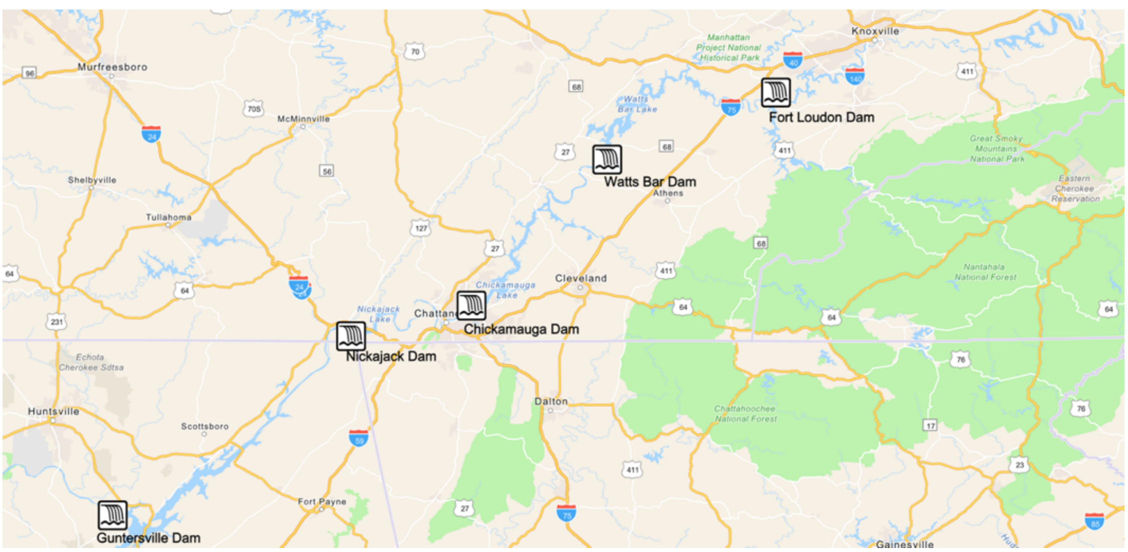


Figure 22: Map showing locations of run-of-river facilities

Table 9: Run-of-river facilities studied

Facility	Minimum Elevation (ft)	Maximum Elevation (ft)	No. of units	Net dependable capacity (MW)
Watts Bar	735	745	5	196
Chickamauga	675	685	4	142
Guntersville	593	595	4	123
Fort Loudon	807	813	4	151
Nickajack	632	635	4	107

Table 10: Pre- whitening of Inflow time series- run-of-river facilities

Facility	SARIMA fit	Pre-whitening filter for Inflow
Watts Bar	(4,0,4) (1,1,1) [7]	$\left[\frac{(1-0.64\mathrm{B}-0.99\mathrm{B}^2+0.33\mathrm{B}^3-0.3\mathrm{B}^4)(1-0.97\mathrm{B}^7)}{(1-1.51\mathrm{B}-0.17\mathrm{B}^2+0.91\mathrm{B}^3-0.23\mathrm{B}^4)(1-0.03\mathrm{B}^7)}\right]\mathrm{Inflow}_\mathrm{t}$
Chickamauga	(3,0,3) (1,1,1) [7]	$\left[\frac{(1+0.27\mathrm{B}-0.64\mathrm{B}^2-0.28\mathrm{B}^3)(1-0.94\mathrm{B}^7)}{(1-0.6\mathrm{B}-0.6\mathrm{B}^2+0.22\mathrm{B}^3)(1-0.04\mathrm{B}^7)}\right]\mathrm{Inflow}_\mathrm{t}$
Guntersville	(4,0,4) (0,1,1) [7]	$\left[\frac{(1-1.41\mathrm{B}+1.06\mathrm{B}^2-0.14\mathrm{B}^3-0.27\mathrm{B}^4)(1-0.94\mathrm{B}^7)}{(1-2.06\mathrm{B}+2.11\mathrm{B}^2-1.13\mathrm{B}^3+0.1\mathrm{B}^4)}\right]\mathrm{Inflow}_\mathrm{t}$
Fort Loudon	(3,0,2) (1,1,1) [7]	$\left[\frac{(1-0.29\mathrm{B}-0.44\mathrm{B}^2)(1-0.97\mathrm{B}^7)}{(1-0.95\mathrm{B}-0.19\mathrm{B}^2+0.16\mathrm{B}^3)(1-0.08\mathrm{B}^7)}\right]\mathrm{Inflow}_\mathrm{t}$
Nickajack	(3,0,4) (1,1,1) [7]	$\left[\frac{(1+1.2\mathrm{B}-0.19\mathrm{B}^2-0.56\mathrm{B}^3-0.08\mathrm{B}^4)(1-0.96\mathrm{B}^7)}{(1+0.69\mathrm{B}-0.87\mathrm{B}^2-0.67\mathrm{B}^3)(1-0.08\mathrm{B}^7)}\right]\mathrm{Inflow}_\mathrm{t}$

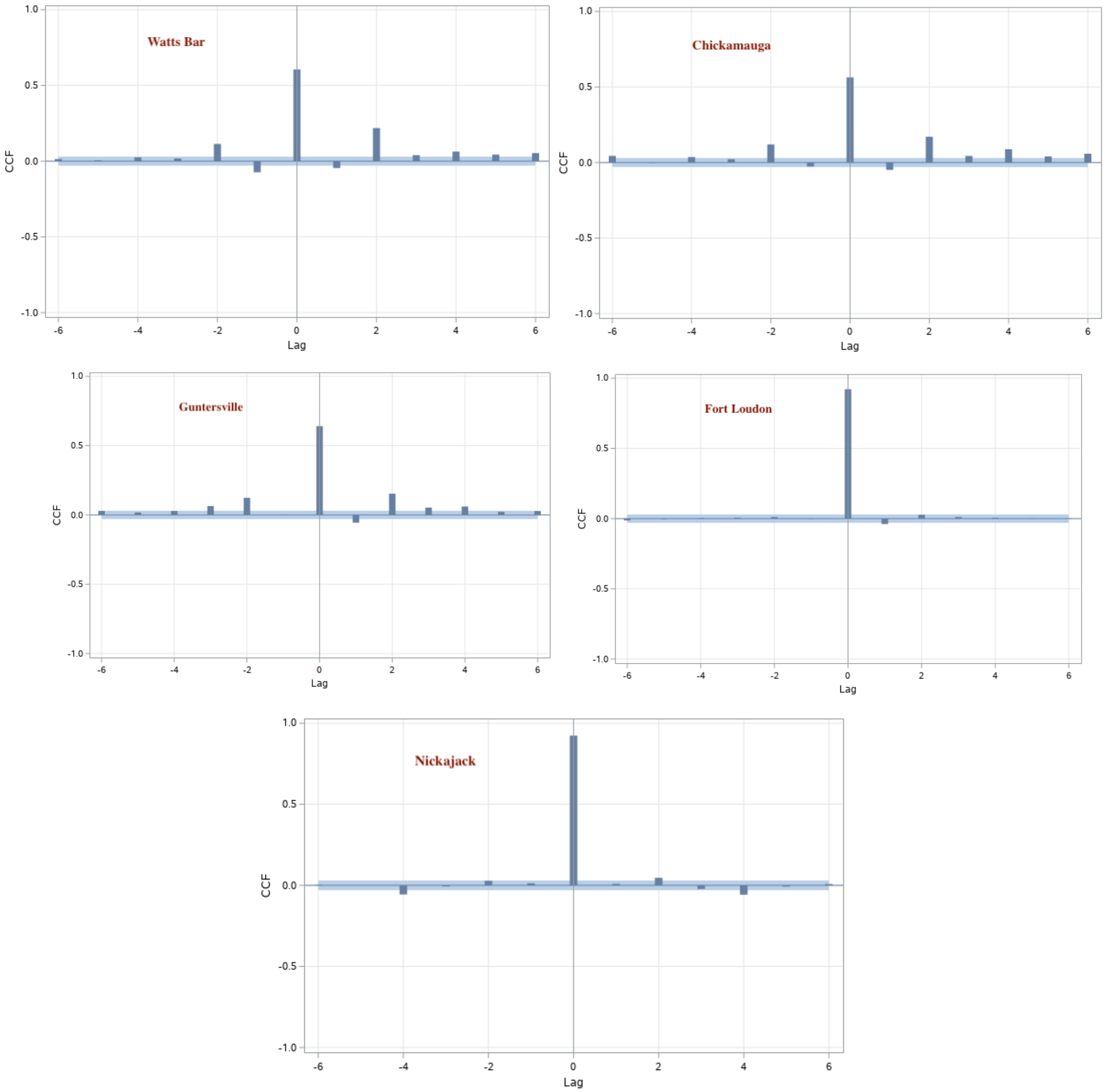


Figure 23: CCF plots of run-of-river facilities

Table 11: Identified Transfer- Function Model: run-of-river

Facility	$(\boldsymbol{r}, \boldsymbol{s}\ \boldsymbol{b})$	**TFN Model**
Watts Bar	(1,2,0)	$Y_t = -14.84 + 0.48\left[\frac{1 + 0.42\mathrm{B}^2}{1 + 0.07\mathrm{B}}\right] X_t + \left[\frac{1 + 0.37\mathrm{B}}{(1 - 0.31\mathrm{B})}\right] \mathrm{a_t}$
Chickamauga	(2,2,0)	$Y_t = 5.67 + 0.44\left[\frac{1 - 0.13\mathrm{B}^2}{1 - 0.52\mathrm{B}^2}\right] X_t + \left[\frac{(1 + 0.56\mathrm{B})}{(1 - 0.19\mathrm{B}^2)}\right] \mathrm{a_t}$
Guntersville	(1,2,0)	$Y_t = -12.61 + 0.6\left[\frac{1 + 0.24\mathrm{B}^2}{1 + 0.06\mathrm{B}}\right] X_t + \left[\frac{(1 + 0.49\mathrm{B})}{(1 - 0.008\mathrm{B})}\right] \mathrm{a_t}$
Fort Loudon	(1,0,0)	$Y_t = -0.92 + \left[\frac{1}{1 + 0.039\mathrm{B}}\right] X_t + \left[\frac{(1 + 0.59\mathrm{B})}{(1 - 0.08\mathrm{B})}\right] \mathrm{a_t}$
Nickajack	(1,0,0)	$Y_t = 0.66 + 0.99\left[\frac{1}{1 - 0.008\mathrm{B}}\right] X_t + \left[\frac{(1 - 0.32\mathrm{B}^2)}{(1 + 0.24\mathrm{B})}\right] \mathrm{a_t}$

4.7 Checking the fitted models

Once the transfer function model is identified, the final step is to check the adequacy of the model chosen by undergoing various diagnostic tests. Additionally, the model should also meet all the criteria below:

i. The model should only include a few parameters, adhering to the parsimony principle.
ii. A stable linear dynamic system must be represented by the transfer function components of the model.
iii. There should be no autocorrelations within the residuals of the model, and it should be independent of the input variable.
iv. The ARIMA noise component n_t should be stationary.

Chapter 5: Results and Discussion

In a transfer function model, the dynamic relationship between output Y_t and input X_t is:

$$Y_t = C + v(B)X_t + n_t$$

As the intercept C and noise term n_t are independent of the input X_t, we can learn how Y_t responds to changes in X_t by utilizing the individual v weights. The positive and negative signs of the weights indicate how much or how little the output increases or decreases when there is a change in input. The transfer function weights can take on a large variety of patterns in practice, thus we can't be certain which pattern is best for a given data set. In the previous chapter some assumptions were made regarding the order of $(r, s\ b)$ and a transfer function model was developed. Before using the results to characterize the relationship between inflow and outflow, a Maximum Likelihood Estimation (MLE) method is utilized to understand the significance of the different coefficients (δ, ω) obtained. The standard error, t-ratio, and p-value that the MLE test in SAS produces allow us to determine the significance of the coefficient in the model.

5.1 General results of transfer function modeling

5.1.1 Storage Facilities

The initial CCF plot of the five storage facilities provides evidence of correlation between outflow at time t and the inflow at a time t and its several lags. The parameter estimates of individual facilities with the results of MLE methods are shown in Table 12. The relationship obtained between outflow and inflow are also summarized in Table 13. The information obtained from the transfer function relationship can be categorized as:

1. Past Information: It is clear from the relationship obtained that outflow at time t is related to the past values of inflows and outflows. The inflow from the day before affects the outflow in three of the five storage facilities under investigation. Regarding the effects of prior outflow values themselves, the results are varied, with Cherokee and Fontana flows depending on the outflow of the previous two days while for the other three, the outflows from the previous day had more impact on the value of the present day.
2. Current Information: In all the storage facilities studied, the current day inflow information has the higher impact on the outflow values. In Cherokee, Fontana, and Douglas, the magnitude of increase in outflow with a change in inflow is higher than the other two facilities analyzed.

Table 12: Diagnostic checks on parameter estimates- storage facilities

Facility	Parameter Estimate	SE	*t*-Ratio	p-value
Norris	-0.03	0.16	-0.17	0.8636
	-0.64	0.1	-6.3	<0.0001
Cherokee	0.3	0.03	11.36	<0.0001
	0.07	0.02	3.27	0.0011
Fontana	0.36	0.02	13.98	<0.0001
	0.07	0.01	3.52	0.0004
Douglas	0.24	0.06	4.07	<0.0001
	-0.29	0.04	-7.09	<0.0001
Blue Ridge	-0.38	0.06	-6.44	<0.0001

Table 13: Inflow-Outflow relationship of storage facilities

Facility	Inflow-Outflow Relationship
Norris	$Y_t = 0.08X_t - 0.64Y_{t-1}$
Cherokee	$Y_t = 0.81X_t - 0.24X_{t-1} + 0.07Y_{t-2}$
Fontana	$Y_t = 0.59X_t - 0.21X_{t-1} + 0.07Y_{t-2}$
Douglas	$Y_t = 0.6X_t - 0.14X_{t-1} - 0.28Y_{t-1}$
Blue Ridge	$Y_t = 0.08X_t - 0.38Y_{t-1}$

5.1.2 Run-of-River Facilities

The CCF plots of the run of river facilities displays a significant difference from the storage facilities. The parameter estimates of individual facilities with the results of MLE methods are shown in Table 14. The relationship obtained between outflow and inflow are also summarized in Table 15. The information obtained from the transfer function relationship can be categorized as:

1. Past Information: In the run of river facilities, the relationship between outflow and its past values varies between the facilities under consideration. For instance, in all the projects except Chickamauga, the present-day outflow has a negative correlation with the past values, but the magnitude of this change is very small. For three projects, Watts Bar, Chickamauga, and Guntersville, the outflow at current day also depends on the value of inflow two days prior ($t-2$). Additionally, in Chickamauga dam, both the outflow and inflow the of day ($t-2$) has influence on the current day outflow.
2. Current Information: The current day outflow is significantly influenced by the equivalent inflow values, just like with storage facilities. Fort Loudon and Nickajack both exhibit a nearly identical outflow equals inflow relationship.

5.2 Discussion – Interpretation of Box Jenkins model results

The main goal of this thesis is to examine how well the Box Jenkins methodology categorizes various kinds of hydropower facilities. From the review of the different types of classification of hydropower facilities in section 1.1 it can be seen that the classification categories are not mutually exclusive. Storage facilities, for instance, can be used as a base load or peak load plant, and most large hydropower plants have a high head. Run-of-river facilities often fall under the base load category, however, certain plants with pondage water storage during off-peak periods and use this water during peak period to meet the hourly demand swings.

