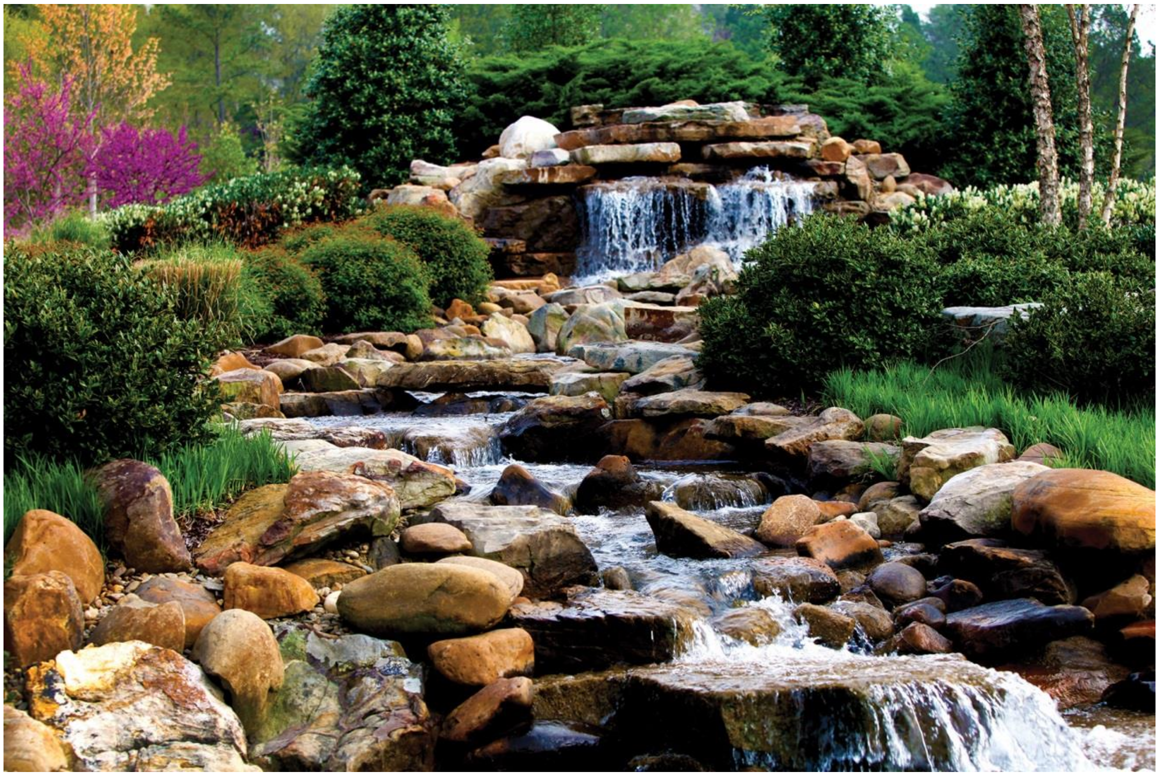
ORNL/SPR-2024/3298

A Hydropower Facility as an Energy Water Signal Processor



Asha Shibu
Srijib Mukherjee
Brennan Smith
Fangxing Li
Russell Zaretzki

April 2024

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/SPR-2024/3298

Environmental Sciences Division

A HYDROPOWER FACILITY AS AN ENERGY WATER SIGNAL PROCESSOR

Asha Shibu
Srijib Mukherjee
Brennan Smith*
Fangxing Li
Russell Zaretzki

April 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

*Hydropower Consultant

CONTENTS

CONTENTS iii
ABBREVIATIONS v
EXECUTIVE SUMMARY vi
1. INTRODUCTION 1
1.1 PURPOSE 1
1.1.1 Water System Perspective of Hydropower Operation 2
1.1.2 Power System Perspective of Hydropower Operation 2
1.2 RESEARCH OBJECTIVES 3
1.2.1 Using Data to Validate Representation 3
2. LITERATURE REVIEW 5
2.1 HYDROPOWER MODLING AND FIDELITY 5
2.1.1 Watershed Models 6
2.1.2 Dispatch Models 7
2.1.3 Production Cost Models 7
2.2 HYDROPWER REPRESENTATION: RiverWare vs. PLEXOS 8
2.2.1 Representation of Hydropower in RiverWare 8
2.2.2 Reservoir Modeling in RiverWare 8
2.2.3 Representation of Hydropower in PLEXOS 9
2.2.4 Reservoir Modeling in PLEXOS 10
2.2.5 Drawbacks in the Representation 11
2.3 SUMMARY 11
2.4 SIGNIFICANCE OF THE STUDY 12
2.4.1 Characterization of Hydropower – Need for a Common Rubric/Nomenclature 12
2.4.2 Modeling Water and Energy Systems 12
2.4.3 Relevance of Time Series Analysis 13
2.5 STATISTICAL MODELING TECHNIQUES 13
2.5.1 Supervised Learning 13
2.5.2 Unsupervised Learning 14
2.5.3 Selecting the Required Model 14
2.6 APPLICATION OF TRANSFER FUNCTION MODELS 14
3. METHODOLOGY 17
3.1 RESEARCH HYPOTHESIS 17
3.2 TRANSFER FUNCTION MODELING 18
3.2.1 Identification Stage 19
3.3 IDENTIFICATION OF THE TRANSFER FUNCTION MODEL 22
3.3.1 Identification of a parsimonious noise model 22
3.3.2 Estimation Stage 24
3.4 DIAGNOSTIC CHECKING 24
4. DATA AND ANALYSIS 25
4.1 DATA AVAILABILITY 25
4.1.1 Fleet Data 25
4.2 DATA PREPROCESSING 25
4.2.1 Description of the Study Area – Norris Hydropower Facility 25
4.2.2 Inflow Computation 26
4.2.3 Choosing the Appropriate Time Step 34
4.3 LIST OF SOFTWARE USED 35
4.4 ANALYSIS 36

4.5 APPLICATION OF THE BOX–JENKINS APPROACH TO DATA FROM STORAGE FACILITIES ... 37
4.5.1 Transfer Function Modeling ... 37
4.6 APPLICATION OF THE BOX–JENKINS APPROACH TO DATA FROM RUN-OF-RIVER FACILITIES ... 41
4.6.1 Transfer Function Modeling ... 41
4.7 CHECKING THE FITTED MODELS ... 44
5. RESULTS AND DISCUSSION ... 45
5.1 GENERAL RESULTS OF TRANSFER FUNCTION MODELING ... 45
5.1.1 Storage Facilities ... 45
5.1.2 Run-of-River Facilities ... 46
5.2 DISCUSSION – INTERPRETATION OF BOX JENKINS MODEL RESULTS ... 47
5.2.1 The Intuition behind the Transfer Function Coefficients: Inflow and Outflow Dynamics ... 47
5.2.2 Comparing Transfer Function Coefficients – Dominance Analysis ... 52
5.2.3 Practical Application of this Research ... 56
5.2.4 Limitations and Scope for Future Studies ... 57
5.3 SUMMARY ... 58
6. CONCLUSIONS AND RECOMMENDATIONS ... 59
6.1 OUTCOMES FROM BOX–JENKINS REPRESENTATION ... 59
6.2 RESEARCH CONTRIBUTIONS ... 59
6.2.1 Recommendations for Further Research ... 60
7. REFERENCES ... 61

ABBREVIATIONS

DOE	US Department of Energy
DOE-WPTO	Water Power Technology Office
TVA	Tennessee Valley Authority
ARIMA	Auto Regressive Integrated Moving Average
LMP	locational marginal price
FTR	financial transmission rights
ORNL	Oak Ridge National Laboratory
DSS	decision support systems
NERC	The North American Electric Reliability Corporation
ISO	independent system operator
RTO	regional transmission organization
MLE	maximum likelihood estimation
NDA	non-disclosure agreement

EXECUTIVE SUMMARY

Recently, various efforts have been made to address the challenge of adequately representing hydropower systems in modeling frameworks by accounting for the lack of data to represent the multiple constraints in hydropower operation. This research used 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 was used to identify the relationship between reservoir inflows and outflows. To examine the feasibility of this methodology, 5 min fleet data for five storage and five run-of-river facilities were provided by the Tennessee Valley Authority (TVA), and transfer function models were 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 to provide stakeholders with a set of concepts to convey the functionality of hydropower.

1. INTRODUCTION

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 energy system models. 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 (i.e., inflows and outflows) should be considered for better representation within electrical grid models. Moreover, the ever-evolving climate change conditions, which impact river patterns with droughts and floods, further underscore this need.

Multiple optimizations and rule-based water management models are available to assist stakeholders in decision-making. Many initiatives have been established to address the issue of proper representation of hydrological and energy systems in modeling frameworks. However, 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 documented herein used 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, research partners, and advisors. The review includes the following considerations:

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 disparate regions and contexts.
3. The feasibility (including data availability) of automating and scaling the methodology for application to the entire North American hydropower fleet.
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 is intended to 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.

1.1 PURPOSE

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 electricity. Thus, hydropower is a complex system composed of water and power systems with distinct goals. Most 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.

1.1.1 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. Therefore, a wide range of factors must be considered when planning and managing a river basin—including economic development, environmental protection concerns, and water-related issues. The production and storage of hydroelectric projects are always 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 operational constraints. The operational policies of a reservoir are significantly influenced by many public agencies, project beneficiaries, and interest groups—in addition to the organizations that own and operate the reservoir system. The objectives of each project determine 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.

1.1.2 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 balance supply and demand for electricity at all times to provide a dependable electrical system. Demand exceeding supply causes the system frequency of the electrical grid to drop below 60 Hz. If the system frequency drifts slightly from 60 Hz, then the spinning generators will naturally apply greater force to one another to restore the frequency back to 60 Hz. If the deviation is very 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 in which they operate have 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 such that different types of resources supply the power system with different types of services and energy. Thus, power system operators must select among the various services provided by generation resources as efficiently as possible. Hydropower contributes significantly 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 immediate demands. Hydropower operators often must adhere to a range of operational and environmental constraints 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 is 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 accountable only for the distribution of electricity to the consumers; the electricity is produced by other entities. Through competitive power markets such as independent system operators (ISOs) and regional transmission organizations (RTOs), 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.

1.2 RESEARCH OBJECTIVES

1.2.1 Using Data to Validate Representation

Power and water systems are vital in hydropower; they are 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. Whereas 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 for 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, 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 although it is clear what questions existing models can help address, there appears to be a lack of clarity about which model is best for answering facility-specific questions; this discrepancy can lead to incompatibility in decision-making.

Data-driven modeling is based on examining the data that characterize 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 water and power systems, using data to highlight the bidirectional transformation between the two systems can aid in conveying the functionality and value of hydropower from a facility perspective.

