

©Copyright 2018

Nam Jiang Song

Generation Forecasting for Small Run-of-the-River Hydroelectric Systems Using Statistical Learning Models

Nam Jiang Song

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Sciences in Electrical Engineering

University of Washington

2018

Reading Committee:

Daniel Kirschen, Chair

Baosen Zhang

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Generation Forecasting for Small Run-of-the-River Hydroelectric Systems Using Statistical Learning Models

Nam Jiang Song

Chair of the Supervisory Committee:
Daniel Kirschen
Electrical Engineering

With an increasing interest in constructing new run-of-the-river (ROR) hydroelectric generation over the more traditional reservoir-based hydroelectric systems, there is an increasing operational challenge due to the volatility of streamflow. The Snohomish County Public Utilities District (SnoPUD) has recently invested in the construction and operation of 3 new run-of-the-river projects in Northwestern Washington along Calligan Creek, Hancock Creek, and Youngs Creek. In order to effectively plan generation dispatch, SnoPUD has expressed interest in the development of an accurate forecasting tool to predict the generation capacity for these ROR systems.

The following research project aims to use statistical learning models, namely Hidden Markov Models (HMMs), to predict day-ahead generation capacities for the aforementioned ROR systems. These models are constructed using 12 years of historical streamflow data collected at the intake sites and precipitation data recorded at the National Oceanic and Atmospheric Administration (NOAA) Alpine Meadows station. Four methods of constructing the models are studied for their forecast accuracies, and are compared with the persistence model. Despite using only one set of observable variables, the HMMs are shown to have slight improvements in accuracy over the persistence model approach, which shows great optimism for future work.

TABLE OF CONTENTS

| | Page |
|---|------|
| List of Figures | iii |
| List of Tables | iv |
| Chapter 1: Introduction | 1 |
| 1.1 Run-of-the-River Hydroelectric Generation Operation | 4 |
| 1.2 Literature Review | 5 |
| Chapter 2: Site Analysis | 7 |
| 2.1 Instream Flow Requirements | 7 |
| 2.2 Calligan Creek | 8 |
| 2.3 Hancock Creek | 9 |
| 2.4 Youngs Creek | 11 |
| 2.5 Hydrology Summary | 11 |
| 2.6 Precipitation Data | 12 |
| Chapter 3: Model Description | 14 |
| 3.1 Hidden Markov Model Overview | 14 |
| 3.2 Model Construction and Implementation | 20 |
| Chapter 4: Model Testing and Results | 28 |
| 4.1 Model Results | 28 |
| 4.2 Discussion of Results | 40 |
| Chapter 5: Conclusion | 44 |
| 5.1 Future Work | 45 |
| Bibliography | 46 |

| | |
|---|----|
| Appendix A: Calligan Creek Project Overview | 49 |
| Appendix B: Hancock Creek Project Overview | 52 |
| Appendix C: Youngs Creek Project Overview | 55 |

LIST OF FIGURES

| Figure Number | Page |
|---|------|
| 1.1 Fuel Sources Used by SnoPUD [1] | 2 |
| 1.2 Hydropower Project Development Pipeline (As of December 31, 2017)[2] . . | 3 |
| 1.3 Operational Overview of a ROR Generator | 4 |
| 2.1 Generation Curve for Calligan Creek | 9 |
| 2.2 Generation Curve for Hancock Creek | 10 |
| 2.3 Generation Curve for Youngs Creek | 12 |
| 2.4 2013 Precipitation Data At Hancock Creek Intake (Blue) and NOAA Alpine Meadows Station (Red) | 13 |
| 3.1 Visual Representation HMM State Transitions | 17 |
| 3.2 Visual Representation of the Joint Event [3] | 18 |
| 3.3 Observation State V_i Distribution for Calligan Creek | 21 |
| 3.4 Hidden State S_i Distribution for Calligan Creek | 22 |
| 3.5 Visualized Pseudocode Flowchart | 23 |
| 3.6 Seasonal Hidden State S_i Distributions for Calligan Creek | 26 |
| 4.1 Baum-Welch Model Forecast Results (blue) Compared With Actual Historical Values (red) | 30 |
| 4.2 Aggregate Model Forecast Results (blue) Compared With Actual Historical Values (red) | 32 |
| 4.3 Seasonal Model Forecast Results (blue) Compared With Actual Historical Values (red) | 34 |
| 4.4 Probability Distribution Output of the Viterbi Algorithm | 35 |
| 4.5 Hierarchical Model Forecast Results (bbblue) Compared With Actual Historical Values (red) | 38 |
| 4.6 Maximum Model Accuracy Based On Decision Variable Breadth, Calligan Creek | 39 |
| 4.7 Transmission Overview | 41 |
| 4.8 Observation Sequence o_i Error Between NOAA and Intake Sites | 43 |