5.2.1 The intuition behind the transfer function coefficients: inflow and outflow dynamics

Hydrological conditions are always changing, which has a significant effect on hydropower operations. Water inflows-which may be natural inflows as well as discharges from upstream hydropower generation-and outflows-which include losses due to seepage and evaporation—have an impact on reservoirs.

Table 14: Diagnostic checks on parameter estimates- run-of-river facilities

Facility	Parameter Estimate	SE	*t*-Ratio	p-value
Watts Bar	-0.42	0.022	-18.41	<0.0001
	-0.07	0.018	-3.92	<0.0001
Chickamauga	0.13	0.05	2.61	<0.009
	0.52	0.03	14.84	0.0011
Guntersville	-0.24	0.017	-13.89	<0.0001
	-0.06	0.015	-3.93	<0.0001
Fort Loudon	-0.039	0.0059	168.86	<0.0001
Nickajack	-0.38	0.06	-6.44	<0.0001

Table 15: Inflow-Outflow relationship of run-of-river facilities

Facility	Inflow-Outflow Relationship
Watts Bar	$Y_t = 0.48X_t + 0.2X_{t-2} - 0.07Y_{t-1}$
Chickamauga	$Y_t = 0.44X_t - 0.05X_{t-2} + 0.52Y_{t-2}$
Guntersville	$Y_t = 0.6X_t + 0.14X_{t-2} - 0.06Y_{t-1}$
Fort Loudon	$Y_t = X_t - 0.039Y_{t-1}$
Nickajack	$Y_t = X_t + 0.008Y_{t-1}$

The amount of water in a reservoir restricts how much energy can be stored, so an understanding of both of these variables is crucial for proper reservoir operation and management operations in a hydropower project. Figure 24 illustrates the dependence between inflows and outflows derived from the transfer function modeling. The solid line represents the direct pathways of influence between the variable. In areas with regulated flows, the inflow on a given day is influenced by previous values because accurate and reliable forecasting of inflows is essential for the effective use of water resources and reservoir operation (Gragne, Sharma, Mehrotra, & Alfredsen, 2015). The dashed lines represent the influence of the variables and the lags discovered from the transfer function models developed. The different coefficients and their effects at various time lags are listed in Table 16. Regression analysis uses p-values and coefficients to determine the statistical significance of the correlation between the dependent and independent variables as well as the nature of the relationships. The coefficients describe the mathematical relationship between the variables and the sign (positive or negative) indicates whether there is a positive or negative correlation between the variables. The mean of the dependent variable increases as the value of the independent variable increases, according to a positive coefficient, whereas a negative coefficient suggests the contrary. The value of the coefficient represents how much the mean of the dependent variable changes when the independent variable is shifted by one unit while the other variables in the model remain constant. This is essential because it enables the user to evaluate each variable's impact independently of the others (Frost, 2019).

Reservoir outflow is the amount of water leaving per second by outlets, spillways, or water withdrawal. The volume of water that leaves the reservoir in any time period is referred to as withdrawal. The outflow is determined by the reservoir release schedule, which is composed of a number of regulations, rules, and guidelines approved by the water management authorities. The release schedule also includes data about the minimum outflow maintained, which is the minimal amount of water that must be released from the reservoir into a stream to satisfy the demands downstream (Votruba & Broža, 1989). The TVA's water control system, which consists of a network of connected dams and reservoirs on the Tennessee River and its tributaries, is seen in Figure 25. In all the ten facilities studied, outflows at time t are correlated to the corresponding inflows. However, each facility has different coefficients and dependencies at various time lags. Some potential reason for these outcomes are discussed below.

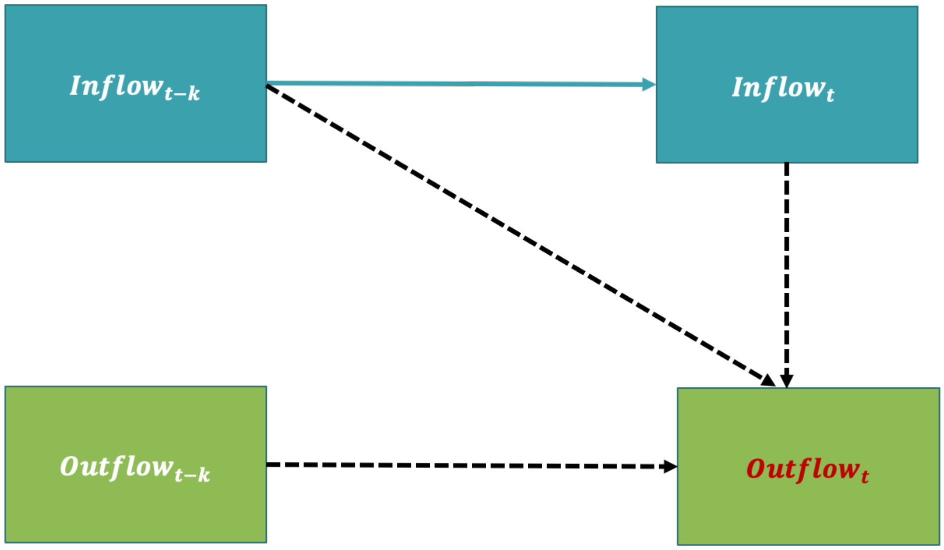


Figure 24: Inflow-Outflow dependance

Table 16: Effect of coefficients and lags

Facility		Contribution to $\mathbf{Outflow_t}$				
		Outflow		**Inflow**		
		$(\mathbf{t-1})$	$(\mathbf{t-2})$	**t**	$(\mathbf{t-1})$	$(\mathbf{t-2})$
Storage	Norris	⇓ 0.64		⇑ 0.08		
	Cherokee		⇑ 0.07	⇑ 0.81	⇓ 0.24	
	Fontana		⇑ 0.07	⇑ 0.59	⇓ 0.21	
	Douglas	⇓ 0.28		⇑ 0.60	⇓ 0.14	
	Blue Ridge	⇓ 0.28		⇑ 0.08		
Run of River	Watts Bar	⇓ 0.07		⇑ 0.48		⇑ 0.2
	Chickamauga		⇑ 0.52	⇑ 0.44		⇓ 0.05
	Guntersville	⇓ 0.06		⇑ 0.60		⇑ 0.14
	Fort Loudon	⇓ 0.039		⇑ 1		
	Nickajack	⇑ 0.008		⇑ 1		

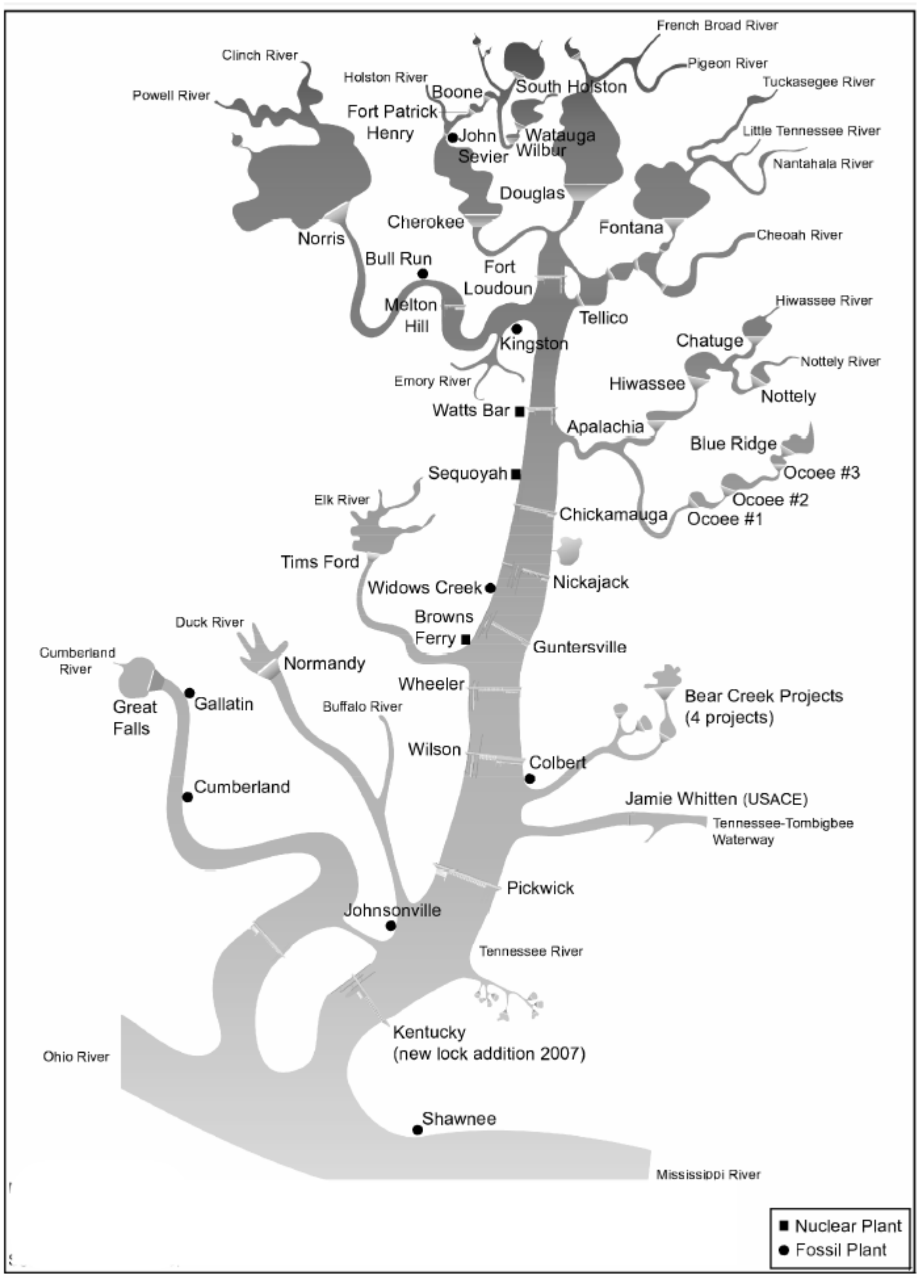


Figure 25: The TVA Water Control System

Adapted from Tennessee Valley Authority Reservoir Operations Study (Authority, 2004)

5.2.1.1 Storage Facilities

Fontana Dam with the largest capacity of the storage facilities under study, can generate 304 MW, whereas Blue Ridge is the smallest with 16 MW. When comparing the characteristics of coefficients of both Cherokee and Fontana, a similar pattern can be seen, with fairly similar coefficient values at the time lags. The same can be stated for Norris and Blue Ridge Dam while Douglas has a different trend. Let's now analyze each facility separately depending on its location and operational objectives inside the TVA water control system.

1. **Norris and Blue Ridge**: Located on the Clinch River basin, in addition to the different operational objectives, Norris dam also serves as a supply of cooling water to the Bull Run Steam Plant located 32 miles downstream. Additionally, this dam is 56.7 river miles upstream from Melton Hill Dam, the only TVA dam on a tributary stream with a navigation lock, (Tomljanovich, Strunk, & Oxendine, 1992). The Blue Ridge Dam is located on the Toccoa River in North Georgia, which flows northwest into Tennessee where it is called the Ocoee River. Additionally, it is the uppermost of the four Ocoee River dams that the TVA manages. The hydroelectric generating capacity of Norris is higher at 126 MW than that of Blue Ridge, which has a capacity of only 16 MW. According to the findings of transfer function modeling, the outflows at time t are dependent on the corresponding inflow as well as the outflow from the day before, with the prior outflow value having higher dependence (negative correlation) than the inflow. One possible explanation of this behavior would be the presence of dams downstream of these facilities. There are dams located downstream from both Norris and Blue Ridge where the discharges contribute to the inflow. For Melton Hill Dam, drainage from 2912 square miles of the watershed is regulated the Norris Dam(Tomljanovich et al., 1992) while outflows from Blue Ridge Dam make for 56 percent of the total inflow to the Ocoee no. 3 dam (Cox, 1990). This could account for the inverse correlation between the past outflow value with the current day's outflow.
2. **Cherokee and Fontana**: Cherokee Dam is located on the Holston River, and it is the largest (by storage volume) of the six dams located on the Holston River and its forks.

Boone, Fort Patrick Henry, and South Holston Dams are located on the South Fork Holston River while Watauga and Wilbur Dam are located on Watauga River. Having a flood storage capacity of 747,400 acre-ft., during a year with normal rainfall, the water level in Cherokee reservoir varies about 30 ft. from summer to winter to provide seasonal flood storage (*Status of Cherokee Reservoir*, 1990). Fontana Dam is located on the Little Tennessee River and is the uppermost of the five dams on the Little Tennessee River: Cheoah Dam downstream, followed by Calderwood Dam, Chilhowee Dam, and Tellico Dam. Fontana dam controls the reservoir levels of Chilhowee, Calderwood and Cheoah all of which operate in a "modified run-of-river mode" in which the inflow and outflow from the facilities balance out daily (Sale, Hall, & Keil, 2016). The Tellico Dam has no hydroelectric facilities and is designed only for storage. Fontana has a total seasonal flood control storage of 771,200 acre-ft. and gives a high degree of control flood control (*Water resources appraisal for hydroelectric licensing: Little Tennessee River Basin, Tennessee, North Carolina, and Georgia. Appraisal report*, 1981). The two dams are components of basin wide multiple reservoir system, and the results of transfer function modeling shows a positive correlation with current day inflow and the outflow two days prior and an inverse correlation with the past day inflow. The outflow value two days prior are disregarded for the time being as its correlation is smaller when compared to the inflow factors. Both these reservoirs are part of multi-reservoir system which offers a substantial degree of flood control. Because of this, the outflow for the current day is heavily dependent on the inflows, which could also explain why there is a negative correlation between the outflow and the inflow from the previous day.

3. **Douglas Dam**: Douglas Dam is the only TVA reservoir located on the Lower Broad River Basin in East Tennessee. A combination of the two scenarios discussed above is seen in the results for Douglas. The outflows at time t are dependent on the corresponding inflow as well as the outflow and inflows from the day before. The Douglas Dam serves as a flood reservoir for the Tennessee River downstream in addition French Broad River. Additionally, the downstream Fort Loudon Dam, which operates in the run-of-river mode, is impacted by the releases from Douglas as well.

5.2.1.2 Run-of-River Facilities

Run-of-river facilities are those where there is no long-term water storage and where you would anticipate that outflows would be dependent only on the inflows. The outcomes of the Box Jenkins approach suggest otherwise. In some under study, correlation at time lags also appears to be significant, even if outflow strongly depends on the corresponding inflow values. Since all these facilities are located on the mainstem of the Tennessee river, it might be challenging to distinguish the distinctive differences between their operations. In this case, it would make more sense to assess facilities based on their location in the mainstem rather than categorizing them based on the outcomes of transfer function modeling.