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

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).

2. LITERATURE REVIEW

Historically, efforts to statistically relate a system's input to its output began with regression analysis. However, a regression model considers only the simultaneous response between the input and output variables. Additionally, regression analysis would be successful only 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 literature review examines the following:

a) Representation of hydropower systems in existing models and the significance of transfer function modeling
b) Application of the Box–Jenkins methodology in non-hydropower-related research

2.1 HYDROPOWER MODLING AND FIDELITY

Decision-making in a hydropower facility falls into three domains: environmental outcomes, operational capabilities, and power system services, as shown in Figure 1.

- Environmental Outcomes: Despite the advantages of hydropower as a relatively clean fuel, the development of hydropower facilities has 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 waters 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 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.

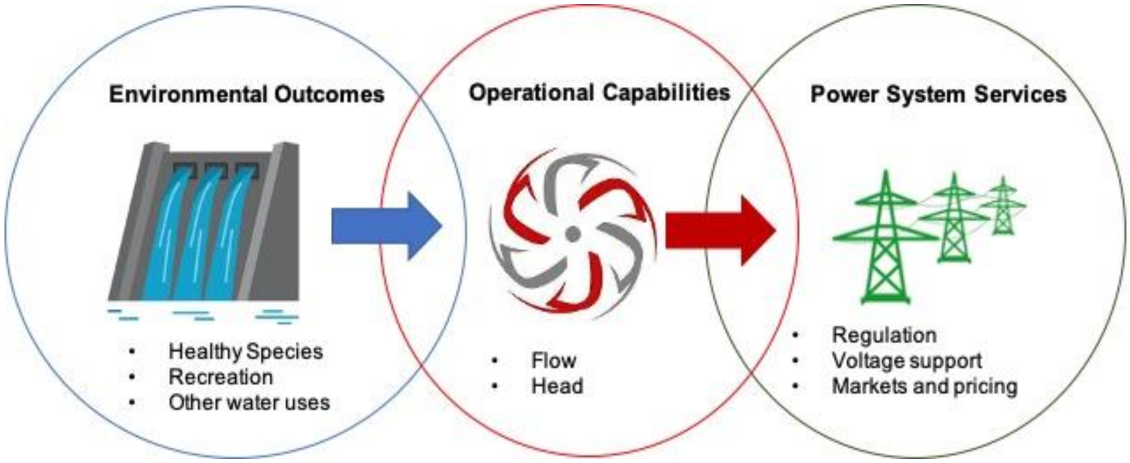


Figure 1: The three major domains in hydropower operation.

2.1.1 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 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, 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 can be inputs, and additional information is provided by prioritized rules that 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 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 composed of nodes and links connecting the nodes. Therefore, in addition to simulation of reservoir allocation and operations, MODSIM can also perform complex water rights accounting without writing scripts or rules (Labadie, 2006).

2.1.2 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). Because the generation must match the load, a set of network constraints as well as security and stability constraints must 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 used 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 are 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).

2.1.3 Production Cost Models

The main objective of production cost models is to minimize 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 reads 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 are 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; 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 generation, 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)

2.2 HYDROPWER REPRESENTATION: RIVERWARE VS. PLEXOS

2.2.1 Representation of Hydropower in RiverWare

RiverWare is a watershed modeling tool developed as part of a collaborative effort by the TVA, the 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 that correspond 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 by importing a database. The objects also contain various methods to model the different processes. There are two method types, namely, dispatch and user selectable. In the dispatch method, the user specifies the input/output configuration to solve the process using conventional algorithms; in user selectable methods, the basin is modeled in accordance with the algorithm/model that the user selects (Zagona & Magee, 1999). The main objects associated with hydropower modeling are listed in Table 1. 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 has no 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 has no 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).

2.2.2 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 in which 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 user. Optimization offers a universal solution across all objects and all the time steps taken—in contrast to the other two procedures, which solve each item individually, one time step at a time. This enables the optimization solution to trade off objectives both spatially and temporally. Modeling inflow, storage, and outflow in the power reservoir (facilities with power generating capability) objects is accomplished using a mass balance approach: the common equations for reservoir mass balance in RiverWare are:

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

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

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

Hydrological inflows are inflows into a reservoir that are not a part of the main stream and/or are 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.

Table 1: 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.

2.2.3 Representation of Hydropower in PLEXOS

The PLEXOS system consists 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 have 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*, whereas *memberships* refer to the relationships between the objects. *Properties* are the characteristics of objects, 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 storage 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 and define the individual or combinations of elements in the hydro system. The different constraints can be set for generators, waterways, and storage in any combination.

There are four different simulation options in PLEXOS: Long Term (LT) Plan, Projected Assessment of System Adequacy (PASA) Plan, Medium Term (MT) schedule, and Short Term (ST) schedule. LT Plan performs long-term expansion planning to minimize the net present value of the overall system expenditure. This is done by building new generating plants, retirements, and transmission upgrades (and retirements). Models for LT plans are user-defined and are often anticipated to last between 10 and 30 years. The PASA plan distributes and optimizes the generating capacity among the different regions to create a reasonable schedule of maintenance outages for each year based on the maintenance rates entered by the user. The MT schedule is a model that uses load duration curve analysis and fully incorporates the generating and transmission systems. It can be used on a weekly or monthly basis. The ST schedule is utilized as a tool for simulating production costs since it is made to simulate the dispatch and price of actual market clearing engines. It also possesses additional capabilities such as unit commitment, constraint modeling, and Monte Carlo simulation. The different simulation options can be used independently or in conjunction. These tools are automatically integrated by PLEXOS, allowing information to flow from one plan to another; however, there are only specific sequences in which such information flow may occur.

2.2.4 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, 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 is 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).

2.2.5 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, whereas storage facilities are depicted as level power reservoirs in the RiverWare workspace that represent the TVA system’s reservoirs (Biddle, 2001). Meanwhile, in PLEXOS, all the reservoirs fall under the Storage Class, which is divided into 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 generate 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). Analysis of hydropower representation in both models shows that they are based on general guidelines for water release, which are derived from the prior knowledge of inflows, reservoir elevations, flow and capacity constraints, and demands. However, neither model can account for the reservoir operation events that are the result of manual decisions made by the reservoir operators, no those 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.

2.3 SUMMARY

Hydropower development is frequently subjected to environmental and regulatory constraints, and representing these constraints in existing energy models is a complex process. The literature 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 yet been possible to capture these intricacies in a single model. 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 of the following reasons.

a) Most models are concerned with linear programming optimization, so 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.

2.4 SIGNIFICANCE OF THE STUDY

2.4.1 Characterization of Hydropower – Need for a Common Rubric/Nomenclature

Most available models (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 in 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 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 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.

2.4.2 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) noted 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, Pacific Northwest National Laboratory (PNNL) and the National Renewable Energy Laboratory (NREL) conducted a workshop supported by DOE to better understand the need for research to improve hydropower representation in grid models (Voisin et al., 2020). This research aims to narrow the knowledge gap in several of the research themes outlined in the workshop report released in November 2020. In particular, it assists 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.

2.4.3 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 determining whether 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 examples include precipitation, flowrate, discharge levels, and streamflow (Survey, 2016). Hydrological time series data have 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.

2.4.3.1 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 characterize 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.

2.5 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 fall into two categories: supervised and unsupervised learning (Sathya & Abraham, 2013).

2.5.1 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 a discrete class label. The test data are 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 a continuous real number. Linear regression, logistic regression, and polynomial regression models are all common types of regression models.

2.5.2 Unsupervised Learning

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

- Clustering: In clustering, the data are 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 When the dimensions of the features are too high this technique is applied. Dimensionality reduction attempts to maintain data integrity while lowering the number of 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.

2.5.3 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 because data from the hydropower fleet are labeled and are in the form of historical time series. Furthermore, the appropriate time to apply unsupervised learning is when there are 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 are in continuous time series format, regression models are believed to be the best choice among the types of models available in supervised learning.

2.6 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 the application of Box–Jenkins transfer function models, and over the years transfer function modeling has 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 United States 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 of 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 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. performed 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 term 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 because 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 can supplement existing models by providing robust data management and modeling tools for water managers, policymakers, and other stakeholders.

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 2. 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 that may influence these signals.

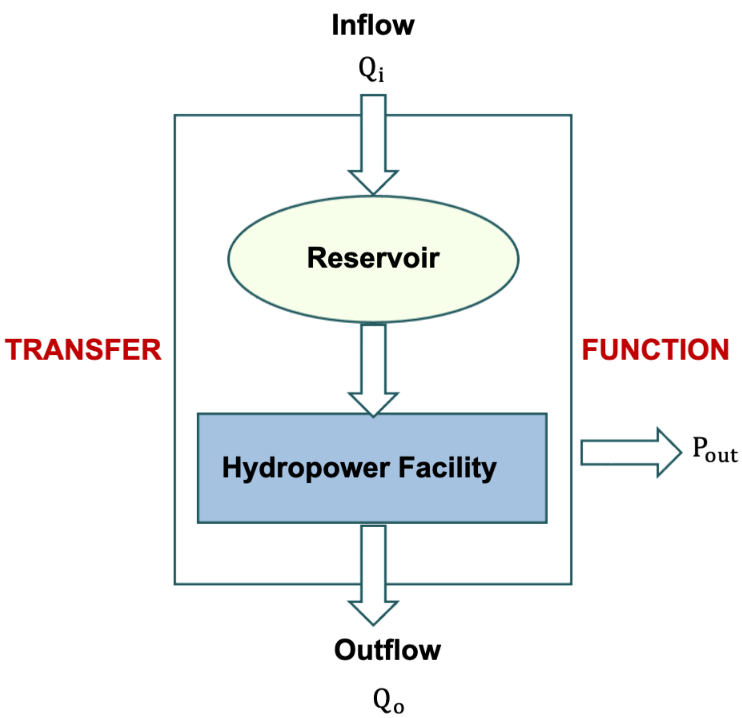


Figure 2. Proposed Transfer Function Model.

3.1 RESEARCH HYPOTHESIS

The main hypothesis of this work is 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.

3.2 TRANSFER FUNCTION MODELING

ARIMA models are univariate time series models, and because 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, then 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 elsewhere (G. E. P. Box et al., 2015).