LIST OF TABLES

| Table Number | Page |
|---|------|
| 2.1 Penstock and Powerhouse Location Information for the ROR Systems | 7 |
| 2.2 IFRs (cfs) for Each Run-of-the-River Site | 8 |
| 2.3 Key Hydrological Differences | 11 |
| 3.1 Observation State V Definitions | 20 |
| 3.2 Hidden State S Definitions | 22 |
| 3.3 Date Periods Used to Construct and Validate the Hidden Markov Models . . | 23 |
| 3.4 Season Definitions for the Seasonal Model | 27 |
| 4.1 Theoretical Maximum Accuracy Using Hierarchical HMM with Increasing n | 40 |
| 4.2 Cumulative Accuracy Results for Running All 4 Models over the 3 ROR Systems | 40 |

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to the Hydro Research Foundation for providing the opportunity to work on this research project.

As well, the author wishes to express gratitude to Scott Spahr and Bill Harris at the Snohomish Public Utility District for all their support along the way.

Finally, the author wishes to thank his advisor Daniel Kirschen for his guidance and advice.

Chapter 1

INTRODUCTION

As the nationwide demand for renewable energy continues to grow and utilities start making the transition toward a network with high renewable energy penetration[4], the Snohomish County Public Utility District (SnoPUD) has maintained itself as a regional leader in renewable power generation. To meet the growing interest by its customers to achieve a 100% carbon-free power supply mix, SnoPUD has committed to continue investing in the construction of new renewable generation sources, and purchasing exclusively hydroelectric power from the Bonneville Power Administration (BPA). In addition to their existing research work in tidal and geothermal energy generation, some of the facilities SnoPUD have invested in include[5]:

- Small hydropower plants including the Jackson Hydroelectric Project, Woods Creek Hydroelectric Project, and Packwood Hydroelectric Project
- Purchases from the Hampton Lumber Mill co-generation plant
- Power purchase agreements with Northwest wind projects including White Creek Wind Project, Wheat Field Wind Project, and Hay Canyon Wind Project
- Operation of the Qualco Energy Biodigester which consumes biowaste fuel from local farms

While SnoPUD purchases more than 80% of its consumed power from the BPA, the majority of the fuel sources used by SnoPUD for generation is still predominantly hydroelectric generation (Figure 1.1). With the current SnoPUD generation capacity at a fraction of its

average demand (120MW generation capacity compared to 1448MW peak demand)[6], there is always a push for the construction of new hydroelectric generators to meet the increasing demand.