1. **Fort Loudon Dam**: Fort Loudon is the uppermost in the chain of nine TVA operated reservoirs that form a continuous navigable channel from Tennessee to Kentucky. The French Broad River, Little River, and Holston River are three major tributaries that are flowing into the reservoir (Anderson, 1984). From the transfer function modeling, outflows at time t are correlated to inflows on a 1:1 basis in addition to a small, but potentially unimportant, negative correlation to the outflow from the day before.
2. **Watts Bar Dam:** Located on the Tennessee River, the Watts Bar Dam extends 72.4 miles northeast from the dam to Fort Loudon Dam. Watts Bar Lake receives unregulated inflows from the 1,790 square mile local drainage region in addition to releases from Melton Hill and Fort Loudon Dam. The outflow at time t are dependent on the corresponding inflow in addition to the outflows from the day before and inflows two days prior. The outflows have positive correlation to the inflow values while a negative correlation is observed for the past outflow observations.
3. **Chickamauga Dam:** Chickamauga Dam is located in the Tennessee river, 7 miles above Chattanooga. It maintains a navigation channel approximately 59 miles up to the river to Watts Bar Dam and along the Hiwassee River to Charleston, Tennessee. Between the Chickamauga and Nickajack Reservoirs, the dam contains one lock that is 60 feet wide by 360 feet long and can lower barges up to 50 feet (Authority, 2017a). Cooling water for the Watts Bar nuclear reactor, which is situated on the west side of the reservoir, flows from the dam through the plant intake channel to the intake pump station.(No, 2011). The results from transfer function are different compared to other runoff facilities studied. The outflow

at time t are dependent on the corresponding inflow as well as the outflow from two days earlier, with the past outflow value having higher positive correlation than the inflow. As it has a lower correlation than the other two, the negative correlation with prior inflows is temporarily disregarded.

4. **Nickajack Dam:** Located in Southeastern Tennessee, Nickajack Dam was built in 1967 to replace Hales Bar Dam. The dam impounds the Nickajack Lake and feeds into the Guntersville Lake. Between these two lakes, a 600 by 110-foot auxiliary that can lift or lower nine big barges at a time serves Nickajack Dam. The Raccoon Mountain Pumped Storage project is situated adjacent to the Nickajack Reservoir, and during times of low electricity demand, water is pumped from the dam at the base of Raccoon Mountain to the reservoir constructed at the top. (Authority, 2017b). From the transfer function modeling, outflows at time t are correlated to inflows on a 1:1 basis and a small, but potentially insignificant positive correlation to the outflow from the day before.
5. **Guntersville Dam:** Guntersville dam is located in Marshall County in Alabama. It impounds the Guntersville Lake, and the releases are fed into the Wheeler Lake(Authority, 2001). About 37,200 cfs of the inflows into Guntersville Reservoir comes from releases from Nickajack Dam which accounts for almost 89 percent of the annual discharge. The transfer function has comparable characteristics to Watts Bar Dam in which the outflows have positive correlation to the inflow values and a negative correlation to the past outflow observations.

These run-of-river facilities are difficult to categorize based on the results of transfer function modeling, unlike the storage facilities. But the findings lead to the following observations:

a. Watts Bar and Guntersville, as well as the dams upstream of them (Fort Loudon and Nickajack), exhibit comparable features in transfer function modeling. Additionally, it has been observed that releases from the upstream dams make up the majority of the inflows into these reservoirs.
b. The primary flow into Chickamauga Dam is from Watts Bar located upstream and as previously indicated, the results of Chickamauga Dam differ from the rest of the facilities. The Watts Bar Nuclear Power Plant uses Chickamauga Dam as a supply of cooling water;

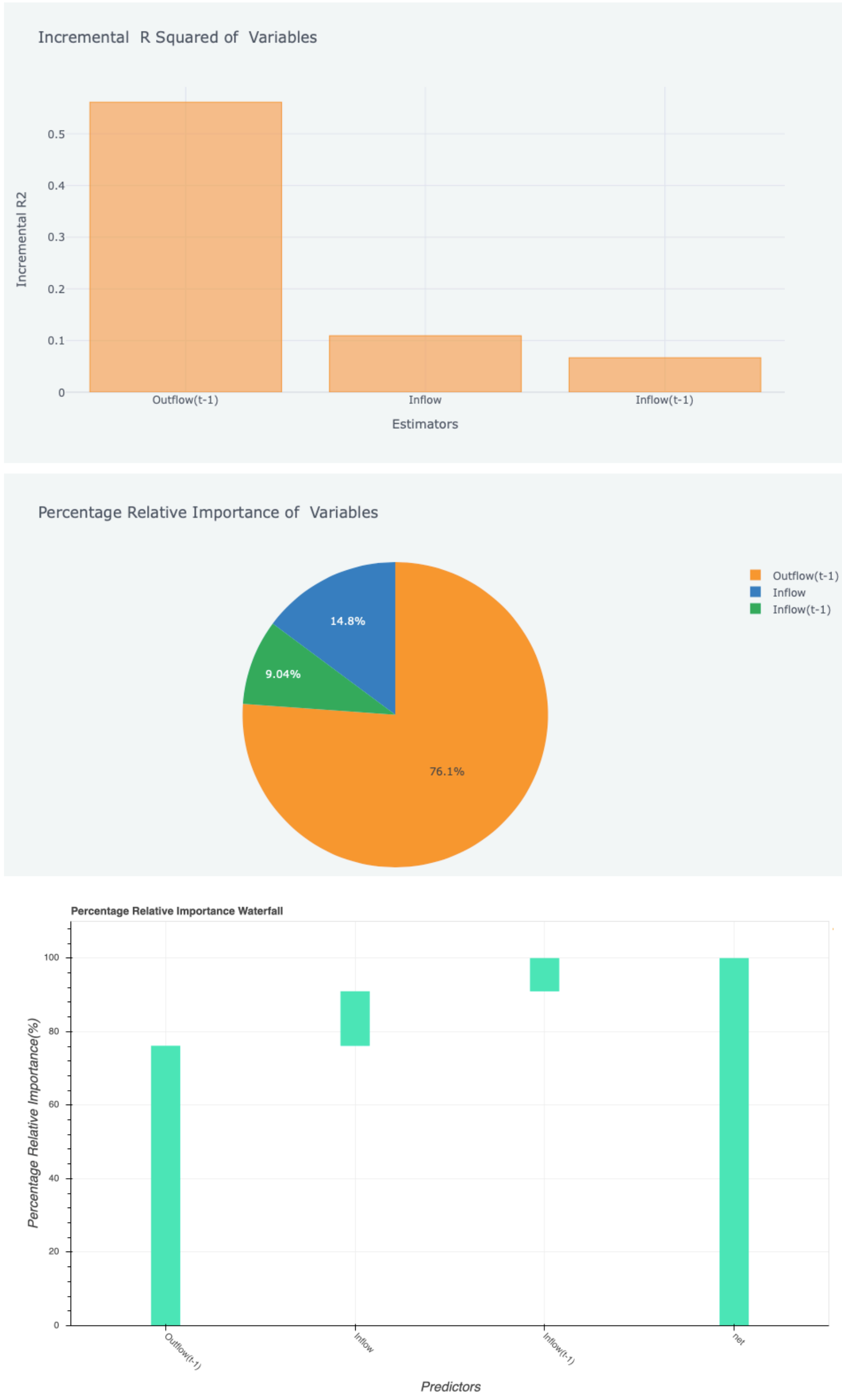
further research is needed to determine whether any attempts have been made to control the flows in order to maintain the ideal reservoir temperature for cooling purposes.

5.2.2 Comparing transfer function coefficients – Dominance Analysis

The relative significance of the independent should be considered whenever multiple regression is used to test and compare models. According to the Box Jenkins models' findings, outflows are related to the equivalent inflows as well as the inflows and outflows from the past. By analyzing the various contributing factors to outflows, some potential reasons of the behavior were examined. It is important to keep in mind that, the many variables influencing the outflows are also inter-correlated and, as a result, cannot be observed separately. Take the example of Douglas Dam, the outflows at t are correlated to inflows as well as outflows at lag $t-1$ which are correlated with each other. As a result, understanding the relative importance of each variable gives reservoir operators a more effective tool to use in operational planning, policy making, and other processes. Dominance analysis is a commonly used technique in statistical models to determine the importance of independent variable (Budescu, 1993). Dominance statistics can be divided into four different types of measures:

- Individual Dominance: Individual dominance provides the variability in the independent (predictor) variable alone on the absence of other variables. Mathematically it is the R^2 vale of the model between the dependent and predictor variable
- Interactional Dominance: When all other predictors are present, the interactional dominance is the impact of variability expressed by a given predictor variable. It is the difference between the R^2 vale of the overall model and the R^2 value of the model calculated without the specified predictor variable.
- Average Partial Dominance: This can be seen as the average impact a predictor has in all combinations with other predictors, excluding the combination in which all predictors are accessible.
- Total Dominance: By averaging all of the conditional values, total dominance compiles the additional contributions of each predictor to all subset models.

The results of dominance analysis conducted on Douglas Dam is shown as an example in Figure 26. The *dominance analysis* package in Python is utilized for finding the relative importance of the different independent variables.



	Interactional Dominance	Individual Dominance	Average Partial Dominance	Total Dominance	Percentage Relative Importance
Outflow(t-1)	0.537093	0.650556	0.495999	0.561216	76.1371
Inflow	0.0853801	0.198384	0.044057	0.109274	14.8246
Inflow(t-1)	0.046204	0.152258	0.001406	0.0666228	9.03835

Figure 26: Results of Dominance Analysis- Douglas Dam

The percentage relative importance computed for all the ten facilities are summarized in Table 17 and 18. In all the storage facilities studied, past outflows were observed to be the dominant predictor when comparing the percentage relative importance values. In run-of-river facilities, current day inflow is observed to have the highest percentage of relative importance. In contrast to storage facilities, where the difference in relative importance was quite noticeable, the past inflow and outflow values are comparable to the current inflow in run-or-river facilities. Determining the relative importance of the different dependence variables obtained from the Box Jenkins analysis allows the practitioners to decide which variable to pay closer attention to. For instance, the results of dominance analysis for Douglas Dam reveals that previous day outflows have the highest dominance over the other inflow values, despite the fact that Box Jenkins results suggested a strong positive correlation to the current day inflow. Furthermore, it should be highlighted that dominance analysis is only used in this study as an additional concept that adds value to the Box Jenkins technique, and that additional research is required to determine the true benefit of combining the outcomes of the Box Jenkins approach and dominance analysis.

5.2.3 Practical application of this research

Reservoir operation frequently involves complex and undocumented decision processes. Decisions made by reservoir operators are based on available hydrologic information, such as historical outflows, reservoir water level, and forecasted inflows, which have a significant impact on the regulated outflows from a reservoir (Chen et al., 2018). The volume of water released from a reservoir is thus determined by the expertise of the reservoir operators, and reservoir simulation models are frequently used to estimate these releases. Many sophisticated reservoir simulation models, such RiverWare and WRAP to mention a couple, were developed and have since gained the favor of numerous governmental organizations and fleet owners. These reservoir models, however, are only valuable if the operating laws or policies included in the simulation could accurately reflect the actual operation. It was also commonly accepted that system operators regularly deviate from the operating guidelines in order to respond to particular circumstances or constraints that can arise at different times (Oliveira & Loucks, 1997). The difference in representation of hydropower in both RiverWare and PLEXOS were briefly discussed in the section 2.3.

Table 17: Relative importance: run-of-river facilities

Facility	Percentage Relative Importance (%)			
	Inflow $(\mathbf{t})$	**Inflow** $(\mathbf{t-2})$	**Outflow** $(\mathbf{t-1})$	**Outflow** $(\mathbf{t-2})$
Watts Bar	38.6	24.5	36.8	
Chickamauga	44.6	25.4		30
Guntersville	47	20	33	
Fort Loudon	66		33.2	
Nickajack	68.2		31.8	

Table 18: Relative importance: storage facilities

Facility	Percentage Relative Importance (%)			
	Inflow$(\mathbf{t})$	**Inflow** $(\mathbf{t-1})$	**Outflow** $(\mathbf{t-1})$	**Outflow** $(\mathbf{t-2})$
Norris	16.8		83.2	
Cherokee	35.8	17.5		46.8
Fontana	29.6	13.6		56.8
Douglas	14.8	9.04	76.1	
Blue Ridge	8.96		91	

Deriving transparent reservoir policies that are based on both hydrological condition and other decision variables is therefore necessary in order to understand the actual release decisions. The novel aspect of this work is the incorporation of statistical techniques in an attempt to address the crucial research problem of characterizing reservoir operation under the ever-evolving grid conditions and the ensuing decisions. The use of Box Jenkins models has also been a focus of this thesis, with the emphasis on time series representation rather than forecasting, which is what they are well known for. We begin our examination of Box-Jenkins procedure by applying it to the time series data on inflow and outflow and tested its ability understand the factors contributing to actual release decisions. To accomplish this task, this study posed the following research questions.

a) Can different hydropower facilities be classified using the Box Jenkins methodology proposed?
b) What information/insights gained from the transfer function model developed might be transferable to other fleets?

Although emphasis in this study was on the value of Box Jenkins approach, the findings revealed a number of additional factors that were found to be helpful in improving the representation of hydropower generation within various reservoir models. Figure 27 below is an extension of Figure 24 with the dashed lines representing the relationship derived from transfer function modeling. By considering the magnitude and direction of the coefficients, the transfer function model developed evaluates the effects of various factors contributing to the reservoir outflow.

1. *Relevance of past outflows*

   Rivers used to generate hydropower typically exhibit significantly higher day-to-day and intraday flow changes, more frequent and faster than those that define free-flowing rivers. The hydraulic parameters including water level, flow rate, water quality, and river morphology also change along with changes in flow regimes, having a substantial impact on the downstream watershed (Bejarano, Sordo-Ward, Alonso, & Nilsson, 2017). Consider reservoir operator making the outflow decision yesterday. In fact, yesterday, they also had the next-day forecasts of the different hydrological parameters. As a result, the operator made the outflow decision by naturally taking forecasting and the state of the reservoir into account, demonstrating that historical outflow data contains much more than just what is initially obvious.

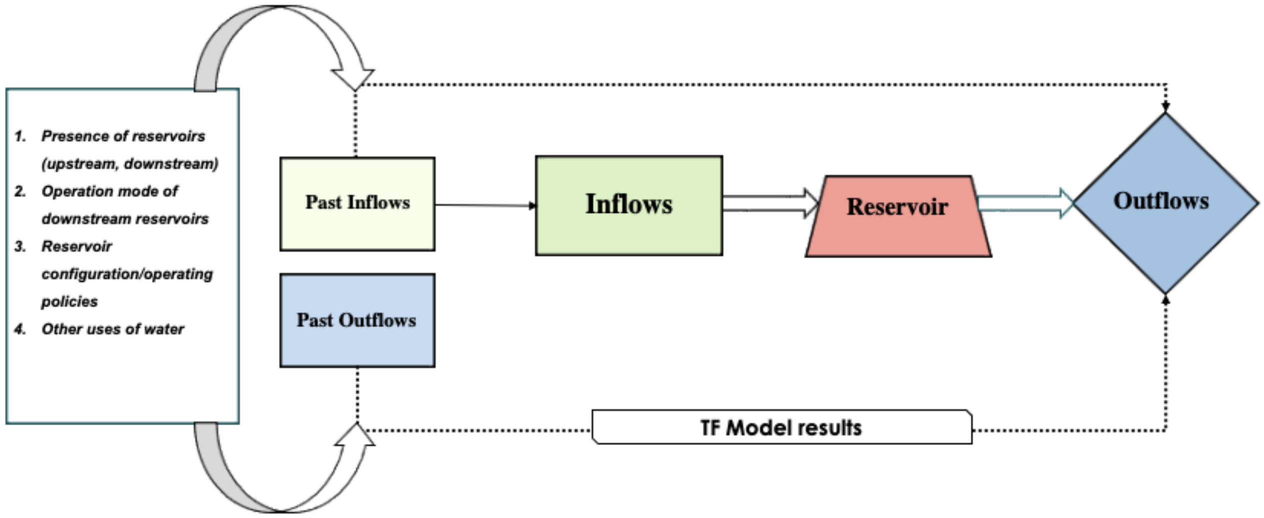


Figure 27: Factors significant for model interpretation

Additionally, as seen with the example of Norris and Blue Ride, the presence of reservoirs downstream also influences the outflows to be dependent on past values. The results of dominance analysis also suggest a significance reliance on past outflow values in storage facilities.