Fig. 3 gives a schematic diagram of a linear system where the input x_t and output y_t are assumed to be stationary. According to the Box–Jenkins approach, a transfer function model with one input variable x_t may be split into two components,

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

where y_t is the dependent variable, which is transformed to achieve stationarity, u_t is the portion of y_t that can be described in terms of the input variable x_t, and 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 x_t and u_t can be represented as

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

i.e.,

$$\begin{aligned} u_t &= \frac{\omega_0 - \omega_1 B - \cdots - \omega_s B^s}{1 - \delta_1 B - \cdots - \delta_r B^r} x_{t-b} \\ &= \frac{\omega(B)}{\delta(B)} x_{t-b} \\ &= v(B) x_t \end{aligned}$$

where $v(B) = \frac{\omega(B)}{\delta(B)} B^b$. The polynomial operator v(B) is also referred to as the *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 v_0 ,v_1, v_2 etc. of the polynomial operator v(B).

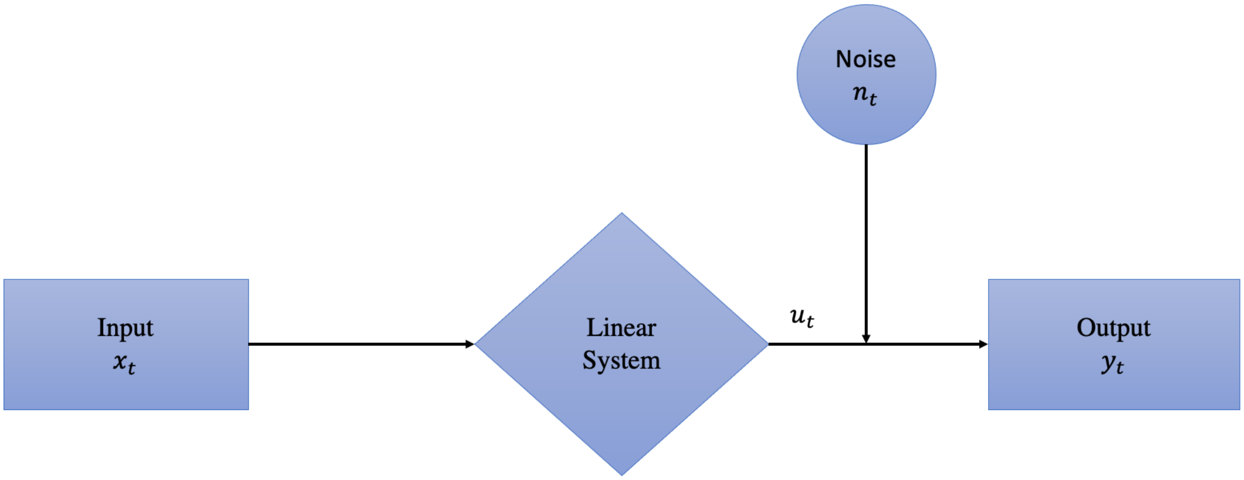


Figure 3. Schematic representation of a linear system.

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

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

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 the following terms:

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

$$\mathrm{v_j} = 0 \qquad ; j < b$$
$$\mathrm{v_j} = \delta_1 \mathrm{v_{j-1}} + \delta_2 \mathrm{v_{j-2}} + \cdots \delta_r \mathrm{v_{j-r}} + \omega_0 \qquad ; j = b$$
$$\mathrm{v_j} = \delta_1 \mathrm{v_{j-1}} + \delta_2 \mathrm{v_{j-2}} + \cdots \delta_r \mathrm{v_{j-r}} - \omega_{j-b\mathrm{Com}} \qquad ; j = b+1, b+2, .. b+s$$
$$\mathrm{v_j} = \delta_1 \mathrm{v_{j-1}} + \delta_2 \mathrm{v_{j-2}} + \cdots \delta_r \mathrm{v_{j-r}} \qquad ; j > b+s.$$

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

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, whereas the 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 $\hat{\mathrm{v}}_{\mathrm{j}}$ 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. 4.

3.2.1.2 Pre-Whitening of the Input

If the input series $\mathrm{x_t}$ is autocorrelated, then 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:

$$\Phi\,(B)\mathrm{x_t} = \theta(\mathrm{B})\alpha_{\mathrm{t}} \qquad (3.4)$$

Given the fitted model the residuals $\mathrm{a_t}$ is computed by Equation (3.5)

$$\Phi\,(B)\theta^{-1}(\mathrm{B})\mathrm{x_t} = \alpha_{\mathrm{t}} \qquad (3.5)$$

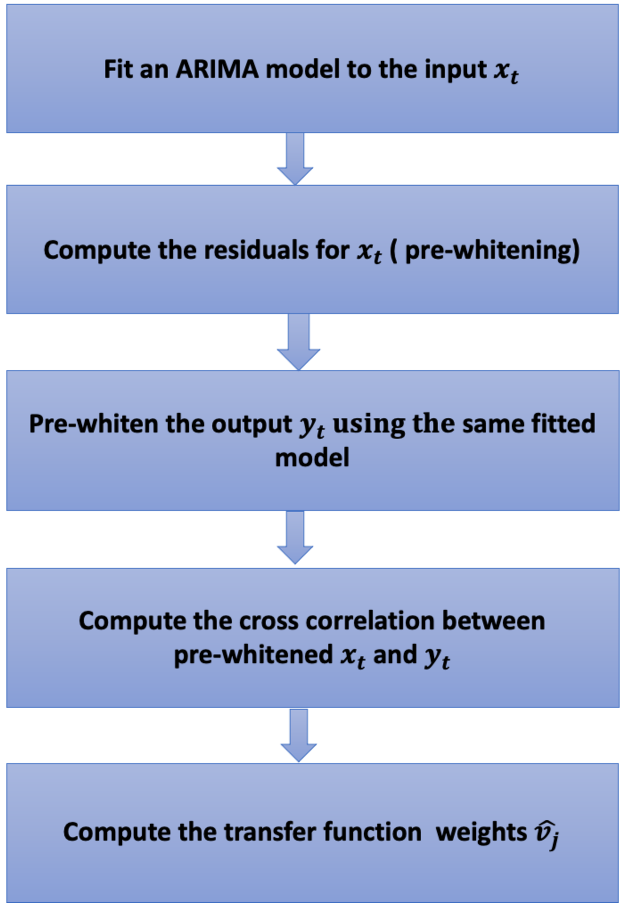


Figure 4. Procedure for estimating impulse response function.

3.2.1.3 Transformation of the Output

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

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

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 $\hat{v}_j$, 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 $\hat{v}_j$ can be computed as

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

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 $\hat{v}_j$ can be determined.

3.3 IDENTIFICATION OF THE TRANSFER FUNCTION MODEL

Identification of the 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 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. 5.

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)

$$\mathrm{v}(B) = \frac{\omega(B)}{\delta(B)} \mathrm{B}^{\mathrm{b}}$$
$$\delta(B)\mathrm{v}(B) = \omega(B)\mathrm{B}^{\mathrm{b}}$$
$$(1 - \delta_1 B - \cdots .. - \delta_r B^r)(\mathrm{v}_0 + \mathrm{v}_1 B + \cdots) = (\omega_0 - \omega_1 \mathrm{B} - \cdots - \omega_s B^s)\mathrm{B}^{\mathrm{b}}$$

This gives

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

Once the values of r, s and b are identified, it is then substituted in Eq. (3.8) 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.

$$\mathrm{Y_t} = \mathrm{v(B)x_t} + \mathrm{n_t}$$
$$\mathrm{n_t} = \mathrm{y_t} - \mathrm{v(B)x_t}$$

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

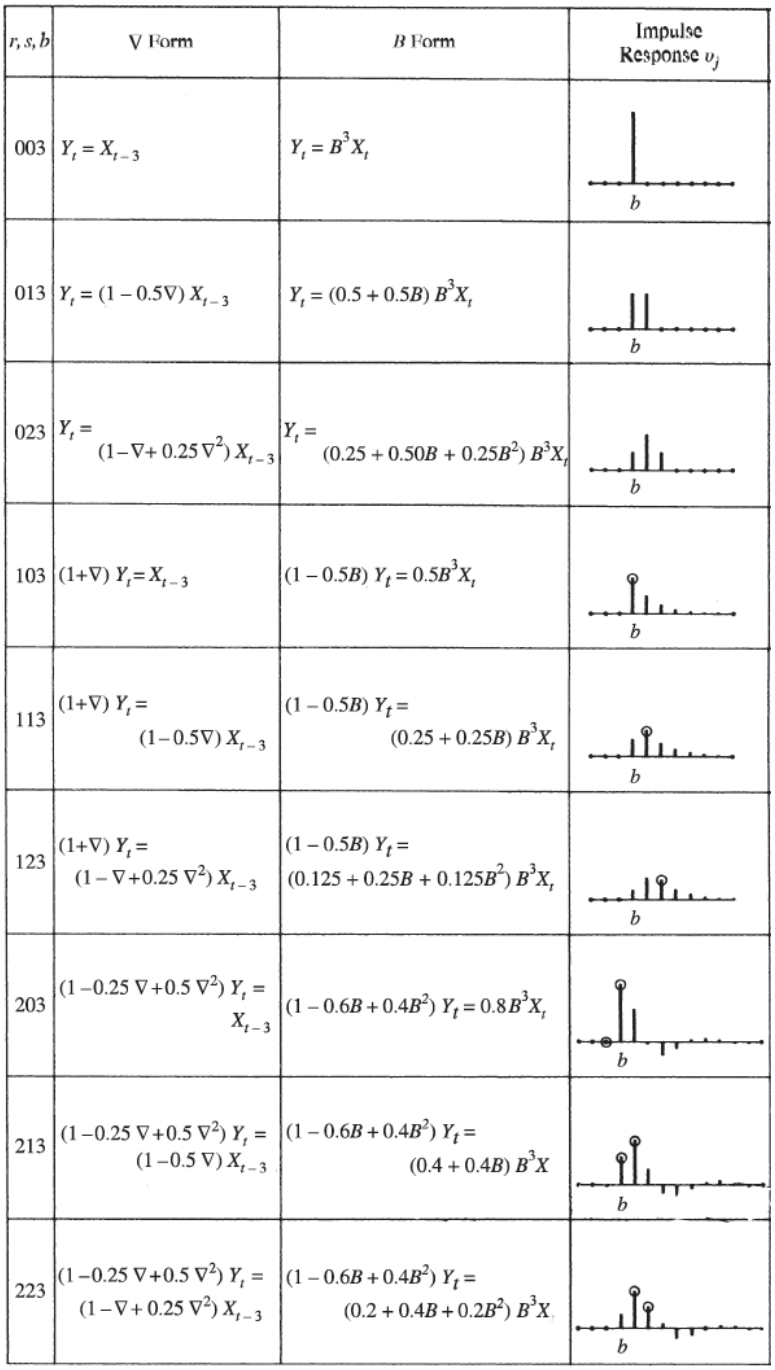
$$\mathrm{n_t} = \mathrm{y_t} - \mathrm{v_0 x_t} - \mathrm{v_1 x_{t-1}} - \mathrm{v_2 x_{t-2}} - \cdots$$

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

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

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

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



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

Figure 5. Examples of impulse response functions of commonly adopted transfer functions.