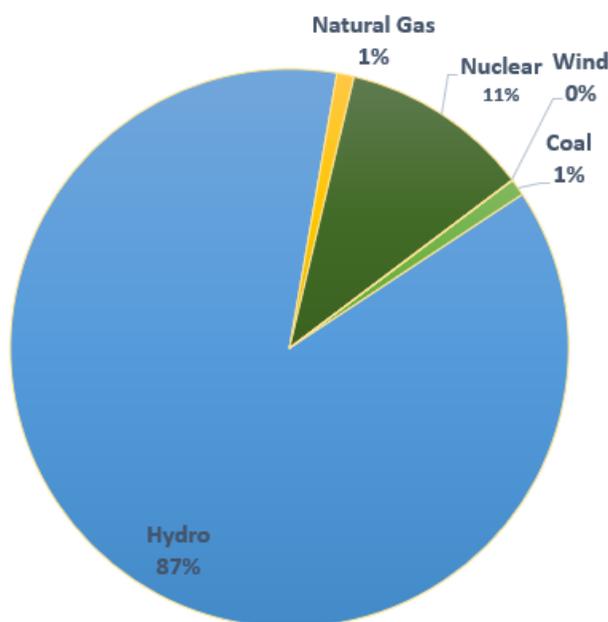


Figure 1.1: Fuel Sources Used by SnoPUD [1]

Of the 118 new nationwide hydropower plants that have started operation since 2006, SnoPUD has led the nation in the new stream-reach development projects through the construction of new hydroelectric plants [2].

In 2015, SnoPUD acquired licenses from the Federal Energy Regulatory Commission (FERC) to start development of two new run-of-the-river (ROR) hydroelectric plants at Calligan and Hancock Creeks, located approximately 10 miles south of the pre-existing Youngs Creek plant¹. Through the usage of these plants, SnoPUD has been able to increase its generation capacity by up to 19MW. However, while using ROR generation has its environmental advantages, SnoPUD has expressed interest in tackling its operational challenges

¹While the construction of the Young Creek project began in the mid-1990s, the plant was not put online until 2011.

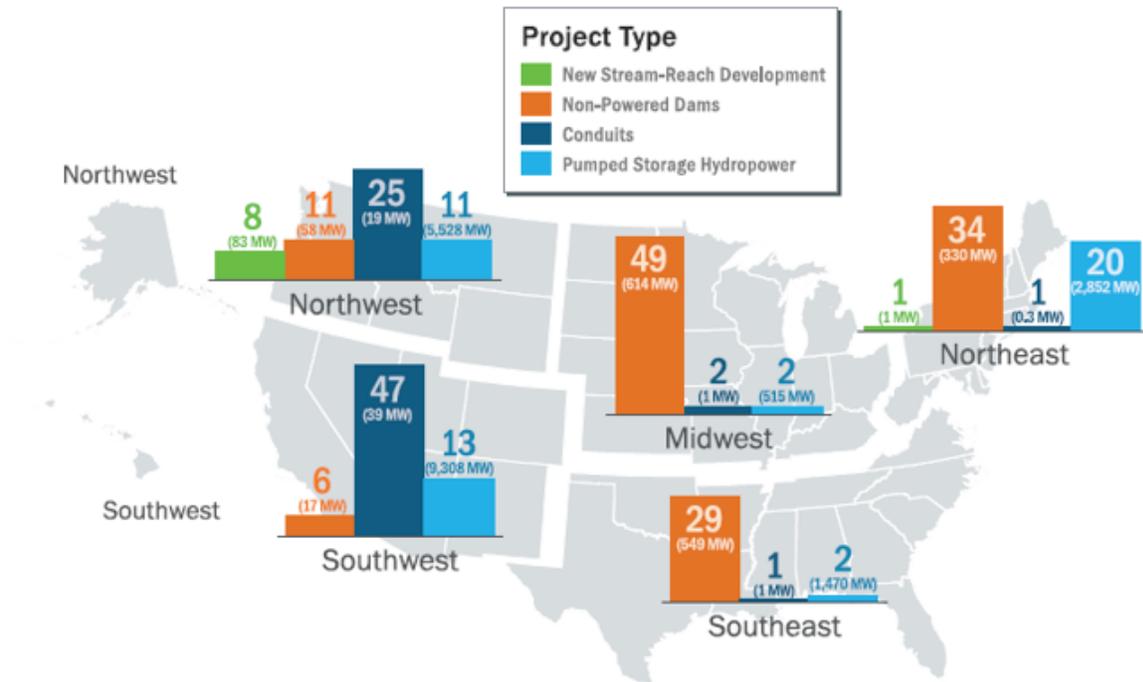


Figure 1.2: Hydropower Project Development Pipeline (As of December 31, 2017)[2]

through the development of a reliable forecasting tool for planning day-ahead dispatch levels and achieving cost-effective operations for the future.