2. *Relevance of past inflows*

   Reservoir operators must be aware of how reservoir inflows change under various hydrological conditions in order to operate them efficiently and the priority of operating objectives varies by reservoirs and regions. Consider reservoir operator making the outflow decision today. The timing and magnitude of outflows are based on the changes in the timing and volume of inflows which was forecasted the day before. This illustrates that the outflow decision is indirectly dependent on the past inflows. In reservoirs built largely to regulate flood flows, that goal take precedence when determining outflows, and past inflows have a significant impact on decisions in such circumstances.

5.2.4 Limitations and scope for future studies

It should be emphasized that the results from five storage and five run-of-river facilities should not be viewed as an exhaustive evaluation of Box Jenkins models for characterizing hydropower operation. These results should only be interpreted within the context of the following qualifying factors:

a) All the facilities owned and operated by a single utility
b) The particular set of input-output data used in the model (reservoir inflows and outflows)
c) The arbitrary daily-time step with a weekly seasonality chosen for analysis
d) Modified inflow values to account for negative results encountered during computation

This research was presented solely as a preliminary step in investigating the feasibility of Box-Jenkins models in characterizing hydropower facilities. The list of qualifying factors makes it evident that there are a variety of modifications of the present analysis that could be investigated. The results of this study show that it is possible to produce Box-Jenkins models to assess reservoir operation. Box Jenkins models have high computational complexity, and the model used in this study was created for daily flows, making analysis of hourly or 5-minute flows time consuming. Applying the Box Jenkins methodology is also subjective, and the researcher's expertise and experience may influence how reliable the model they choose is.

For this research, a prior arrangement was already in place with the TVA, making it possible to obtain the fleet data. The very sensitive nature of fleet data, however, prevents all utilities from sharing their operational data. Would data availability be a challenge if this methodology is to be made transferable?

All the reservoirs in the analyzed for this study are operated by the TVA, whose ownership extends to all supply chain levels, including generating, transmission, and distribution. Future research must be done using fleet data from reservoirs serving in ISO/RTO regions. Not only will this address the performance of the methodology but will assist identifying any additional limitations that affect decisions about outflow that are outside the control of reservoir operators. Lastly, this study develops a single input single output transfer function (SISO) model. Box Jenkins technique, however, is able to handle multiple input variables, and for subsequent research, other elements impacting inflow including temperature and precipitation might be modeled.

5.3 Summary

The transformation between inflows and outflows in five storage and five run-of-river facilities operated by the TVA was modeled using a Box Jenkins methodology. The major results obtained from this pilot data-driven methodology may be summarized as follows:

- The correlation between outflows and inflows demonstrates that present-day outflows are significantly dependent on previous values addition to the corresponding inflows.
- When transfer function coefficients were compared, certain facilities exhibited coefficients that were comparable to one another. Facilities with comparable transfer function patterns were examined to determine the root of the resemblance.
- From the perspective of an individual facility, it was discovered that the following factors were significant interpreting the model results: location and operation mode of downstream reservoirs, release pattern of the upstream reservoirs and their contribution to the inflows, other relevant uses of water (such as flood control, cooling water source etc.)

Chapter 6: Conclusions and Recommendations

6.1 Outcomes from Box Jenkins representation

Hydrologic information in the form of inflows, storages, and discharges are typically included in models for reservoir operations. However, in practice, reservoir operators might use a combination of any of them, or even all of them, depending on the circumstances. As management of reservoirs frequently reply on complex and unrecorded decision-making procedures, representation of hydropower in energy system models is challenging. Let's look at three key stakeholders: a reservoir operator, a power system dispatcher, and lastly a downstream water user, in order to comprehend the relevance of the transfer function model developed.

From the reservoir operator’s point of view, the Box Jenkins model developed requires incorporation of their knowledge of facility operation, and they are able to correct/justify their actions by analyzing the results. Reservoir operators need hydrologic data to make outflow decisions, and the results of transfer function modeling imply that particular attention should be paid not just to the previous inflow values but also to the past outflow values.

The power system dispatchers are responsible for ensuring the steady operation of the power system. The dispatchers should regularly contact with the reservoir operators, who are responsible for the facility operation, including regulation schedules, transmission and generation constraints and emergencies. Even if there is an unforeseen communication issues with the operators, the transfer function model developed and the different factors contributing the model, can still aid dispatchers in understanding how the facility operates.

From the perspective of downstream water users, the regulated outflows from an upstream reservoir are strongly influenced by the operators' actions rather than natural inflow process. As a result, in order to build effective water management systems, the water users downstream must be aware of how the upstream reservoirs function. Without the need for complex models, the Box Jenkins model may provide the necessary physical interpretation, allowing downstream water consumers to comprehend the operating characteristics of the upstream reservoirs.

6.2 Research Contributions

The following is a summary of the contributions made by the research presented in this thesis:

- The three-step iterative Box Jenkins methodological framework was developed to serve as a guide in achieving the objectives of this study. Specifically, to understand how a transfer function model may act as a tool for parametrizing hydropower.
- The real-world situation was next investigated by utilizing the fleet data from ten TVA operated facilities to estimate degree of the transferability of the methodology.
- The transfer function model generated was used to qualitatively analyze and characterize the operations of the various facilities under study. The contributing aspects important for interpreting model results are also discussed.
- Dominance analysis was introduced to add value to the Box Jenkins model results and provide different stakeholders with an additional set of concepts to convey the functionality of hydropower.

6.2.1 Recommendations for further research

The following recommendations are made for additional research in this area:

1. Development of seasonal transfer function: The Box Jenkins model created for this research includes flows for 12 years between 2004 and 2016. The next stage would be to create distinct transfer function models for the dry, normal, and wet multi-year segments, and then compare the variations in outcomes.
2. Developing multi-input models: The Box Jenkins methodology can handle multiple inputs and subsequent research can include factors influencing the inflow such as temperature and water quality as inputs. The differences in the outcomes of the transfer function model might then be examined.
3. Feature importance of the transfer function coefficients: It's critical to determine which coefficients in the transfer function model are more significant if the Box Jenkins methodology is to be scaled. A dominance analysis is presented to compare the relative importance of the contributing factors. For further research, machine learning methods like feature importance or could be used to compare the relative contributions of each variable by collecting additional data over more years.

4. Conversion to frequency domain: The dynamic regression model can be analyzed in the frequency domain which involves the frequency analysis of ARMA models. Because spectral approaches are generally easier to interpret, analyze, and understand and can almost completely reveal the dynamics of the system, they are generally more useful than time domain methods.

References

Adamowski, J. F. (2008). River flow forecasting using wavelet and cross-wavelet transform models. Hydrological Processes: An International Journal, 22(25), 4877-4891.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. Czaki, Akademiai Kiado, Budapest.

Aldeyab, M., Harbarth, S., Vernaz, N., Kearney, P., Scott, M., Elhajji, F., . . . McElnay, J. (2011). The impact of antibiotic use on the incidence and resistance pattern of extended-spectrum beta-lactamase-producing bacteria in primary and secondary healthcare settings. British journal of clinical pharmacology, 74, 171-179. doi:10.1111/j.1365-2125.2011.04161.x

Alemu, E. T., Palmer, R. N., Polebitski, A., & Meaker, B. (2011). Decision support system for optimizing reservoir operations using ensemble streamflow predictions. Journal of Water Resources Planning and Management, 137(1), 72-82.

Anderson, D. W. (1984). RESERVOIR MANAGEMENT PLANNING: AN ALTERNATIVE TO REMEDIAL ACTION. Lake and Reservoir Management, 1(1), 423-426.

ARIMA Modelling. Retrieved from https://www.jmp.com/en_ch/learning-library/topics/time-series/arima-modeling.html

Authority, T. V. Status of Norris R'eservoir.

Authority, T. V. (2001). Guntersville Reservoir- Final Environmental Impact Statement and Reservoir Land Management Plan. Retrieved from https://tva-azr-eastus-cdn-ep-tvawcm-prd.azureedge.net/cdn-tvawcma/docs/default-source/environment/land-management/reservoir-land-management-plans/guntersville/guntersville-reservoir-environmental-review.pdf?sfvrsn=9d66d89_2

Authority, T. V. (2004). Programmatic Environmental Impact Statement: Tennessee Valley Authority Reservoir Operations Study Record of Decision. Retrieved from

Authority, T. V. (2017a). Chickamauga Reservoir- Final Reservoir Land Management Plan. Retrieved from https://tva-azr-eastus-cdn-ep-tvawcm-prd.azureedge.net/cdn-tvawcma/docs/default-source/default-document-library/site-content/environment/environmental-stewardship/land-management/land-plans/2017-updates/02_chickamauga-rlmp_final_vol-ii.pdf?sfvrsn=b5a1b82a_0

Authority, T. V. (2017b). Nickajack Reservoir - Final Reservoir Land Management Plan. Retrieved from https://tva-azr-eastus-cdn-ep-tvawcm-prd.azureedge.net/cdn-tvawcma/docs/default-source/default-document-library/site-content/environment/environmental-stewardship/land-management/land-plans/2017-updates/06_nickajack_rlmp_final_vol_vi.pdf?sfvrsn=23054daf_0

Authority, T. V. (2021). Norris Operating Guide. Retrieved from https://www.tva.com/environment/lake-levels/norris/operating-guide

Authority, T. V. (2022). Norris Dam. Retrieved from https://www.tva.com/energy/our-power-system/hydroelectric/norris

Bain, D. M., & Acker, T. L. (2018). Hydropower Impacts on Electrical System Production Costs in the Southwest United States. Energies, 11(2), 368. Retrieved from https://www.mdpi.com/1996-1073/11/2/368

Bejarano, M. D., Sordo-Ward, Á., Alonso, C., & Nilsson, C. (2017). Characterizing effects of hydropower plants on sub-daily flow regimes. Journal of hydrology, 550, 186-200.

Biddle, S. H. (2001). Optimizing the TVA reservoir system using RiverWare. Paper presented at the Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges.

Bisgaard, S., & Kulahci, M. (2011). Time series analysis and forecasting by example: John Wiley & Sons.

Boslaugh, S. (2012). Statistics in a nutshell: A desktop quick reference: " O'Reilly Media, Inc.".

Box, G. E., & Jenkins, G. M. (1976). Time series analysis. Forecasting and control. Holden-Day Series in Time Series Analysis.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control: Wiley.

Bracmort, K., Stern, C. V., & Vann, A. (2013). Hydropower: Federal and nonfederal investment. In: Congressional Research Service.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. Psychological bulletin, 114(3), 542.

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2), 121-167. doi:10.1023/A:1009715923555

Chang, D.-F. (2018). Effects of higher education expansion on gender parity: a 65-year trajectory in Taiwan. Higher Education, 76(3), 449-466. doi:10.1007/s10734-017-0219-9

Chen, K., Guo, S., He, S., Xu, T., Zhong, Y., & Sun, S. (2018). The value of hydrologic information in reservoir outflow decision-making. Water, 10(10), 1372.

Chow, V., Maidment, D., & Mays, L. (1988). 1988, Applied hydrology:(New York: McGraw-Hill).

Cools, M., Moons, E., & Wets, G. (2009). Investigating the Variability in Daily Traffic Counts through use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site Locations. Transportation Research Record, 2136(1), 57-66. doi:10.3141/2136-07

Cotter, J., Hydraulic Engineer, F., District, W., & Zagona, E. INTEGRATION OF RIVERWARE INTO THE CORPS WATER MANAGEMENT SYSTEM.

Cox, J. P. (1990). Water resources review: Ocoee reservoirs, 1990. Retrieved from

Exemplar, E. (2022). PLEXOS® Simulation Software. URL: https://energyexemplar. com/products/plexos-simulation-software/, last accessed on, 25(12), 2022.

Frost, J. (2019). Regression analysis: An intuitive guide for using and interpreting linear models: Statisics By Jim Publishing.

Goel, M., Jain, S. K., Rani, D., & Chalisgaonkar, D. (2018). Methodology to deal with computed negative inflows in Indian reservoirs.

Gragne, A. S., Sharma, A., Mehrotra, R., & Alfredsen, K. (2015). Improving real-time inflow forecasting into hydropower reservoirs through a complementary modelling framework. Hydrology and Earth System Sciences, 19(8), 3695-3714.

Gulliver, J. S., & Arndt, R. E. (1991). Hydropower engineering handbook: McGraw-Hill, Inc.

Harvey, A. C., & Pierse, R. G. (1984). Estimating Missing Observations in Economic Time Series. Journal of the American Statistical Association, 79(385), 125-131. doi:10.1080/01621459.1984.10477074

Hirth, L. (2016). The benefits of flexibility: The value of wind energy with hydropower. Applied Energy, 181, 210-223.

Huang, F., & Wu, B. (2014). Using the Transfer Function Model in Analyzing and Forecasting Students' Studying Achievement. Journal of Business and Economics, 5, 2052-2056. doi:10.15341/jbe(2155-7950)/11.05.2014/009

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice: OTexts.

Ibanez, E., Magee, T., Clement, M., Brinkman, G., Milligan, M., & Zagona, E. (2014). Enhancing hydropower modeling in variable generation integration studies. Energy, 74, 518-528.

Ingram, M. (2019). Hydropower Value and Opportunities in Future Grid States. Paper presented at the Conference: Presented at the 2019 CEATI Hydropower and Stations Conference, 19-20 March 2019, Tucson, Arizona, United States. https://www.osti.gov/biblio/1504620 https://www.osti.gov/servlets/purl/1504620

Jakaša, T., Andročec, I., & Sprčić, P. (2011). Electricity price forecasting—ARIMA model approach. Paper presented at the 2011 8th International Conference on the European Energy Market (EEM).

Johnson, K., Hadjerioua, B., & Martinez, R. U. (2015). Small hydropower in the United States. ORNL/TM-2015/326, Oak Ridge, TN: Oak Ridge National Laboratory.

Krispin, R. (2019). Hands-On Time Series Analysis with R: Perform time series analysis and forecasting using R: Packt Publishing.

Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review, 7(3), 21. doi:10.1007/s12544-015-0170-8

Labadie, J. W. (2006). MODSIM: decision support system for integrated river basin management.

Liu, L.-M. (1991). Dynamic relationship analysis of us gasoline and crude oil prices. Journal of Forecasting, 10(5), 521-547. doi:https://doi.org/10.1002/for.3980100506

Love, B. C. (2002). Comparing supervised and unsupervised category learning. Psychonomic bulletin & review, 9(4), 829-835.

Lumia, R., & Moore, R. (1983). Computation of inflows and outflows of eight regulated lakes in the Oswego River Basin, New York, 1930-79 (Vol. 82): US Geological Survey.

Majumder, M., & Ghosh, S. (2013). Decision making algorithms for hydro-power plant location: Springer.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). Forecasting methods and applications: John wiley & sons.

McGinnis, H. (1994). Determining the Impact of Economic Factors on Local Government Growth Policy: Using Time-series Analysis and Transfer Function Models. Urban Studies, 31(2), 233-246. doi:10.1080/00420989420080221

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? Science, 319(5863), 573-574.

Monnet, D., López-Lozano, J.-M., Campillos, P., Burgos, A., Yagüe, A., & Gonzalo, N. (2001). Making sense of antimicrobial use and resistance surveillance data: application of ARIMA and transfer function models. Clinical microbiology and infection, 7, 29-36.