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 $\theta(\mathrm{B})$ 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 $\mathrm{a_t}$ of the tentative model are examined to see whether they correlate with the unsystematic changes in the input $\mathrm{x_t}$. Statistical methods, such as the Ljung–Box Chi-Square test, can be used to establish that the residuals are random.

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 US (Bracmort, Stern, & Vann, 2013). The major federal entities include the US Army Corps of Engineers (Corps), the Bureau of Reclamation, and the 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 shown in Table 2. The data were recorded on a 5 min 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 from 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 are not repeated in the other facilities evaluated except when warranted.

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. 6). 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 (TVA).

Table 2. 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 6. 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% 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. 7, and a snapshot of the data is shown in Fig. 8. (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 are 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),

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

where

- Q_i = Reservoir inflow
- Q_r = Reservoir release

- V_i = Initial storage volume
- V_{i+1} = Final storage volume
- ΔT = Time period under consideration
- Q_l = Water losses (including evaporation and seepage losses)

The reservoir storage V_I is estimated by the reservoir stage-elevation relationship.

$$V_t = f(HW)$$

Here, HW is the observed water level. It should be noted that the water losses are not usually quantified; thus, for this calculation, it was assumed to be negligible. The data for computing the relationship between the reservoir storage and elevation were provided by TVA through direct correspondence; they were then plotted to obtain the function shown in Fig 9. A second-degree polynomial relationship is identified between the elevation level and reservoir volume, expressed as:

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

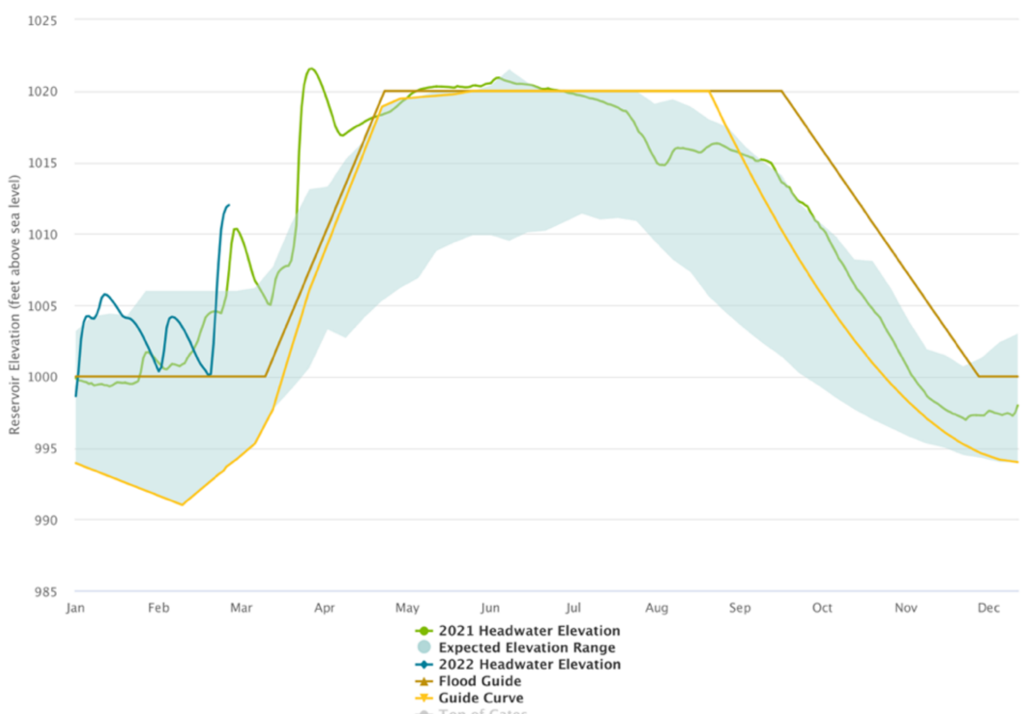


Figure 7. Operating Guide of Norris Dam.

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 8. Synopsis of operating data.

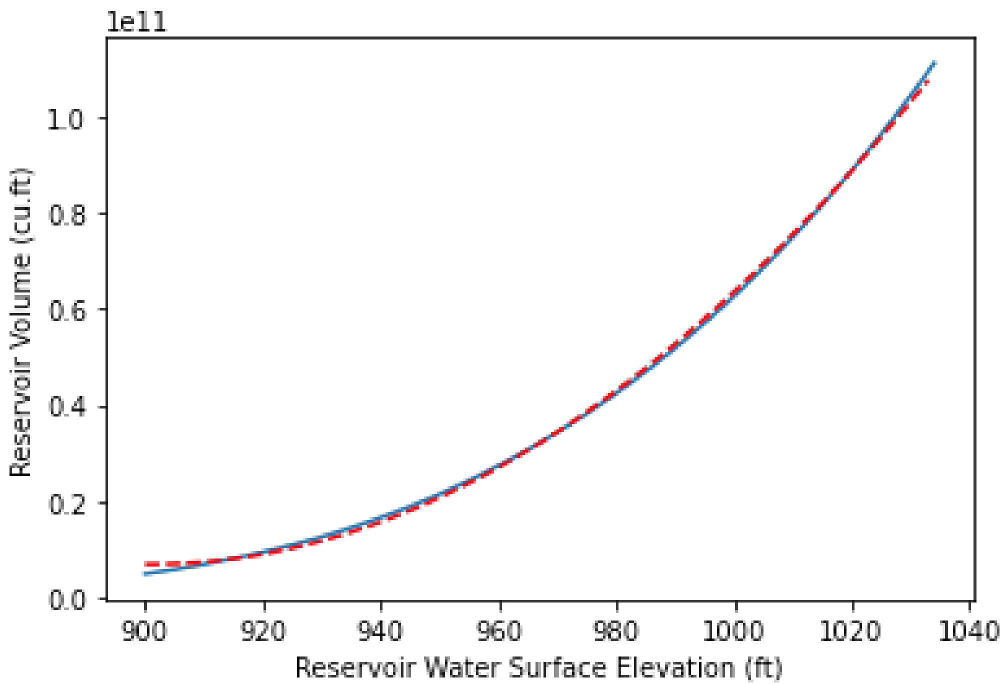


Figure 9. Curve-fitting – Reservoir volume as a function of elevation.

4.2.2.1 Negative Inflow Values: Sources of Error

The initial inflow estimation produced inflow values with significant negative terms as shown in Fig.10 and 11. 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 12).
b) Due to the lack of recent data for computing the reservoir volume–elevation curve, the reservoir capacity at specific elevations could be miscalculated.
c) The estimated uncertainties in computed outflows for lakes having continuous-record gauging 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.13.

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 is 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; Fig. 12 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. 13. 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).

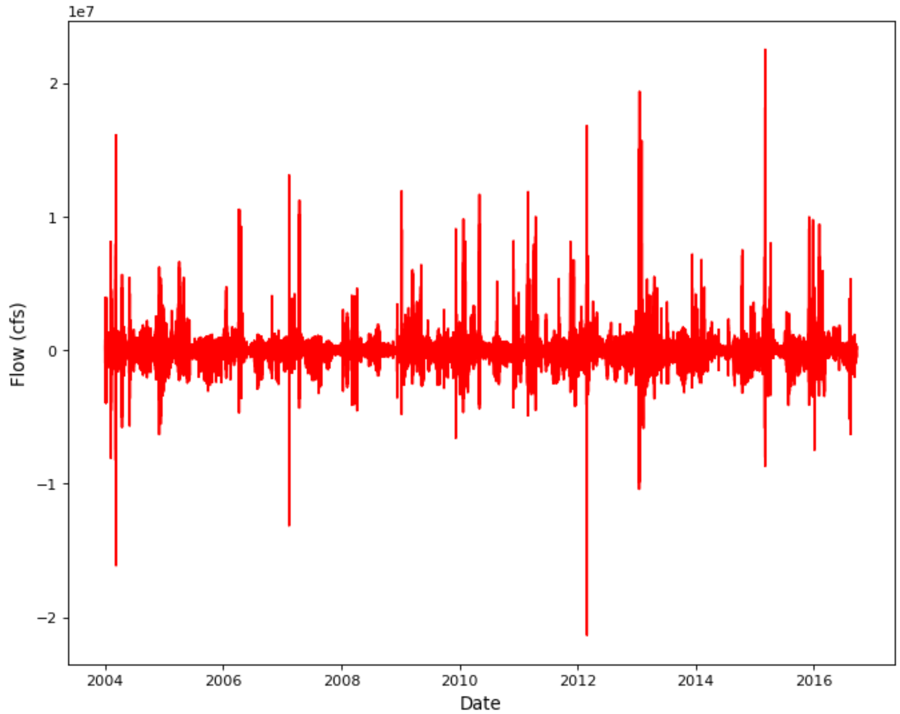


Figure 10. Computed reservoir inflow.

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 11. Computed reservoir inflow data.

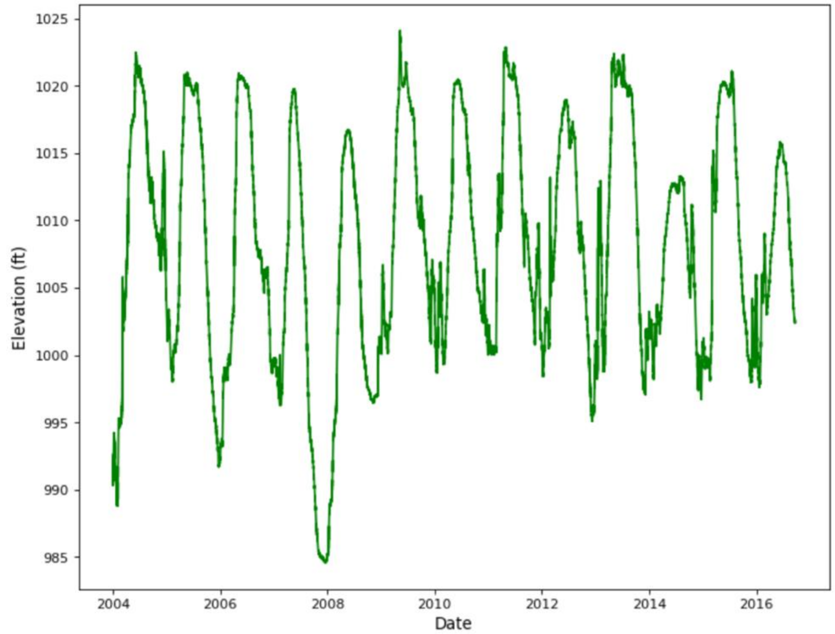


Figure 12. Recorded elevation levels.