This research project investigates the use of statistical learning models, namely the Hidden Markov Model (HMM), to provide an accurate day-ahead forecast for the amount of power that can be generated at the Calligan Creek, Hancock Creek, and Youngs Creek ROR plants. The models are constructed using from both the historical streamflow data collected at the intake sites provided by SnoPUD, and meteorological data collected by the National Oceanic and Atmospheric Administration (NOAA). Finally, the models are validated over a 12 to 20 month long period to determine the accuracy of the forecast model within the acceptable tolerance level provided by SnoPUD.

1.1 Run-of-the-River Hydroelectric Generation Operation

While the use of conventional hydroelectric generation in the United States contributes to the largest share of renewable energy consumption, the majority of the recent increase in hydropower generation capacity does not stem from the construction of new reservoir systems. Instead, this increase is achieved from the addition of hydropower generation equipment to pre-existing non-powered dams and conduits[2]. Of the new hydroelectric plants that have been constructed, most of them have been localized in the Northwestern region due to the availability of creeks and rivers that can be utilized for ROR generation.

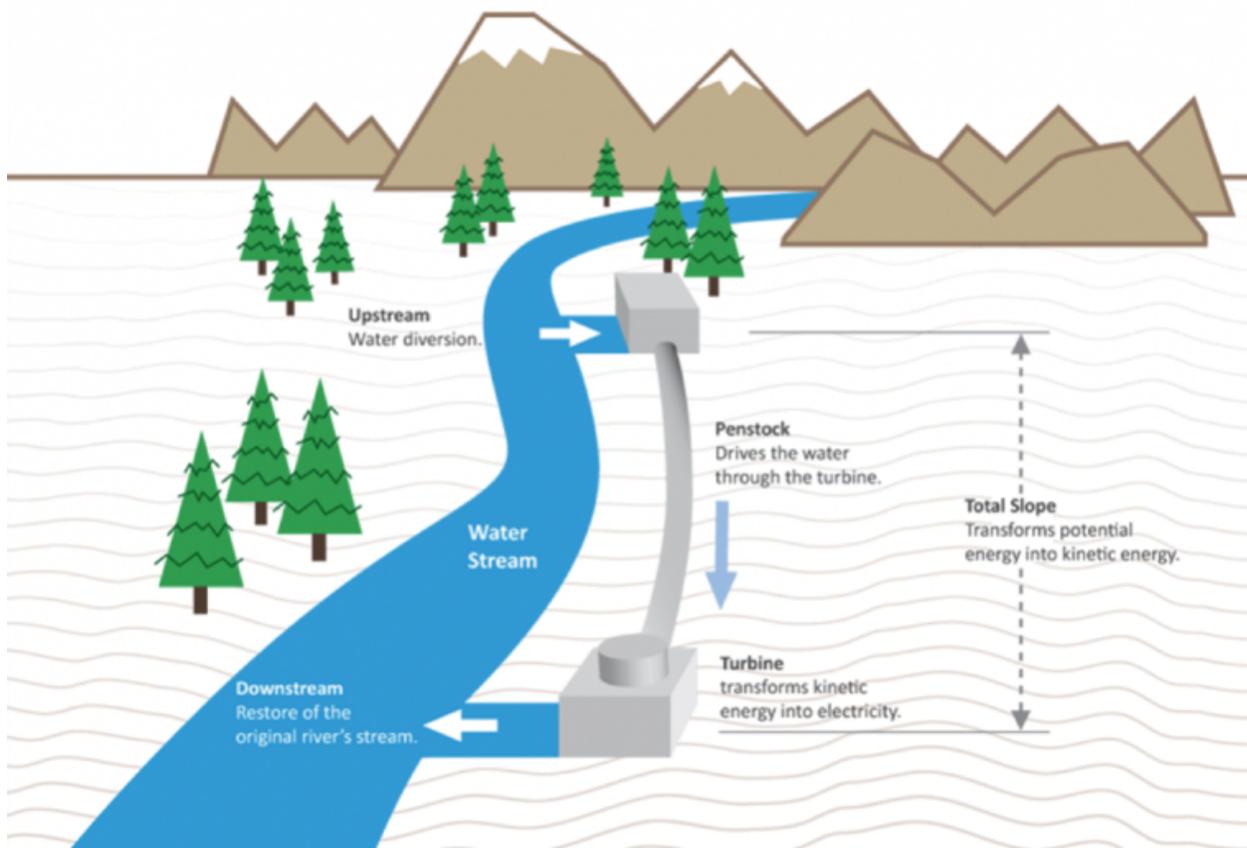


Figure 1.3: Operational Overview of a ROR Generator

ROR systems operate by diverting a fraction of the upstream streamflow through a penstock which diverts the streamflow towards a remote powerhouse. Once the water passes through the turbine, it is then restored to the original creek downstream. The effective head for ROR systems is equivalent to the altitude difference between the upstream and downstream sites.

The primary advantage of operating ROR systems over conventional reservoir-based systems is the avoidance of constructing a dam, which necessitates flooding of the surrounding lands to create the reservoir. By flooding the area, it can cause large environmental impacts by displacing the local ecosystem and any agricultural lands. In addition, dams can obstruct fish migration routes[7], and cause increased greenhouse gas emissions due to the buildup of decaying rotting vegetation caught in the dams[8].