Moran, E. F., Lopez, M. C., Moore, N., Müller, N., & Hyndman, D. W. (2018). Sustainable hydropower in the 21st century. Proceedings of the National Academy of Sciences, 115(47), 11891-11898.

Nathalie Voisin, S. T., Konstantinos Oikonomou , Hongyan Li ,Jinxiang Zhu. (2021). Weekly hydropower datasets to support reliability studies over the Western Interconnect. Retrieved from https://www.wecc.org/Administrative/Voisen-WECC_B1_2021-07-20.pdf

Nekooie, B. (2018). Hydro and pump storage modelling in ABB Ability™ PROMOD® Benchmark report. Retrieved from https://library.e.abb.com/public/2bb611278afe4d8bac8555addc262bfb/Hydro-Pump-Storage-Modeling-PROMOD_9AKK107046A5365-A4.pdf

Nihan, N. L., & Holmesland, K. O. (1980). Use of the box and Jenkins time series technique in traffic forecasting. Transportation, 9(2), 125-143. doi:10.1007/BF00167127

No, N. D. (2011). Unit 2 Engineering Manager7.

Nogales, F. J., & Conejo, A. J. (2006). Electricity price forecasting through transfer function models. Journal of the Operational Research Society, 57(4), 350-356. doi:10.1057/palgrave.jors.2601995

O'connor, P., Saulsbury, B., Hadjerioua, B., Smith, B. T., Bevelhimer, M., Pracheil, B. M., . . . Uria Martinez, R. (2016). Hydropower Vision A New Chapter for America's 1st Renewable Electricity Source. Retrieved from

Oliveira, R., & Loucks, D. P. (1997). Operating rules for multireservoir systems. Water Resources Research, 33(4), 839-852.

Pankratz, A. (2009). Forecasting with Univariate Box - Jenkins Models: Concepts and Cases: Wiley.

Pankratz, A. (2012). Forecasting with Dynamic Regression Models: Wiley.

Papadopoulos, C., Johnson, R., Valdebenito, F., & Exemplar, E. (2014). PLEXOS® integrated energy modelling around the globe. ed: Energy Exemplar, 10.

Parsons, D. G., & Colbourne, E. B. (2000). Forecasting Fishery Performance for Northern Shrimp (<i>Pandalus borealis</i>) on the Labrador Shelf (NAFO Divisions 2HJ). Journal of Northwest Atlantic Fishery Science, 27, 11-20. doi:10.2960/j.v27.a2

Pereira-Cardenal, S. J., Mo, B., Riegels, N. D., Arnbjerg-Nielsen, K., & Bauer-Gottwein, P. (2015). Optimization of multipurpose reservoir systems using power market models. Journal of Water Resources Planning and Management, 141(8), 04014100.

Petri, M. C. (2009). National power grid simulation capability: need and issues. Retrieved from

Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., . . . Westerhoff, R. (2013). Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting. Bulletin of the American Meteorological Society, 94(6), 776-785.

Ragavan, K., & Satish, L. (2007). Closure on “An Efficient Method to Compute Transfer Function of a Transformer From Its Equivalent Circuit”. Power Delivery, IEEE Transactions on, 22, 1261-1262. doi:10.1109/TPWRD.2007.893940

RiverWare. Retrieved from https://www.colorado.edu/cadswes/creative-works/riverware

Sale, M., Hall, D., & Keil, J. (2016). Low Impact Hydropower Institute Certification Handbook. In: Harrington Park, NJ: Low Impact Hydropower Institute.

SAS, S., & ETS, R. (2014). 13.2 User’s guide the ARIMA procedure. SAS Institute Inc., Cary, NC, USA, 7.

Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. International Journal of Advanced Research in Artificial Intelligence, 2(2), 34-38.

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461-464, 464. Retrieved from https://doi.org/10.1214/aos/1176344136

Singh, V. P., & Frevert, D. K. (2010). Watershed models: CRC press.

Skjelbred, H. I. (2020). The role of AI in Hydropower Optimization. Retrieved from https://www.ntnu.no/documents/1269211504/0/9.+Session1_Skjelbred%2C+Hans+Ivar_Role+of+AI.pdf/070c22e6-6b25-16e4-d642-c19e539b00b7?t=1581586931541

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics, 10(1), 3-22.

Status of Cherokee Reservoir. (1990). Retrieved from United States: https://www.osti.gov/biblio/6284528

https://www.osti.gov/servlets/purl/6284528

Stoll, B., Andrade, J., Cohen, S., Brinkman, G., & Brancucci Martinez-Anido, C. (2017). Hydropower modeling challenges. Retrieved from

Survey, U. S. G. (2016). National Water Information System data available on the World Wide Web (USGS Water Data for the Nation). Retrieved from http://waterdata.usgs.gov/nwis/

Tomljanovich, D. A., Strunk, J. W., & Oxendine, L. B. (1992). Water resources review: Melton Hill Reservoir, 1992. Retrieved from

Turner, S. W., & Voisin, N. (2022). Simulation of hydropower at subcontinental to global scales: a state-of-the-art review. Environmental Research Letters.

Uria-Martinez, R., Johnson, M. M., Shan, R., Samu, N. M., Oladosu, G., Werble, J. M., & Battey, H. (2021). US Hydropower market report. Retrieved from

USGS NWIS, G. S. N. W. I. S. (2010). USGS surface-water data for the nation. In.

Veselka, T., Ploussard, Q., & Christian, M. (2020). Historical Hydropower Operations and Economic Value. Retrieved from

Veselka, T. D. (2009). Estimating Colorado River Storage Project Power Economics with the GTMax Model Retrieved from http://www.riversimulator.org/Resources/GCDAMP/AMP1999toPresent/TWG/2009/2009_12/Attach_10.pdf

Voisin, N., Bain, D., Macknick, J., & O'Neil, R. S. (2020). Improving Hydropower Representation in Power System Models - Report Summary of PNNL-NREL Technical Workshop Held March 6-7, 2019 in Salt Lake City, UT (PNNL-29878 United States 10.2172/1726280 PNNL English). Retrieved from https://www.osti.gov/servlets/purl/1726280

Voisin, N., Kintner-Meyer, M., Skaggs, R., Nguyen, T., Wu, D., Dirks, J., . . . Hejazi, M. (2016). Vulnerability of the US western electric grid to hydro-climatological conditions: how bad can it get? Energy, 115, 1-12.

Votruba, L., & Broža, V. (1989). Water management in reservoirs: Elsevier.

Waage, M. D., Baldwin, C. K., Steger, R. G., & Bray, T. J. (2001). Incorporating seasonal stream flow forecasts into operational decision making. Paper presented at the Proceedings of the 16th Annual Western Snow Conference, Sun Valley, Idaho.

Wang Fa-Qiang, M. X.-K. (2013). Transfer function modeling and analysis of the open-loop Buck converter using the fractional calculus. Chin. Phys. B, 22(3), 30506-030506. doi:10.1088/1674-1056/22/3/030506

Water resources appraisal for hydroelectric licensing: Little Tennessee River Basin, Tennessee, North Carolina, and Georgia. Appraisal report. (1981). Retrieved from United States: https://www.osti.gov/biblio/6551134

https://www.osti.gov/servlets/purl/6551134

Winter, T. C. (1981). Uncertainties in estimating the water balance of lakes 1. JAWRA Journal of the American Water Resources Association, 17(1), 82-115.

Wurbs, R. A. (2005). Comparative evaluation of generalized river/reservoir system models.

Yan, X., & Su, X. (2009). Linear regression analysis: theory and computing: World Scientific.

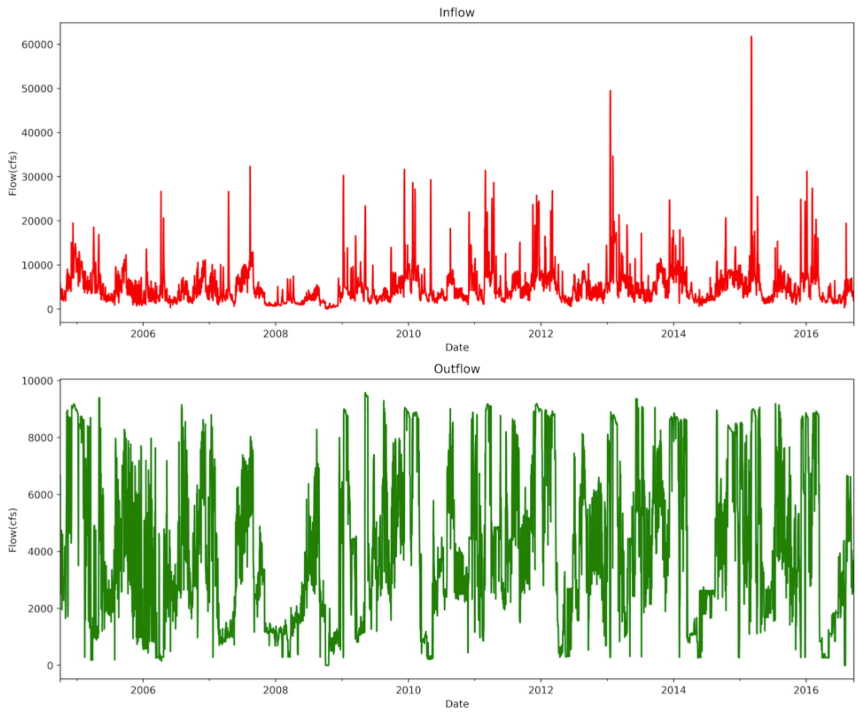
Zagona, E. A., & Magee, T. M. (1999). Modeling hydropower in RiverWare. In Waterpower'99: Hydro's Future: Technology, Markets, and Policy (pp. 1-10).

Zhong, M., & Sharma, S. (2006). Matching Hourly, Daily, and Monthly Traffic Patterns to Estimate Missing Volume Data. Transportation Research Record, 1957(1), 32-42. doi:10.1177/0361198106195700106