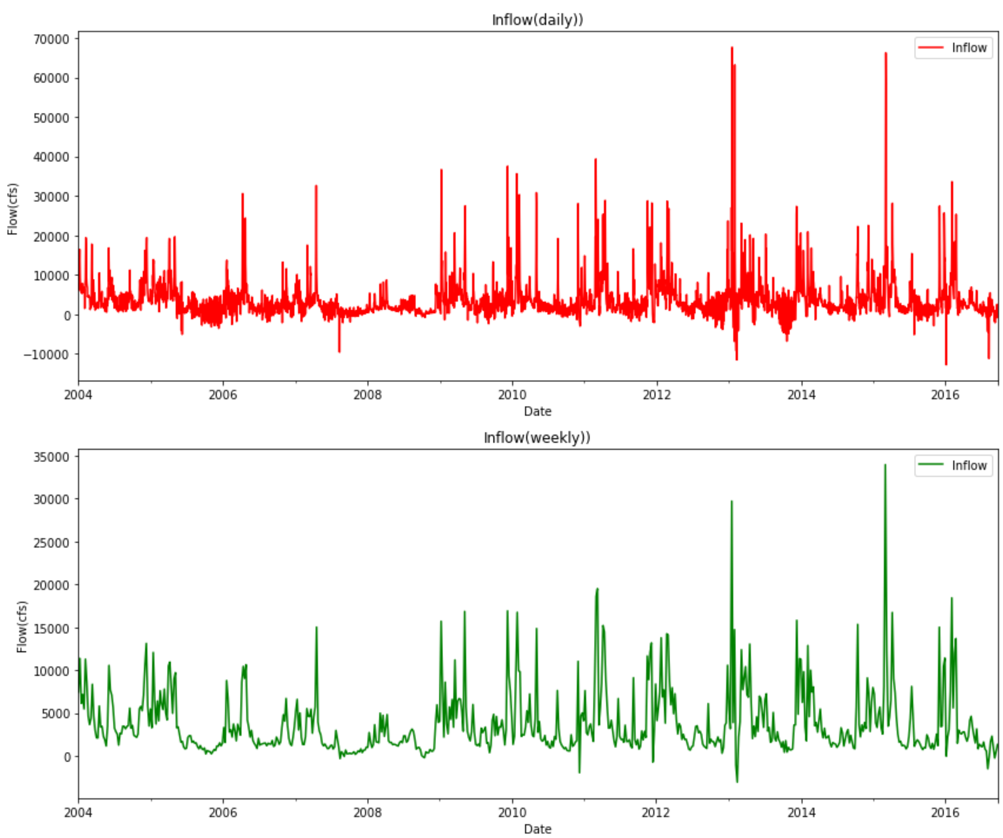


Figure 13. Computed net inflows – daily and weekly.

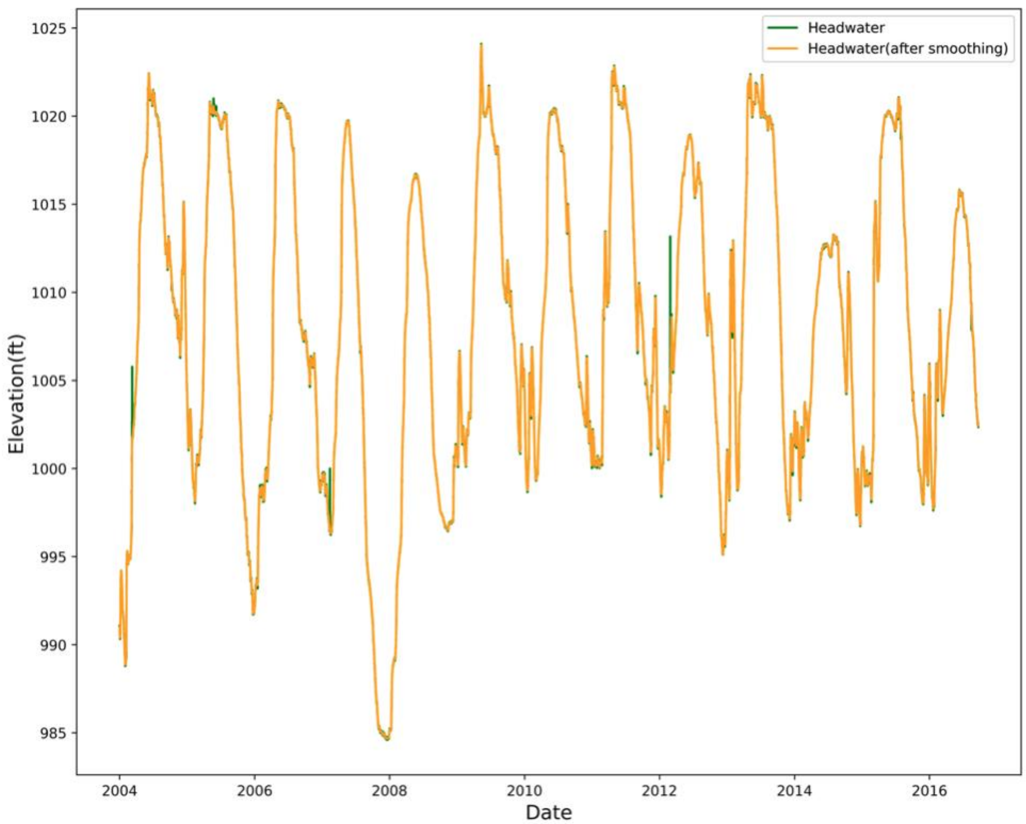


Figure 14. Elevation levels after smoothing.

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, I_u is calculated as

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

where

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
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 because 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% of the total inflow comes from the Clinch River basin, and the Powell River provides the remaining 30% (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 3. 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: the results indicate 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 15.

Table 3. Results of correlation analysis

<table>
<tr><th>Gages used</th><th>Type of inflow data</th><th>Value of r</th><th>Improvement in correlation</th></tr>
<tr><td rowspan="2">Clinch River above Tazewell

Powell River near Arthur</td><td>Inflow (Water Balance Method)</td><td>0.58</td><td rowspan="2">0.16</td></tr>
<tr><td>Inflow (Proposed Methodology)</td><td>0.74</td></tr>
</table>

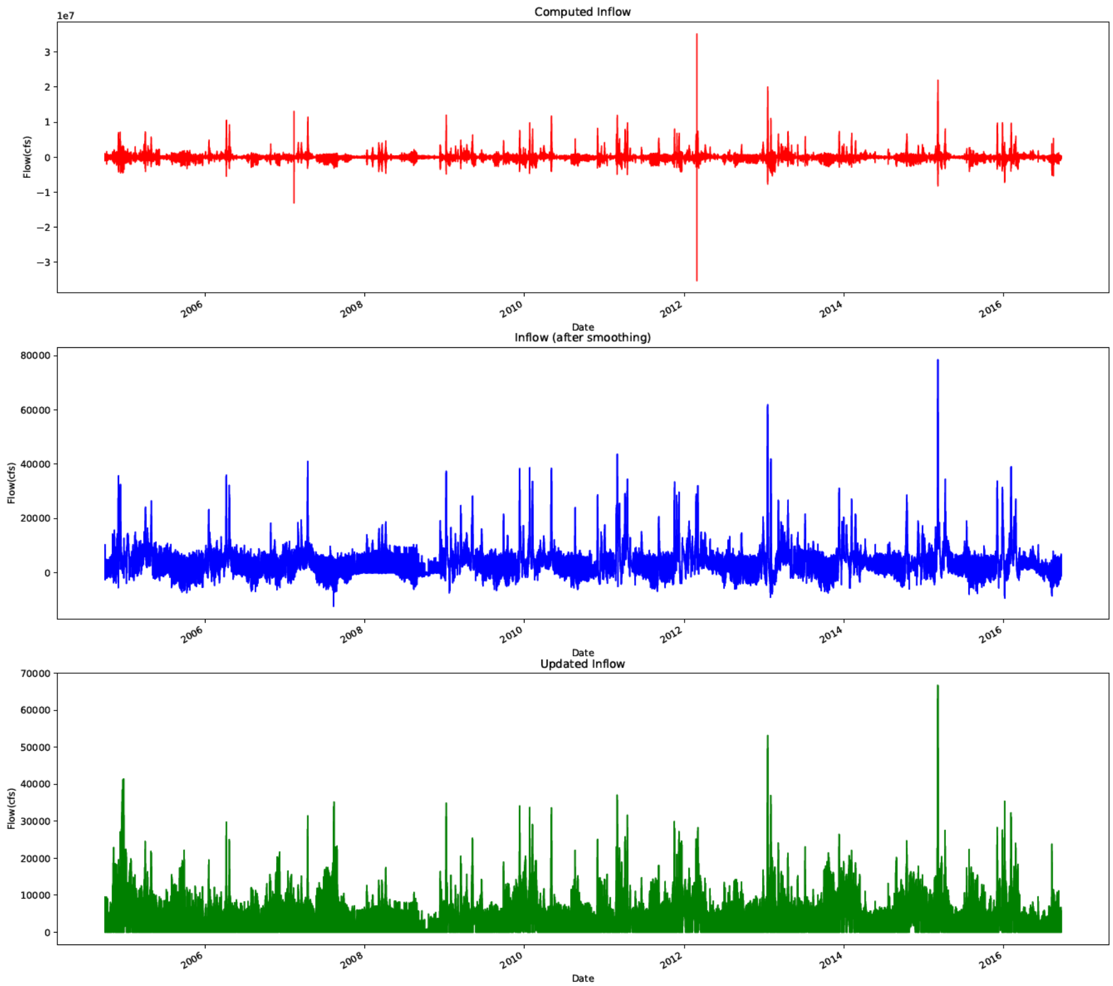


Figure 15. Replacing negative inflow values – different stages.

4.2.3 Choosing the Appropriate Time Step

The time-series data provided by the TVA is recorded at the 5 min interval, and for developing a transfer function model, there are four options available (Fig. 16):

a) Using the dataset as is, with 5 min 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 are less granular and follow 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. 16. It should also be observed that the monthly data show 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.

Although 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 work for developing the Box–Jenkins model:

1. Python: The Pandas library was used to perform the initial data preparation, which includes steps like combining the 5 min fleet, inputting 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 were used to compare various ARIMA models and choose the one with the lowest AIC values.
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 were obtained.