However, the primary drawback of ROR generation is the volatility of the amount of streamflow that is available for generating energy. Since the streamflow is not sourced from a reservoir, the amount of water that can be diverted to the penstock is heavily dependent on both the weather and the amount of water available at the source (examples include lakes and snowpack). In addition, the amount of water that can be diverted is dependent on the Instream Flow Requirement (IFR) which places a restriction on the minimum amount of water which must remain in the creek at all times to sustain local ecosystems. Therefore, the development of a reliable forecasting tool is required to factor in both these operational challenges to predict the amount of flowrate and the amount of available power that can be generated at the upstream sites for each generation plant.

1.2 Literature Review

One of the simplest, and most common, methods for stochastic simulation of river flows is the use of an autoregressive-moving-average (ARMA) model². Models presented by Salas et al.[9] provide the earliest example of using ARMA models for hydrological modeling, and was

²Another similar model that has been utilized is the autoregressive integrated moving average (ARIMA) model.

soon expanded by incorporating seasonal variations[10]. While these initial ARMA models have been typically applied over a large time period by utilizing monthly-averaged flow, daily modeling of streamflow has been implemented by Can et al.[11].

Similarly, the use of Support Vector Machines (SVMs) has been investigated by Shabri et al.[12]. By using cross-validation and grid-search methods, the SVM parameters were detected to model the monthly streamflow behavior of Kinta River in Malaysia.

[13] has investigated the usage of neural networks for optimal operation of run-of-the-river adjustable speed power plants with axial-flow propeller turbines. In this approach, neural networks were constructed to simulate turbine behavior and turbine efficiency, and a maximum-efficiency tracking algorithm was implemented to determine the ideal position for the guide vanes.

Another model construction method was investigated by [14] using rainfall-runoff models. This approaches the problem from a purely hydrological perspective, where the unit hydrograph was constructed for two climatologically different catchments.

An alternate method to the hydrographical approach has been proposed by [15], where an autoregressive multinomial logistic model was constructed by incorporating precipitation data with temperature measurements, and tested over a catchment in the River Tees located in northeast United Kingdom.