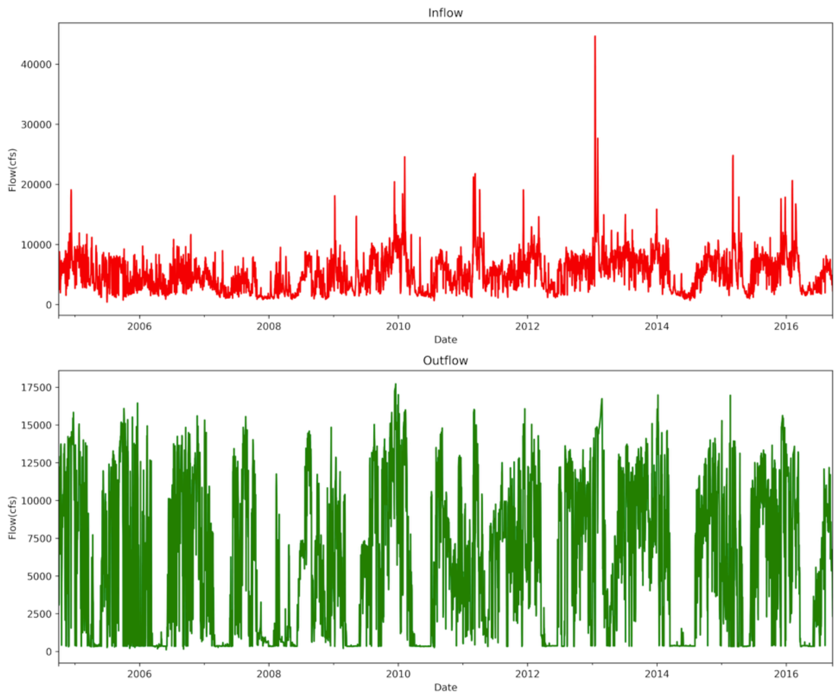
Appendix A: Time series plots

Storage Facilities

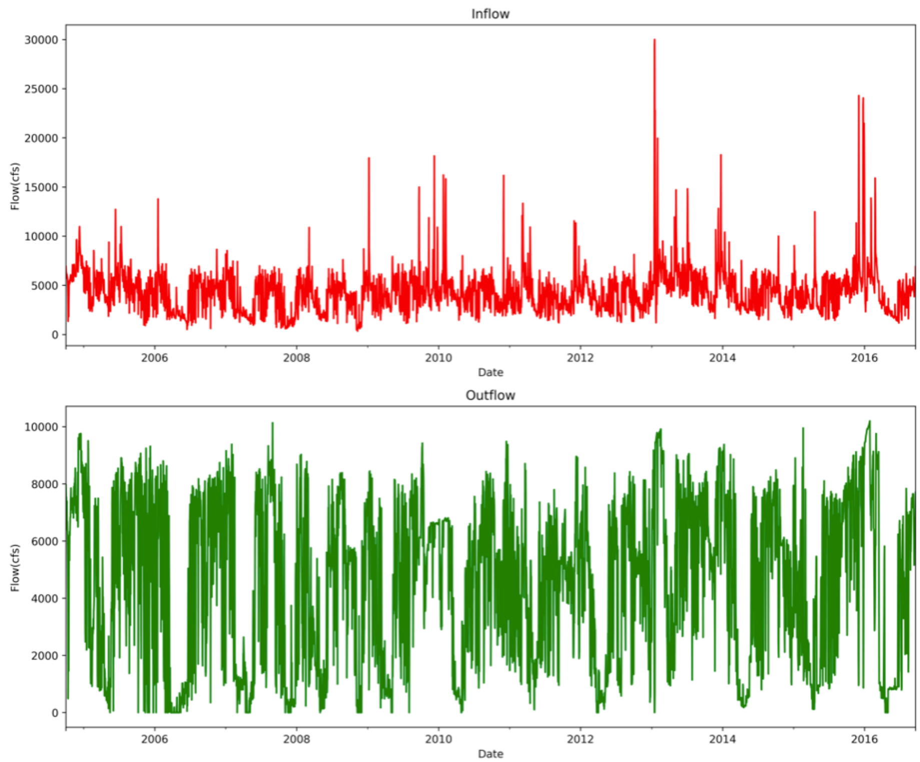
1. Norris Dam



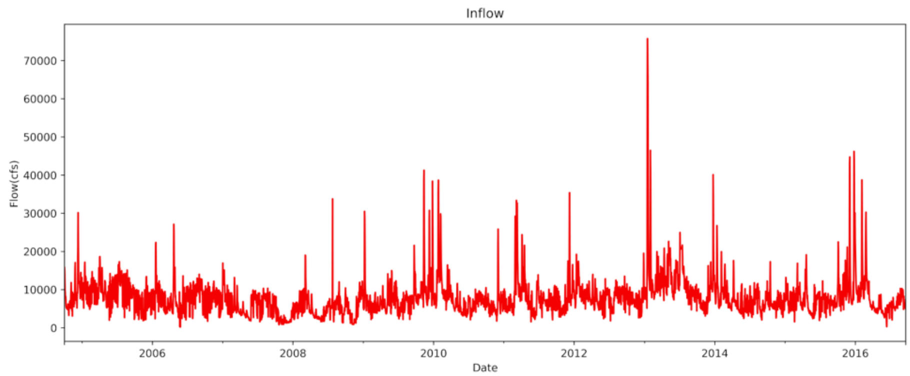
2. Cherokee Dam

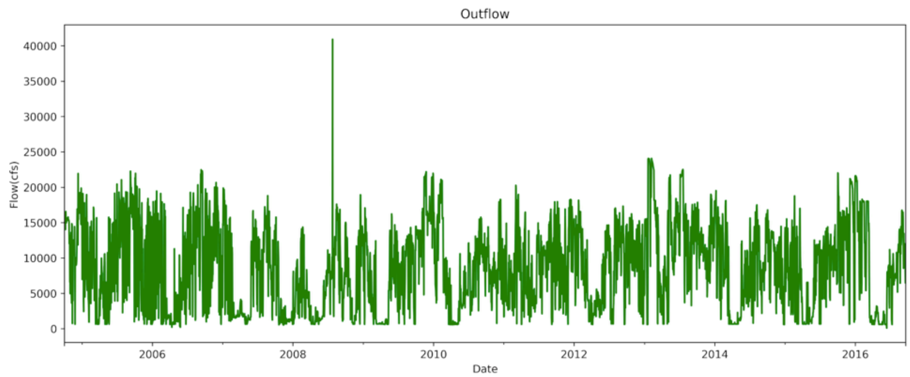


3. Fontana Dam

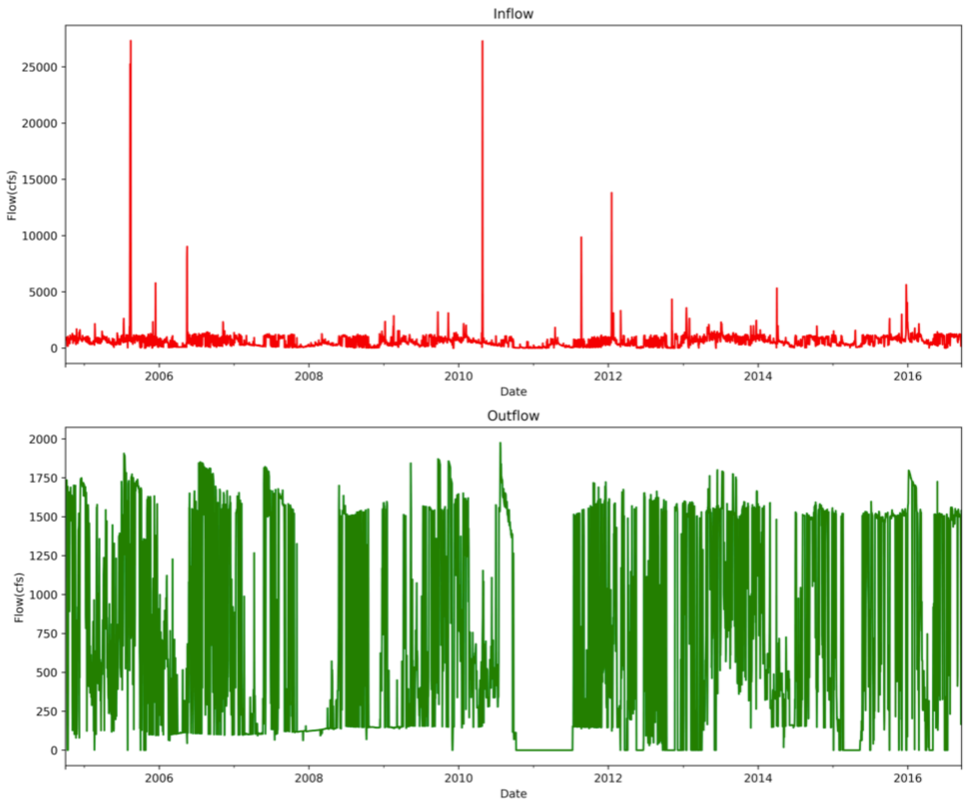


4. Douglas Dam



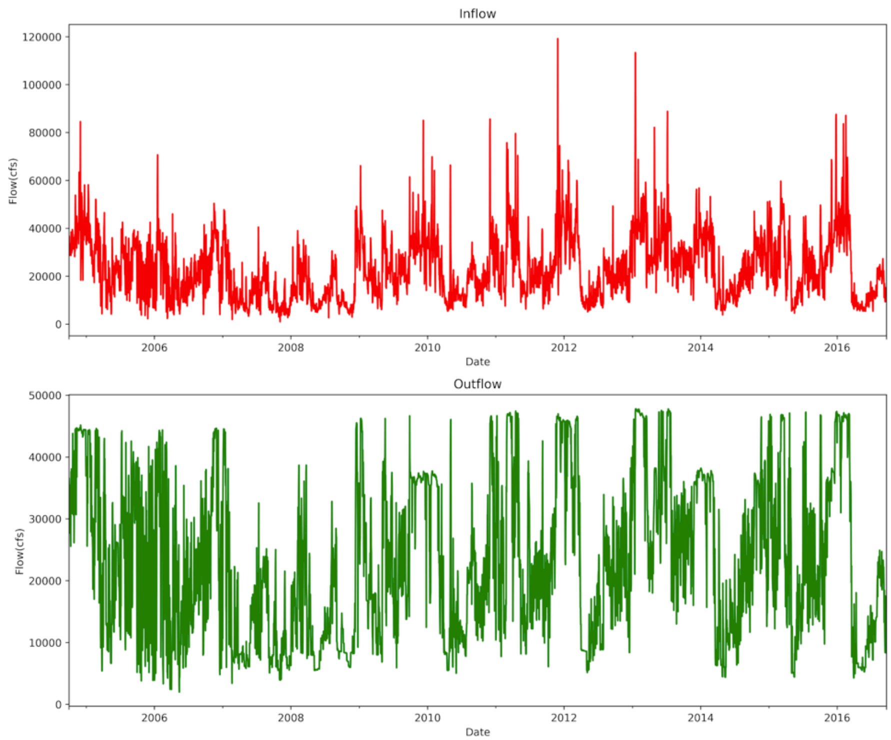


5. Blue Ridge Dam

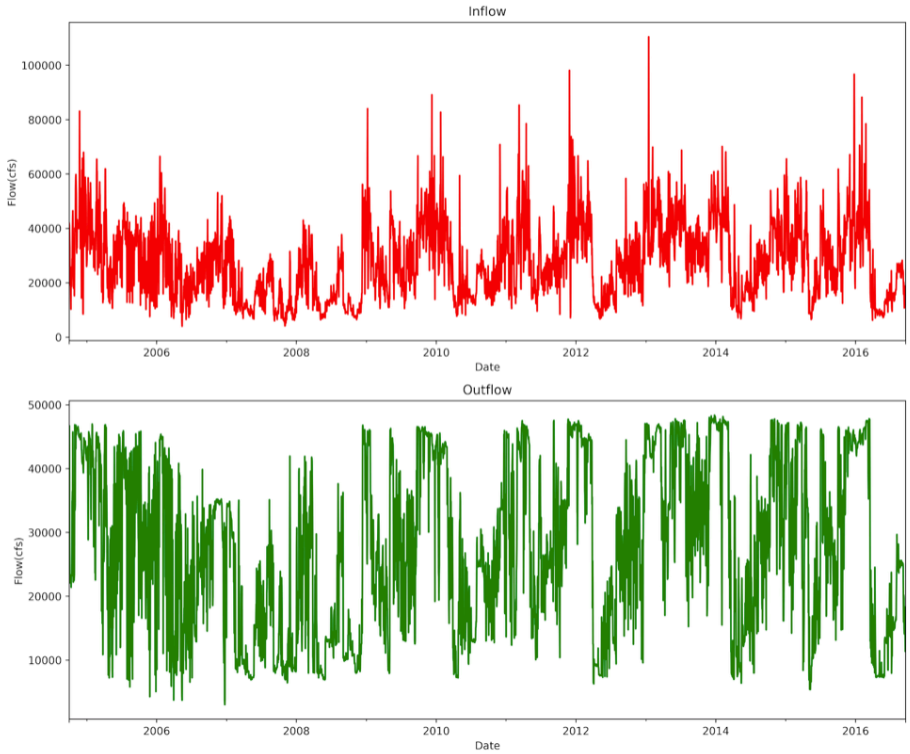


Run-of-river facilities

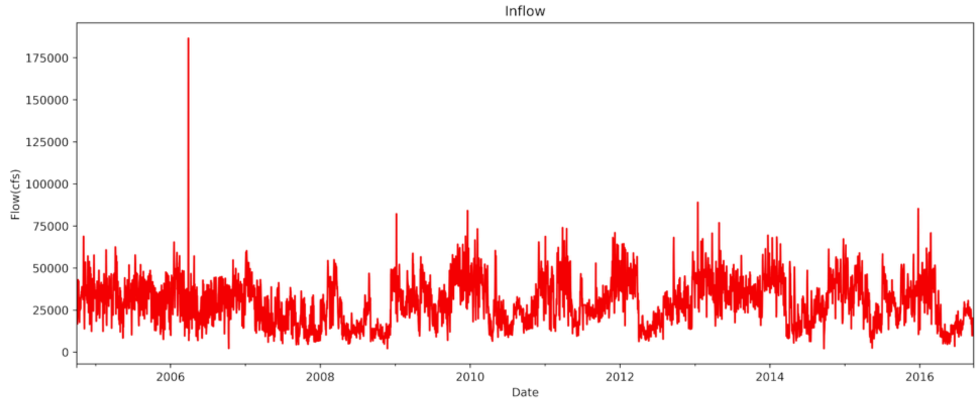
1. **Watts Bar Dam**

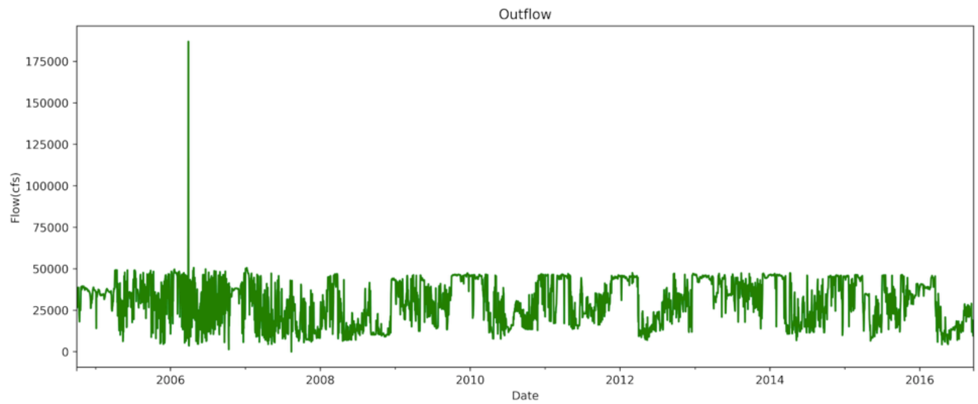


2. Chickamauga Dam

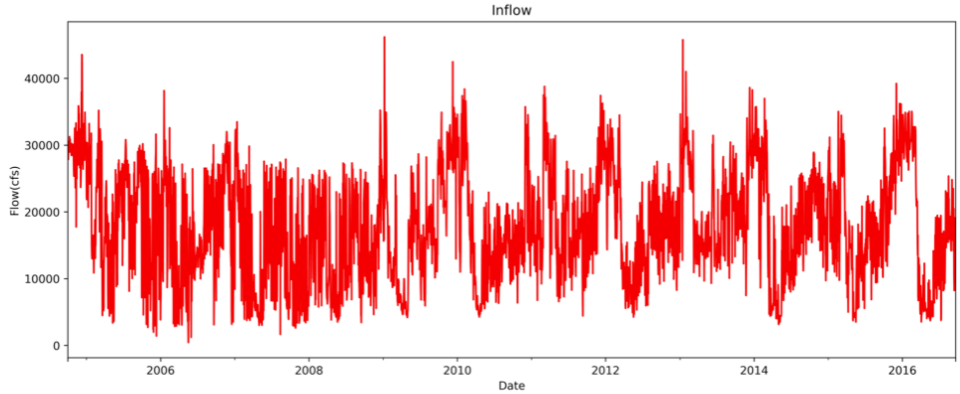


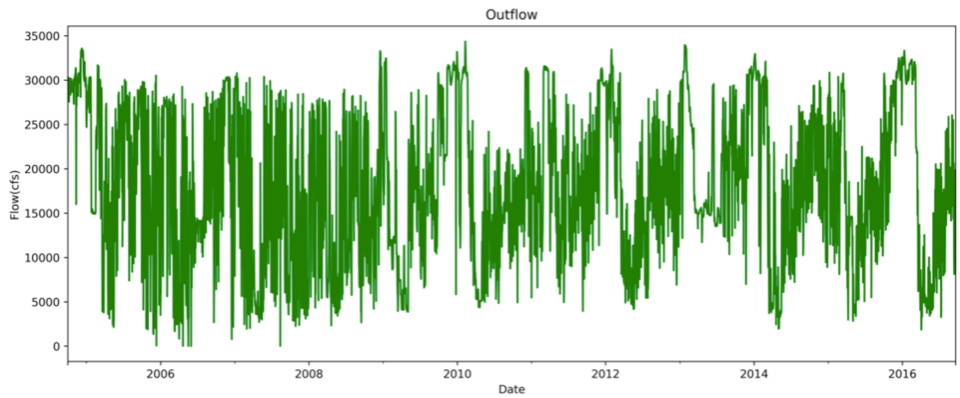
3. Guntersville Dam



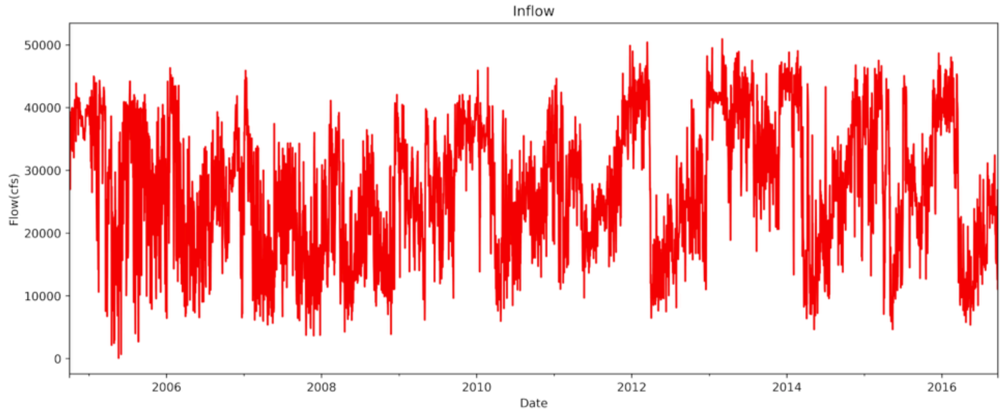


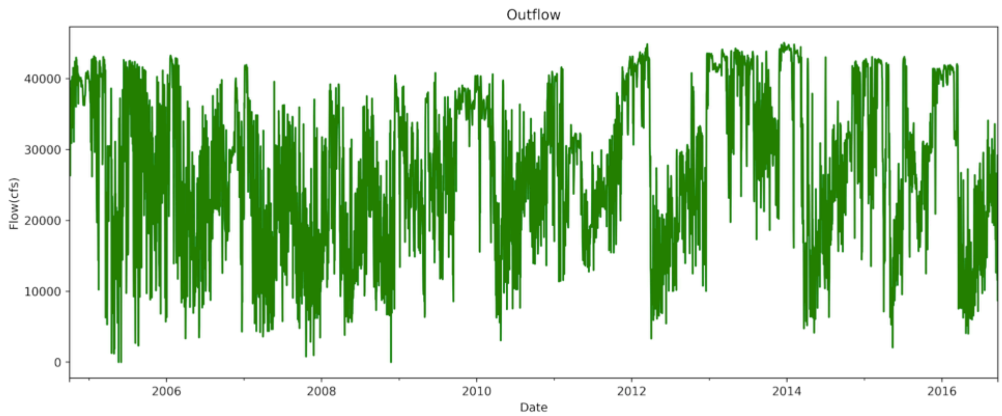
4. Fort Loudon Dam





5. Nickajack Dam

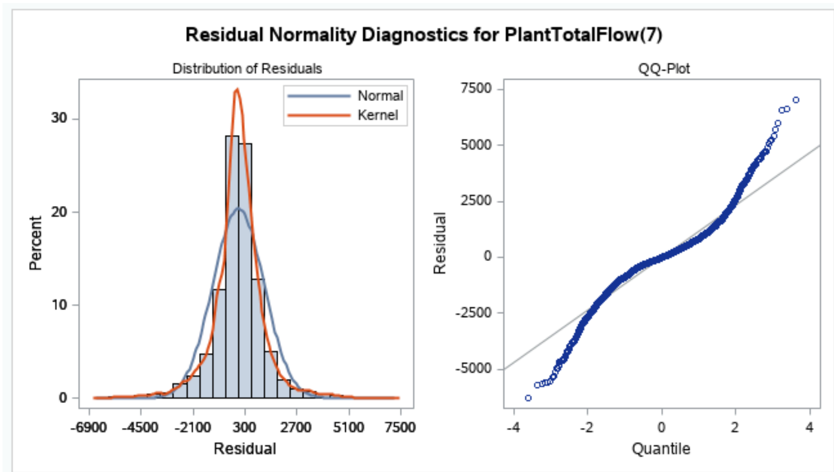


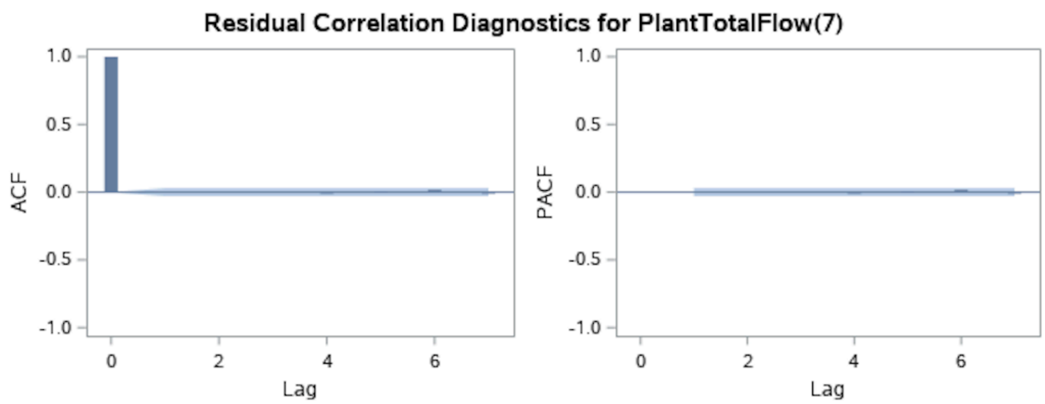


Appendix B: Residual analysis plots

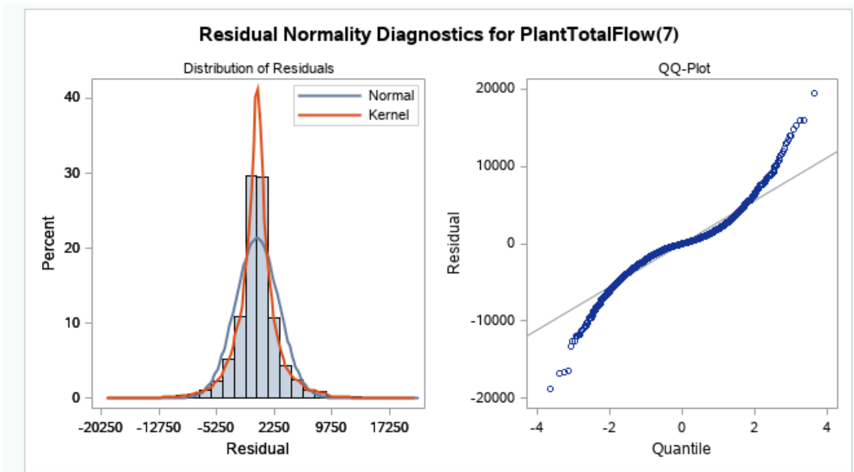
Storage Facilities

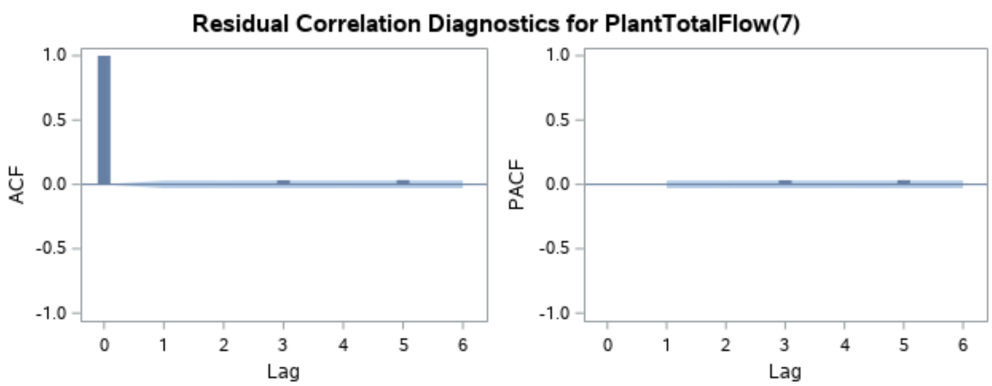
1. Norris Dam



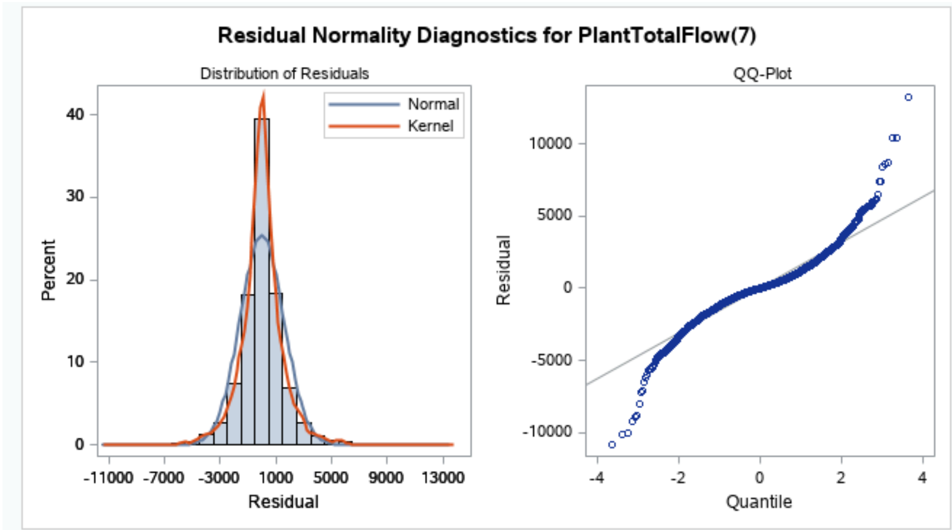


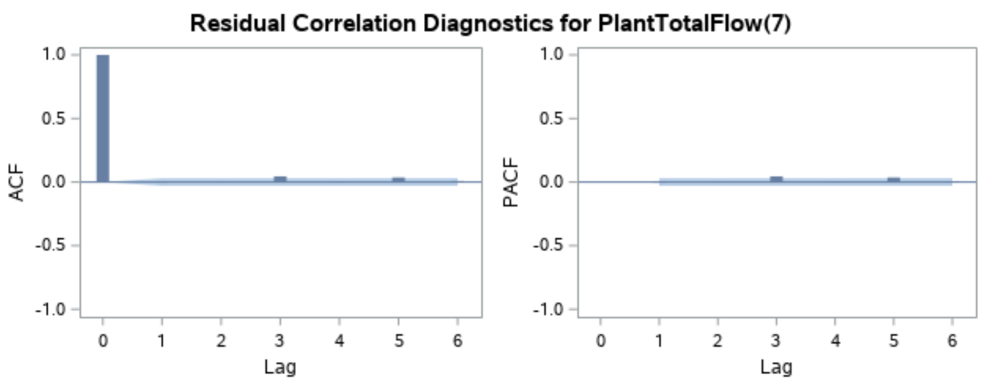
2. Cherokee Dam



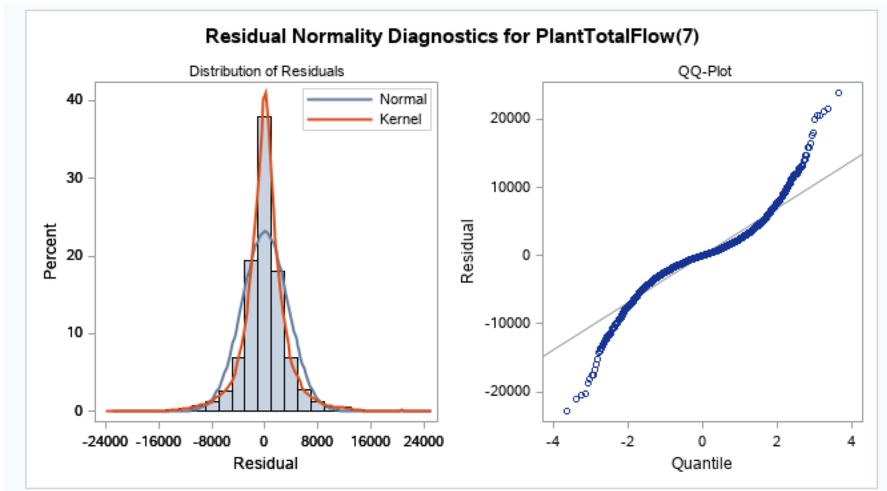


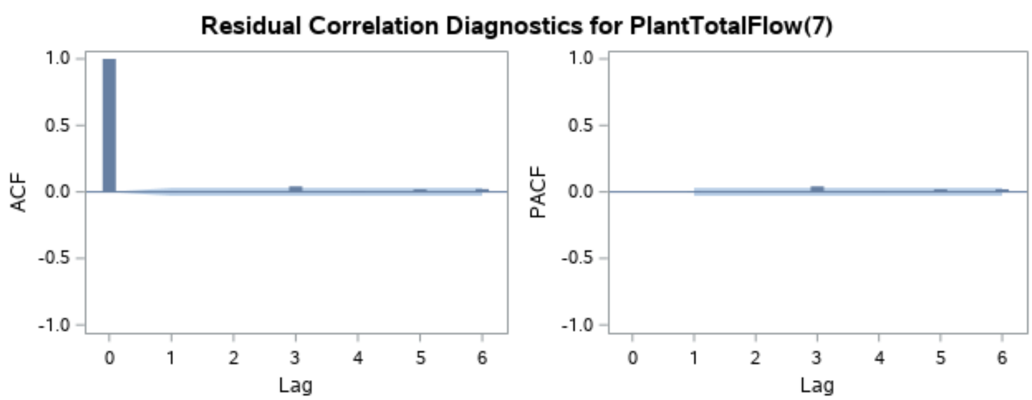
3. Fontana Dam



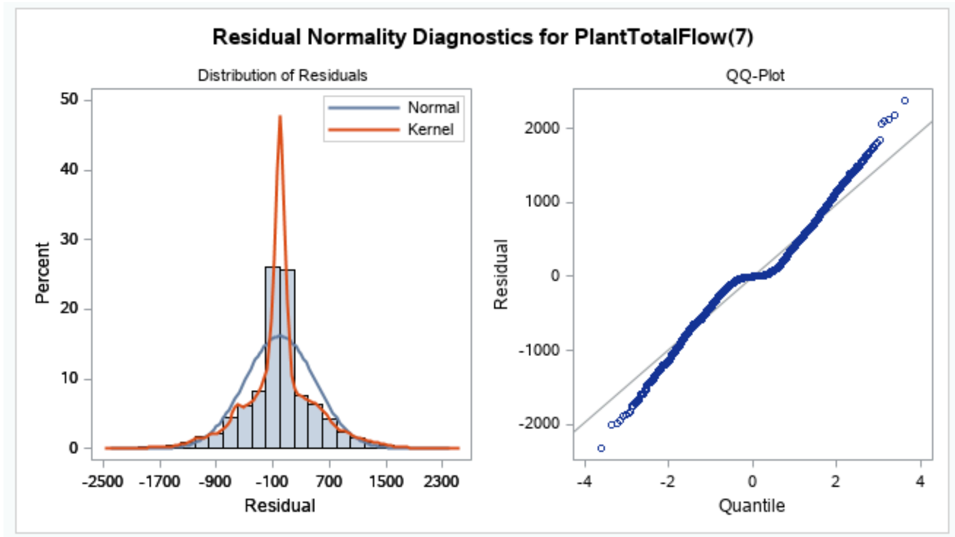


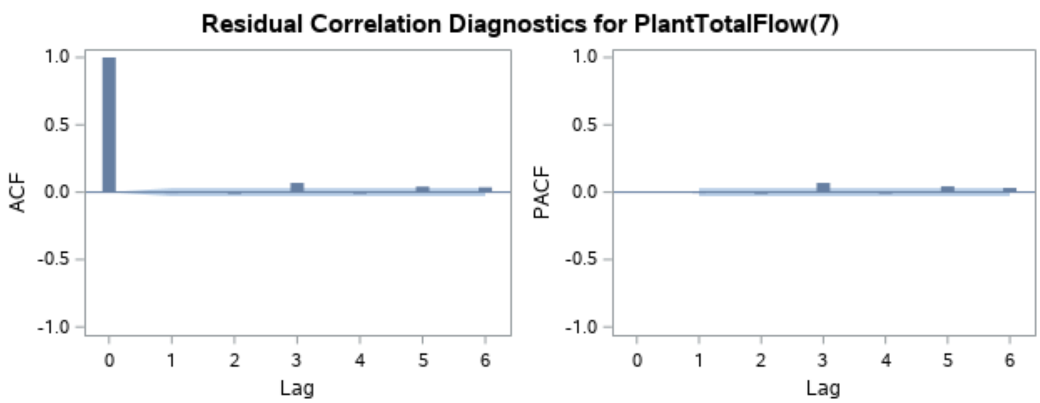
4. Douglas Dam