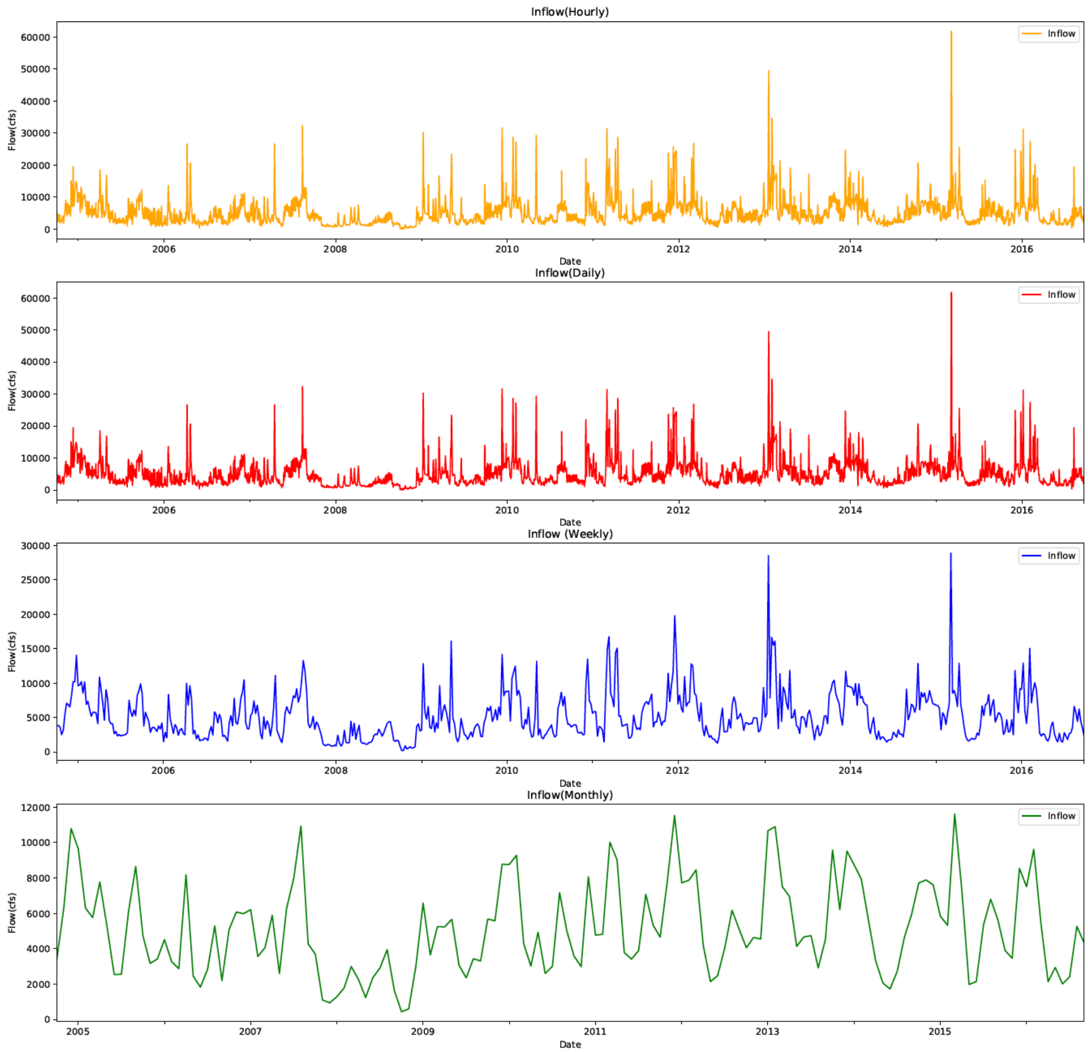


Figure 16. Inflow across different time intervals.

4.4 ANALYSIS

After the data have been pre-processed and the inflow has been calculated, the Box–Jenkins approach discussed in Section 3 was 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 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, whereas the trend is the data's progressive upward or downward movement over time. The seasonal component explains the patterns that repeat at regular intervals, and when seasonality and trend are separated from the time series, the noise or random component remain. 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. The 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. 17, and additional information on the elevation levels, the number of units, and net dependable capacity is given in Table 4. 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 was conducted utilizing the *statsmodels* package in Python, and the results suggest that the inflow time series is stationary; however, 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 Section 3 was applied to the inflow and outflow time series obtained from these facilities. The relevant equations were presented in Section 3 and are only mentioned here for clarity. Mathematical symbols are used if equations are utilized, and they follow the notations used in Section 3.

4.5.1.1 Pre-Whitening Inflow and Outflow

The first step in the 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 5. The coefficients of the pre-whitening filter are 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 17. Map showing location of storage facilities.

Table 4. 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 5. Pre-whitening of Inflow time series – storage facilities.

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

Additionally, significant positive spikes in the CCF plot indicate that changes in the inflow cause changes in outflow to follow suit, whereas 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. 18.

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 the current day was more significant factor in 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 the case of simple exponential decay, $r = 1$ and in compound exponential or sinusoidal decay patterns, $r = 2$

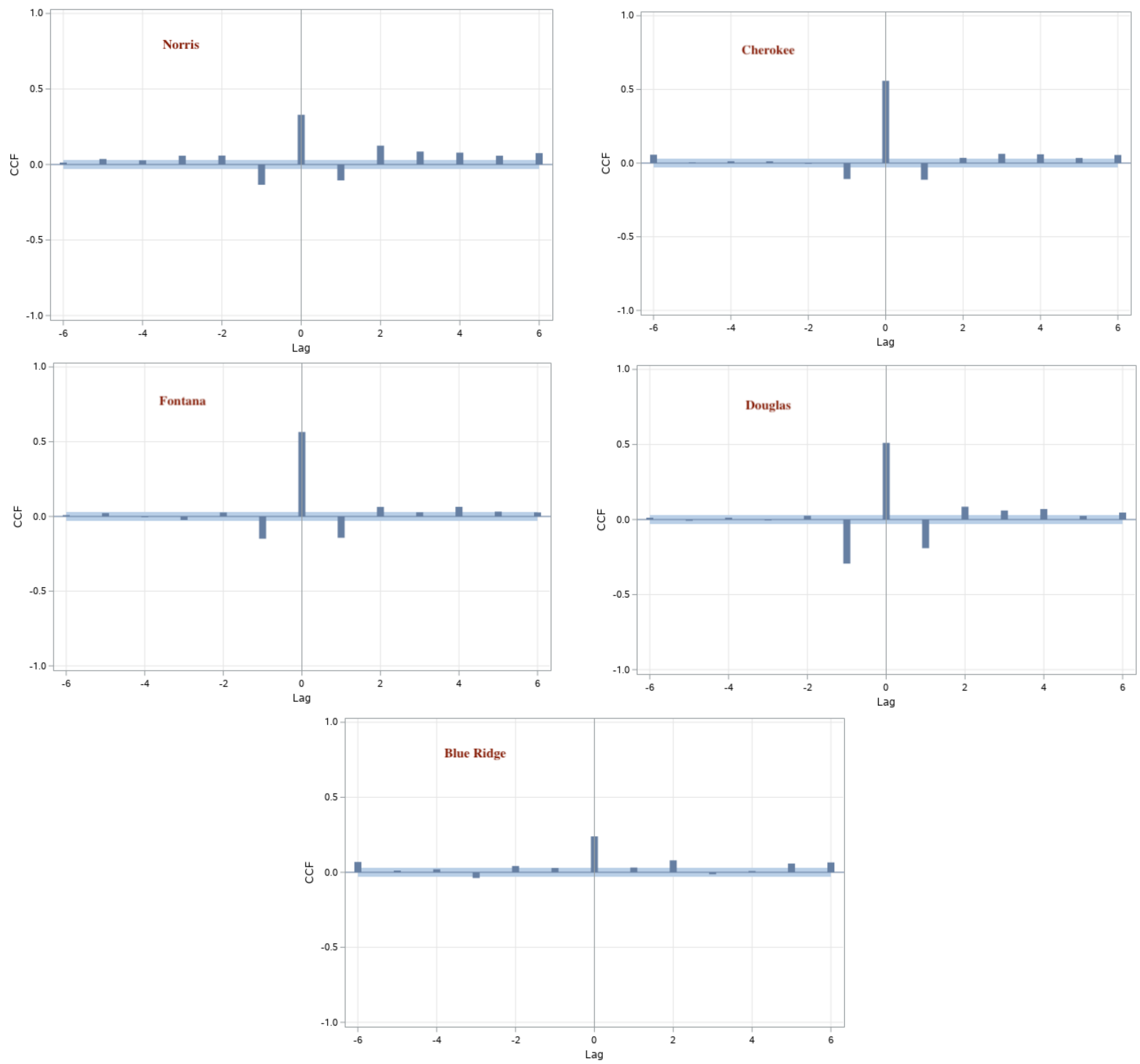


Figure 18. CCF plots of storage facilities.

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, the IDENTIFY and ESTIMATE functions in the proc ARIMA functionality of SAS were 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 6 where Y_t represents outflow and X_t represents inflow.

Table 6. Identified Transfer Function Models: storage

Facility	$(r, s\ b)$	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)	$Y_t = -7.78 + 0.08\left[\frac{1}{1+0.38\mathrm{B}}\right]X_t + \left[\frac{(1+0.22\mathrm{B})}{(1-0.48\mathrm{B})}\right]\mathrm{a_t}$

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 show 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 were modeled using the Box–Jenkins modeling approach.

4.6.1.1 Pre-Whitening Inflow and Outflow

The details of the 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 Interpreting the CCF Plots

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 the current day has a greater effect on the outflow. Additionally, in three of these sites, Watts Bar, Chickamauga, and Guntersville, the lag-2 and lag+2 weights were more significant than other lags. Lastly, in both Fort Loudon and Nickajack facilities, the only significant correlation is at lag 0.

4.6.1.3 Identification of Transfer Function Model

The outcomes of the final transfer function model developed for these five run-of-river facilities studied are listed in Table 7.

Figure 19. Map showing locations of run-of-river facilities.