The usage of HMMs in hydrography was proposed by Pender et al.[16], where the performance of using ARMA models and HMMs were compared by comparing the cumulative statistical characteristic of the streamflow measured over a period of one month. While this approach does not compare the effectiveness of HMMs over shorter time periods, the study also outlines a methodology to test the appropriateness of using HMMs to model the hydrographs.

Chapter 2

SITE ANALYSIS

The scope of this research project, as described in the previous section, encompasses 3 run-of-the-river hydroelectric plants owned and operated by SnoPUD: Calligan Creek, Hancock Creek, and Youngs Creek. Located within 15 miles from Snoqualmie Falls along the west side of the Cascades Range, these facilities provide consumers with a source of local, competitively-priced renewable power. In this section, hydrology analysis will be conducted for each facility to provide a brief background for each of the projects.

| ROR Station | Intake Location | Powerhouse Location |
|----------------|------------------------------|------------------------------|
| Calligan Creek | N47°36'3.54" W121°41'9.44" | N47°36'24.04" W121°42'34.54" |
| Hancock Creek | N47°34'20.81" W121°42'24.32" | N47°34'21.66" W121°42'40.26" |
| Youngs Creek | N47°47'3.75" W121°46'47.84" | N47°48'7.11" W121°49'22.07" |

Table 2.1: Penstock and Powerhouse Location Information for the ROR Systems

2.1 Instream Flow Requirements

The amount of streamflow that can be sent to the turbine for each plant is not only dependent on the weather but also on the Instream Flow Requirements (IFR) that are set by the State of Washington Department of Ecology. The main goal of implementing IFRs for each creek is to set seasonal requirements for the minimum amount of streamflow that must be retained in the creeks to protect instream resources, migrational navigation, and water quality[17]. The monthly IFRs for each stream, expressed in units of cubic feet per second (cfs), is outlined in Table 2.2.

| Period | Calligan Creek | Hancock Creek | Youngs Creek |
|-----------------|----------------|---------------|--------------|
| Jan. | 2.00 | 5.00 | 3.00 |
| Feb. | 2.00 | 5.00 | 3.00 |
| Mar. | 2.00 | 5.00 | 3.00 |
| Apr. | 2.00 | 5.00 | 3.00 |
| May 1-15 | 2.00 | 5.00 | 8.00 |
| May 15-Jun. 30 | 2.00 | 5.00 | 40.00 |
| Jul. 1-15 | 2.00 | 20.00 | 40.00 |
| Jul. 16-Aug. 30 | 2.00 | 20.00 | 22.00 |
| Sept. | 2.00 | 20.00 | 22.00 |
| Oct. | 2.00 | 20.00 | 3.00 |
| Nov. | 2.00 | 5.00 | 3.00 |
| Dec. | 2.00 | 5.00 | 3.00 |

Table 2.2: IFRs (cfs) for Each Run-of-the-River Site

Taking the IFRs for each site into account, the amount of effective streamflow that can be utilized for power generation Q_{eff} can be defined as

$$Q_{eff} = \max(0, Q - Q_{IFR})$$

where Q represents the upstream streamflow, and Q_{IFR} represents the IFR streamflow for the given day in the year.

2.2 Calligan Creek

The Calligan Creek hydroelectric plant operates by bypassing 1.4 miles of Calligan Creek through a 43.1 inch diameter penstock pipe, and discharged to a single 6MW Pelton generator, providing a gross head of 1032 feet. Sourced directly from Calligan Lake, the creek

is consistently supplied year-round. The generator output P_o (MW) can be represented as a function of the amount of the effective streamflow Q_{eff} (cfs), and is visualized on Figure 2.1.