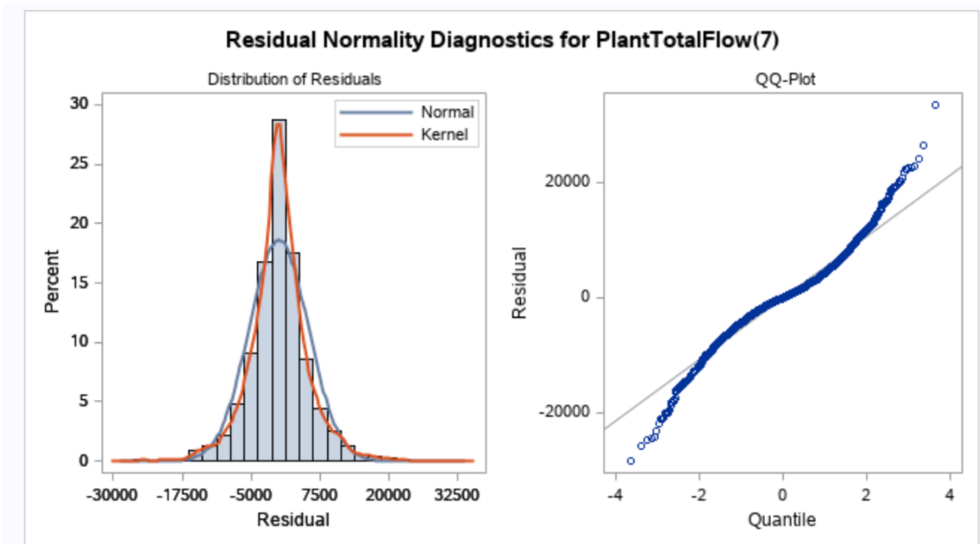
5. Blue Ridge Dam

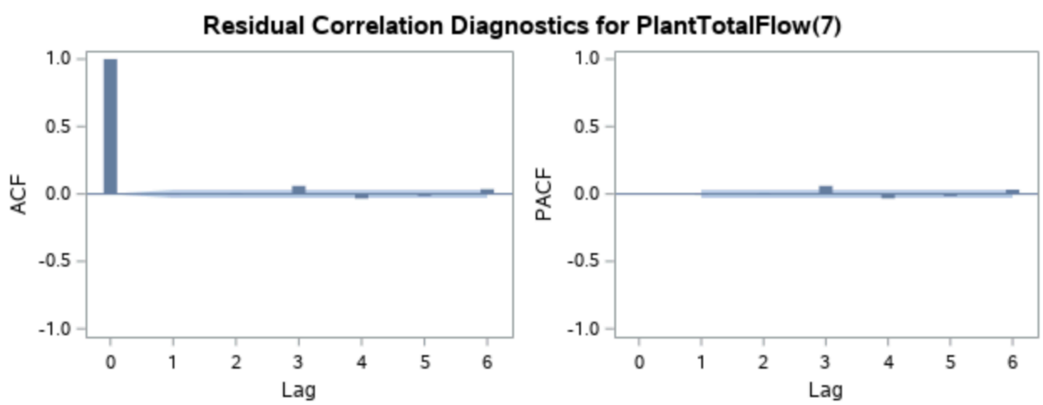




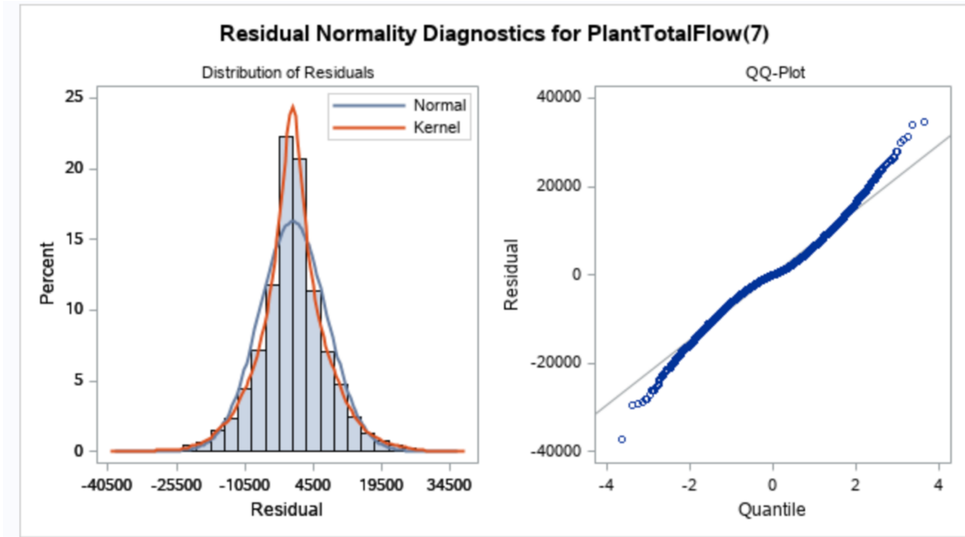
Run-of-river facilities

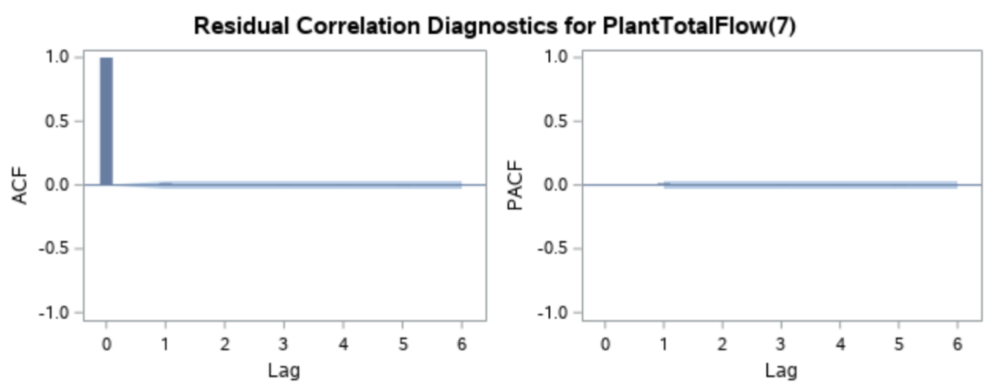
1. **Watts Bar Dam**



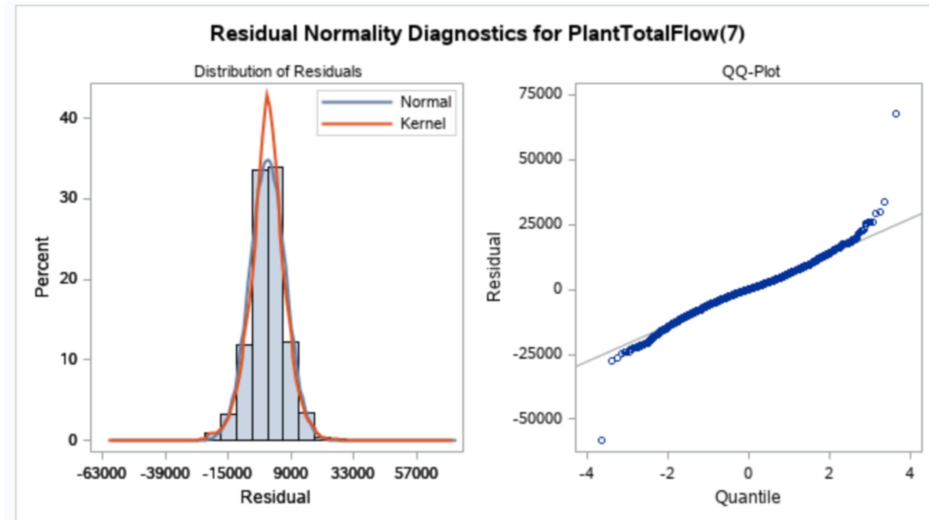


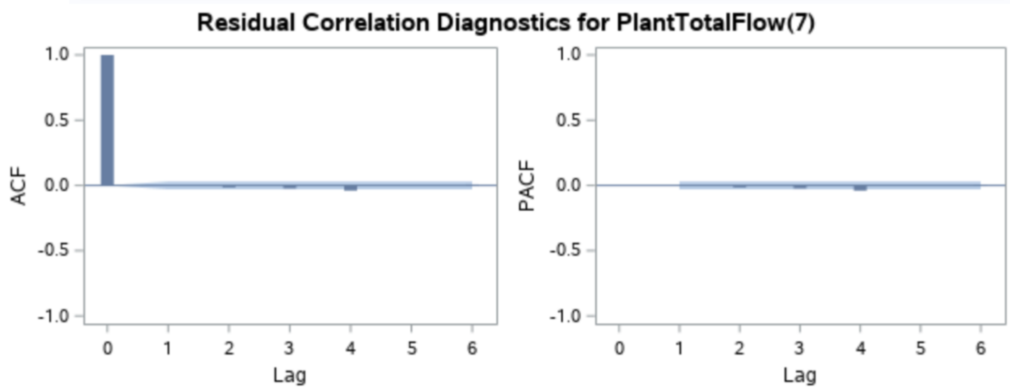
2. Chickamauga Dam



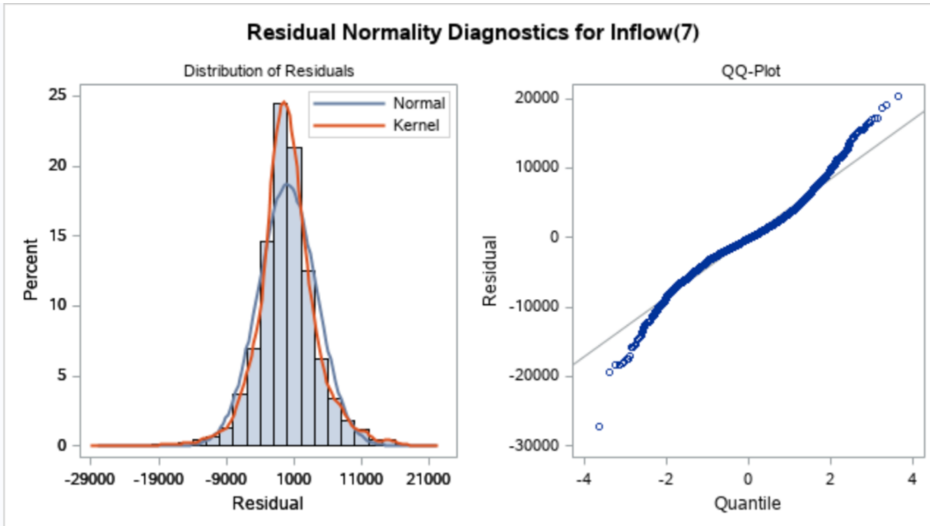


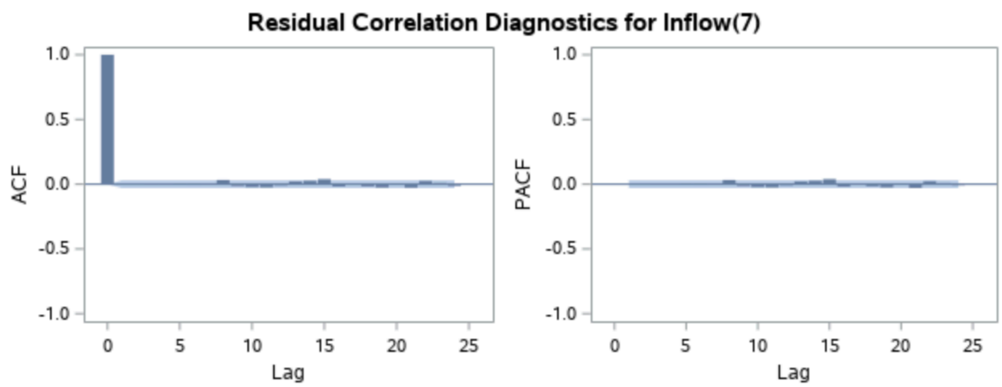
3. Guntersville Dam



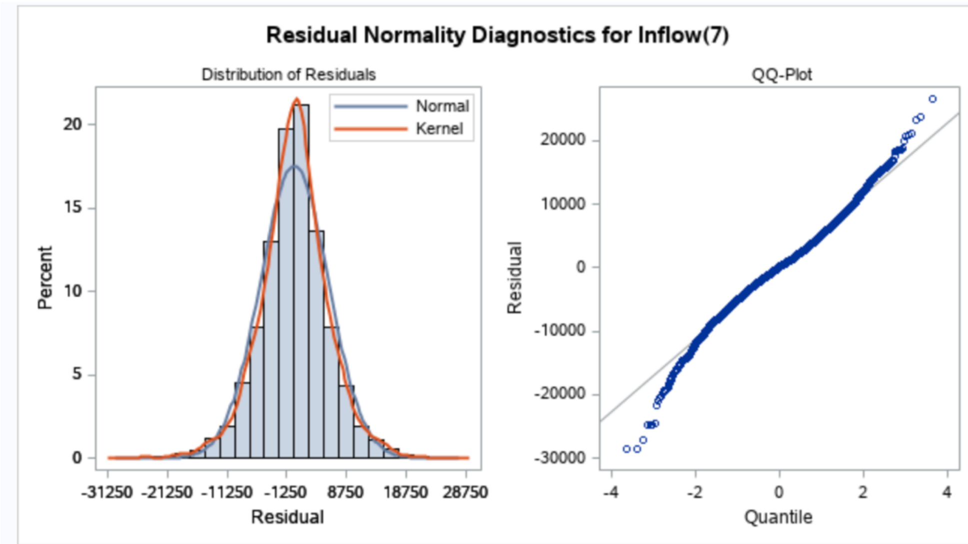


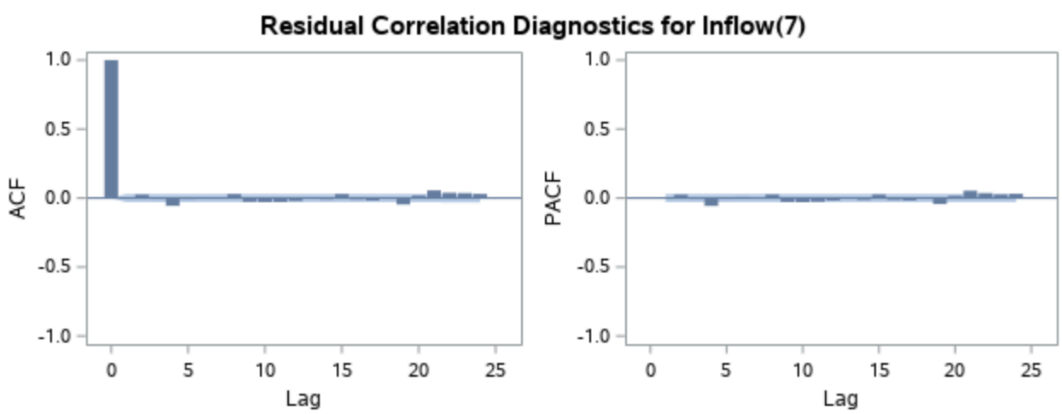
4. Fort Loudon Dam





5. Nickajack Dam





Vita

Born and raised in Kerala, India, Asha Shibu earned her Bachelor's in Electrical and Electronics Engineering from Mahatma Gandhi University. She came to the US in 2016 to pursue a master's degree in Energy from Texas A&M University. As a graduate student at Energy, Asha worked with the Industrial Assessment Center (IAC), a DOE funded Program where she received direct training on energy assessment procedures while her education enhanced her knowledge of energy solution and relevant legislation surrounding energy efficiency and carbon emissions.

After graduation, Asha worked as a Research Intern at the Oak Ridge National Laboratory under the Energy Efficiency Research and Analysis (EERA) group. She was directly involved in three Department of Energy (DOE) projects, and her interaction with the EERA group sharpened her appetite for further knowledge in the field of energy engineering. As a researcher, she realized the importance of collaborating with others in the area and, at the same time, working independently. Her intellectual curiosity and passion for the Energy sector led her to pursue a Ph.D. in Energy Science and Engineering from the University of Tennessee. Asha's dissertation is a part of the project "Hydropower as a signal processor" under the DOE Water-Power Technology Office's HydroWIRES Initiative. Within her research, she focused on developing a transfer function model to identify the relationship between water reservoir inflows and outflows. The research's preliminary findings were quite enlightening, particularly regarding how well the technique can discern between intuitively different facilities (e.g., storage versus run-of-river). In addition to her academic ventures, Asha also served on the Tennessee State Energy Policy Council in the capacity of a graduate student with expertise in energy issues. The experience in the council gave her a sense of the real-life challenges encountered by governing bodies while addressing energy growth and usage issues, including energy production, distribution, and consumption, as well as the relevance of science and policy communication to a broader audience.