Table 7. 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 8. 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]	$\left[\frac{(1-0.64B-0.99B^2+0.33B^3-0.3B^4)(1-0.97B^7)}{(1-1.51B-0.17B^2+0.91B^3-0.23B^4)(1-0.03B^7)}\right]\mathrm{Inflow}_t$
Chickamauga	(3,0,3) (1,1,1) [7]	$\left[\frac{(1+0.27B-0.64B^2-0.28B^3)(1-0.94B^7)}{(1-0.6B-0.6B^2+0.22B^3)(1-0.04B^7)}\right]\mathrm{Inflow}_t$
Guntersville	(4,0,4) (0,1,1) [7]	$\left[\frac{(1-1.41B+1.06B^2-0.14B^3-0.27B^4)(1-0.94B^7)}{(1-2.06B+2.11B^2-1.13B^3+0.1B^4)}\right]\mathrm{Inflow}_t$
Fort Loudon	(3,0,2) (1,1,1) [7]	$\left[\frac{(1-0.29B-0.44B^2)(1-0.97B^7)}{(1-0.95B-0.19B^2+0.16B^3)(1-0.08B^7)}\right]\mathrm{Inflow}_t$
Nickajack	(3,0,4) (1,1,1) [7]	$\left[\frac{(1+1.2B-0.19B^2-0.56B^3-0.08B^4)(1-0.96B^7)}{(1+0.69B-0.87B^2-0.67B^3)(1-0.08B^7)}\right]\mathrm{Inflow}_t$

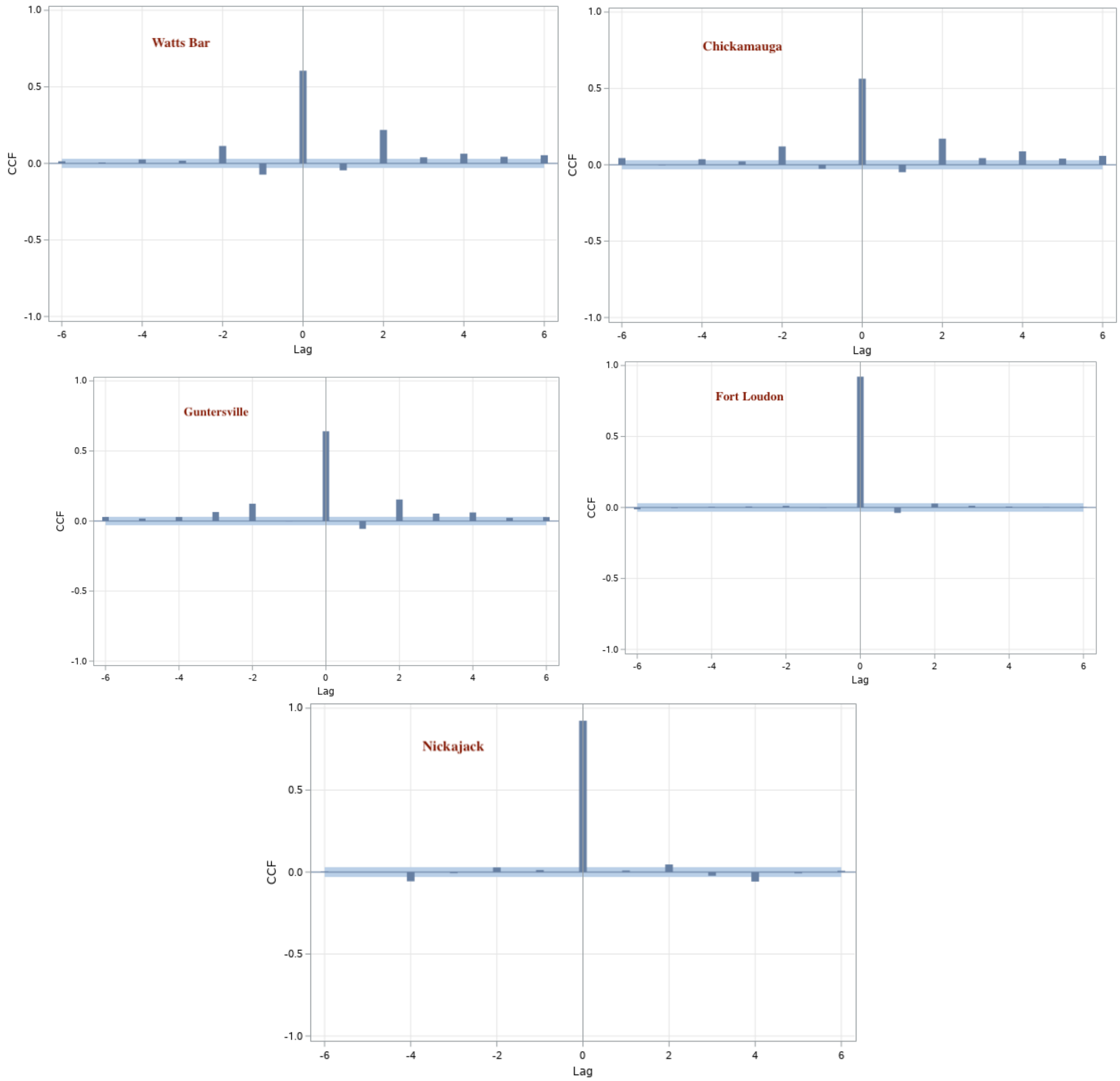


Figure 20. CCF plots of run-of-river facilities.

Table 9. Inflow–Outflow relationship of run-of-river facilities

Facility	$(r, s\ b)$	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 performing various diagnostic tests. Additionally, the model should meet all the criteria below:

i. The model should include only 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 $\mathrm{n_t}$ should be stationary.

5. RESULTS AND DISCUSSION

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

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

Because 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 cannot be certain which pattern is best for a given data set. In the previous section, 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 10. The relationship obtained between outflow and inflow are also summarized in Table 11. The information obtained from the transfer function relationship can be categorized as follows:

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; for the other three, the outflows from the previous day had a greater impact on the value of the present day.
2. Current Information: In all the storage facilities studied, the current day inflow information showed the higher impact on the outflow values. In Cherokee, Fontana, and Douglas, the magnitude of increase in outflow with a change in inflow was higher than that of the other two facilities analyzed.

Table 10. 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 11. 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 display a significant difference from those of the storage facilities. 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 follows:

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 showed 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 the current day also depends on the value of inflow two days prior $(t-2)$. Additionally, for the 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.

Table 12. 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 13. 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	$Y_t = X_t + 0.008Y_{t-1}$

5.2 DISCUSSION – INTERPRETATION OF BOX JENKINS MODEL RESULTS

The main goal of this research is to examine how well the Box–Jenkins methodology categorizes various kinds of hydropower facilities. The review of the different types of classification of hydropower facilities presented in Section 1.1 shows 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 significantly affects 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.

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 21 illustrates the dependence between inflows and outflows derived from the transfer function modeling. The solid line represents the direct pathways of influence between the variables. 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 14. 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, whereas 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).

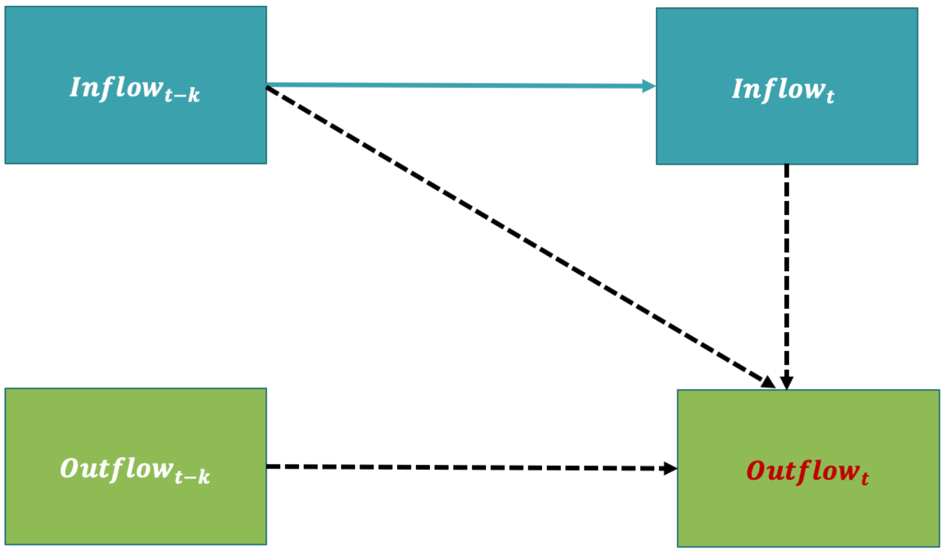


Figure 21. Inflow–Outflow dependence.

Table 14. Effect of coefficients and lags

Facility		**Contribution to $Outflow_t$**				
		Outflow		**Inflow**		
		$(t-1)$	**$(t-2)$**	**t**	**$(t-1)$**	**$(t-2)$**
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		

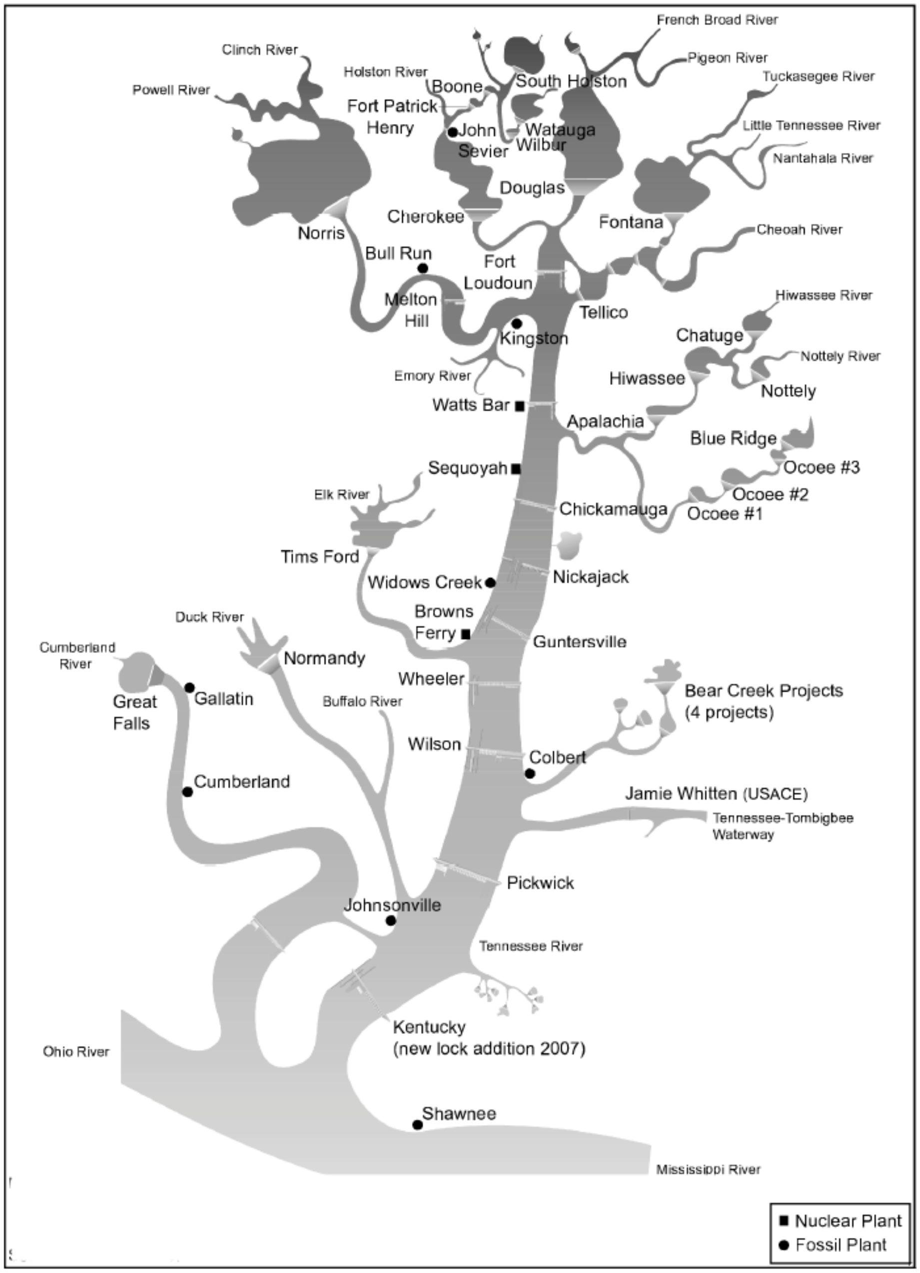


Figure 22. The TVA Water Control System.

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). 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 22. 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 reasons for these outcomes are discussed below.

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, at 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, whereas Douglas has a different trend. In what follows, each facility is analyzed separately depending on its location and operational objectives within 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 from 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), whereas outflows from Blue Ridge Dam make for 56% 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, whereas 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 a basin-wide multiple reservoir system, and the results of transfer function modeling show a positive correlation with current day inflow and the outflow two days prior, as well as 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 a 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 to the 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 one 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 outflows 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, whereas a negative correlation was 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 ft wide by 360 ft long and can lower barges up to 50 ft (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 the transfer function are different compared to other runoff facilities studied. The outflow at time t is 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 come from releases from Nickajack Dam, which accounts for almost 89% 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.

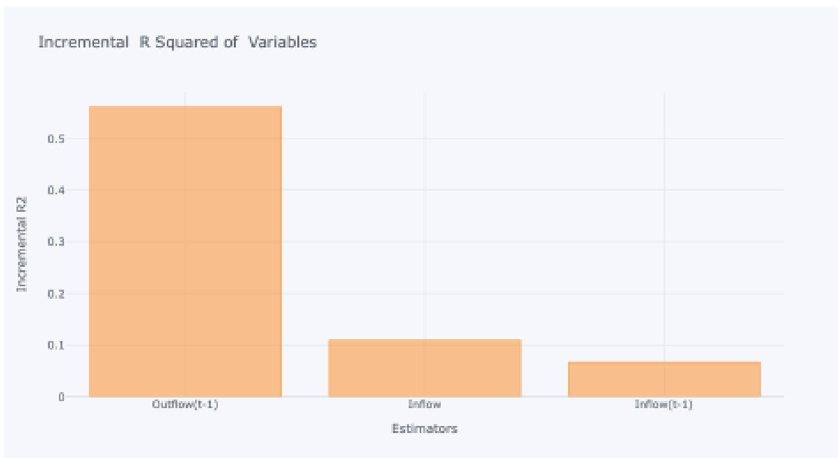
Dominance analysis is a commonly used technique in statistical models to determine the importance of independent variable (Budescu, 1993). Dominance statistics can be divided into four different types of measures:

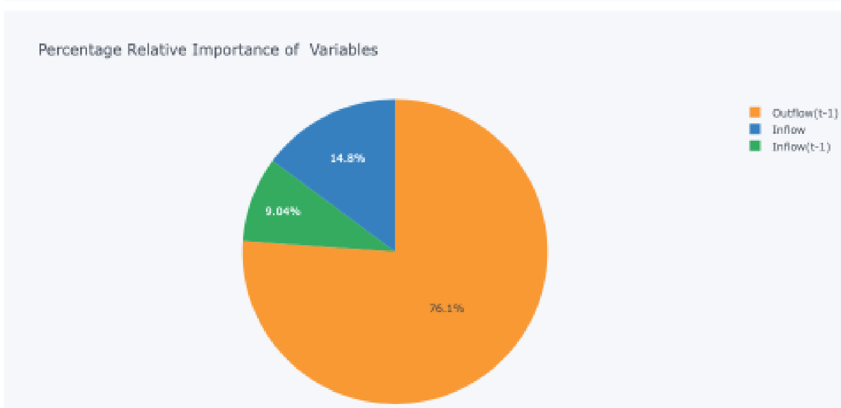
- Individual Dominance: Individual dominance provides the variability in the independent (predictor) variable alone on the absence of other variables. Mathematically it is the R^2 vale of the model between the dependent and predictor variable.
- Interactional Dominance: When all other predictors are present, the interactional dominance is the impact of variability expressed by a given predictor variable. It is the difference between the R^2

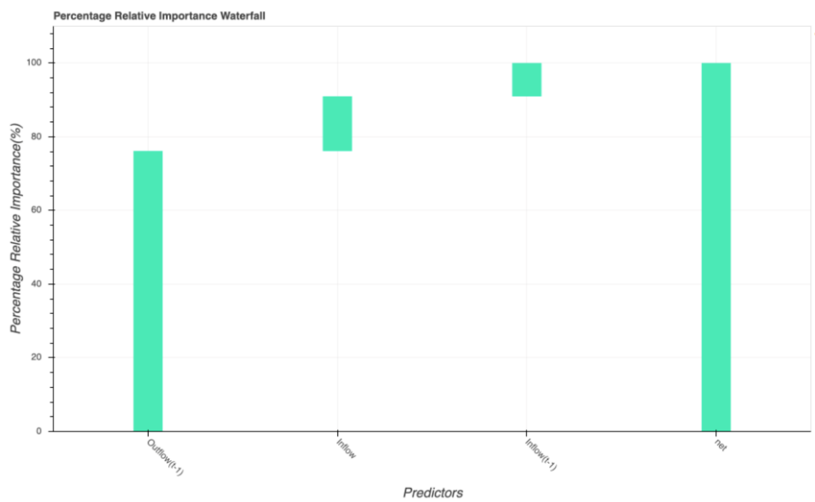
vale of the overall model and the R^2 value of the model calculated without the specified predictor variable.

- 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 are shown as an example in Figure 23. The *dominance analysis* package in Python is utilized for finding the relative importance of the different independent variables.







Facility	Percentage Relative Importance (%)			
	Inflow(t)	**Inflow ($t-1$)**	**Outflow ($t-1$)**	**Outflow ($t-2$)**
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	

	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 23. Results of Dominance Analysis – Douglas Dam.

The percentage relative importance computed for all the ten facilities are summarized in Table 15 and 16.

Table 15. Relative importance: run-of-river facilities

Facility	Percentage Relative Importance (%)			
	Inflow (t)	Inflow $(t-2)$	Outflow $(t-1)$	Outflow $(t-2)$
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 16. Relative importance: storage facilities

Facility	Percentage Relative Importance (%)			
	Inflow(t)	Inflow $(t-1)$	Outflow $(t-1)$	Outflow $(t-2)$
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	

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 reveal 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 was used in this study only 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.

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 the 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 24 below is an extension of Figure 21 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.

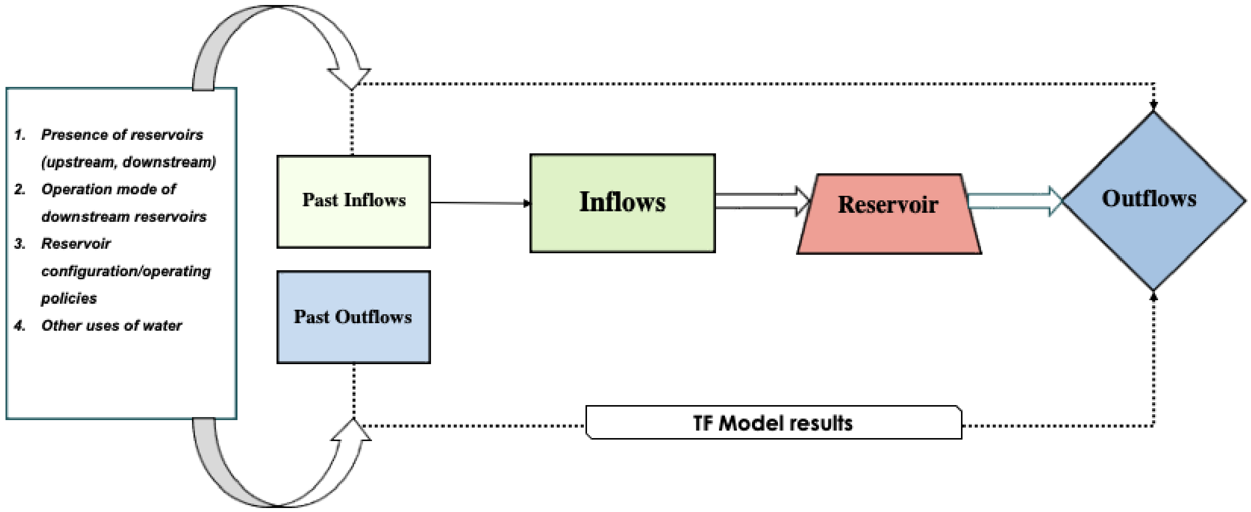


Figure 24. Factors significant for model interpretation.

1. *<u>Relevance of past outflows</u>*
   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 a 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 contain much more than just what is initially obvious.
   Additionally, as seen with the example of Norris and Blue Ridge, 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. *<u>Relevance of past inflows</u>*
   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 a 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 takes 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 min flows time-consuming. Applying the Box–Jenkins methodology is also subjective, and the researcher's expertise and experience may influence the reliability of a chosen model.

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 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. The 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 the application of 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 in 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 in interpreting the model results: location and operation mode of downstream reservoirs, the release pattern of the upstream reservoirs and their contribution to the inflows, and other relevant uses of water (such as flood control, cooling water source etc.).

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 rely on complex and unrecorded decision-making procedures, representation of hydropower in energy system models is challenging. The perspectives of three key stakeholders—a reservoir operator, a power system dispatcher, and lastly a downstream water user—should be examined 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 be in regular contact with the reservoir operators, who are responsible for the facility operation, including regulation schedules, transmission and generation constraints, and emergencies. Even if there are unforeseen communication issues with the operators, the transfer function model developed and the different factors contributing to 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 operator's actions rather than natural inflow process. Therefore, 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 document:

- 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 the 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 is 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 such as feature importance 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.

7. REFERENCES

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

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. (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. (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.

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.

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.

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).

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.

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

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.

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