$$P_o(Q_{eff}) = \begin{cases} 0 & \text{for } Q_{eff} < 5.25 \\ 0.0516Q_{eff} + 0.0004 & \text{for } 5.25 \leq Q_{eff} < 17.77 \\ 0.0761Q_{eff} - 0.0401 & \text{for } 17.77 \leq Q_{eff} < 85.82 \\ 6.587 & \text{for } Q_{eff} \geq 85.82 \end{cases} \quad (2.1)$$

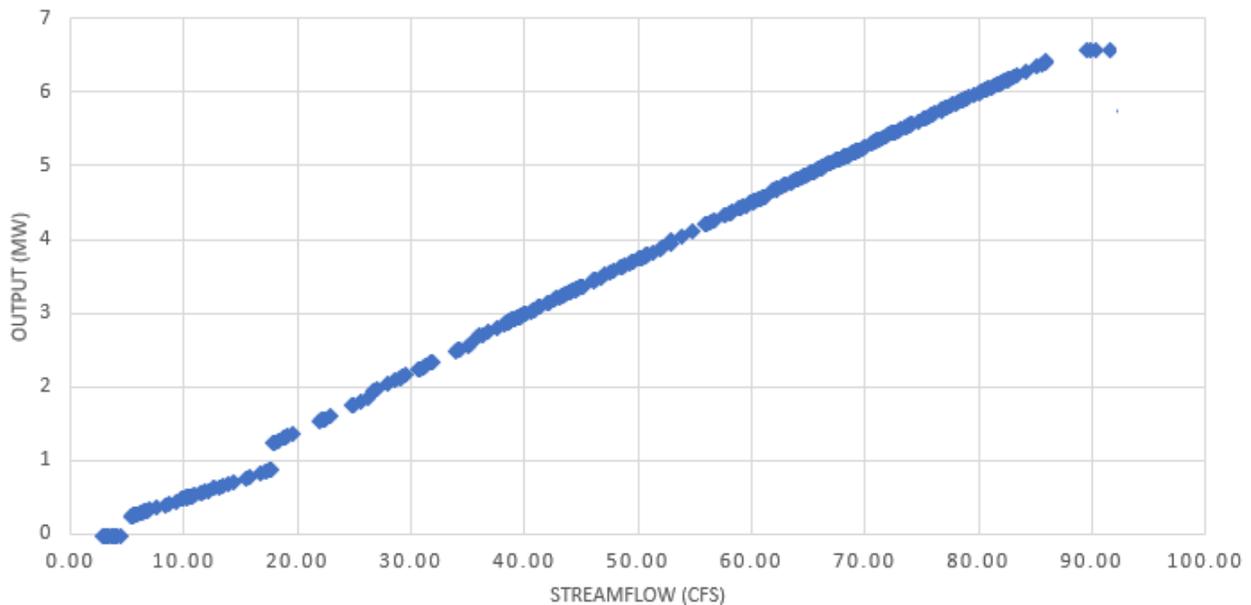


Figure 2.1: Generation Curve for Calligan Creek

2.3 Hancock Creek

Similar to Calligan Creek, the Hancock Creek hydroelectric plant is sourced from Lake Hancock. The diverted streamflow flows through a 41.1 inch diameter penstock pipe, and is

discharged to a single 6MW Pelton generator, providing a gross head of 1113 feet. Unlike Calligan Creek, Hancock Creek has a higher instream flow requirement during the summer months to accommodate for the increased fish population in the area. The generator output function for the plant $P_o(Q_{eff})$ is outlined below, and the generator curve is visualized in Figure 2.2.

$$P_o(Q_{eff}) = \begin{cases} 0 & \text{for } Q_{eff} < 4.18 \\ 0.0535Q_{eff} + 0.0004 & \text{for } 4.18 \leq Q_{eff} < 16.36 \\ 0.0817Q_{eff} - 0.0481 & \text{for } 16.36 \leq Q_{eff} < 80.90 \\ 6.511 & \text{for } Q_{eff} \geq 80.90 \end{cases} \quad (2.2)$$

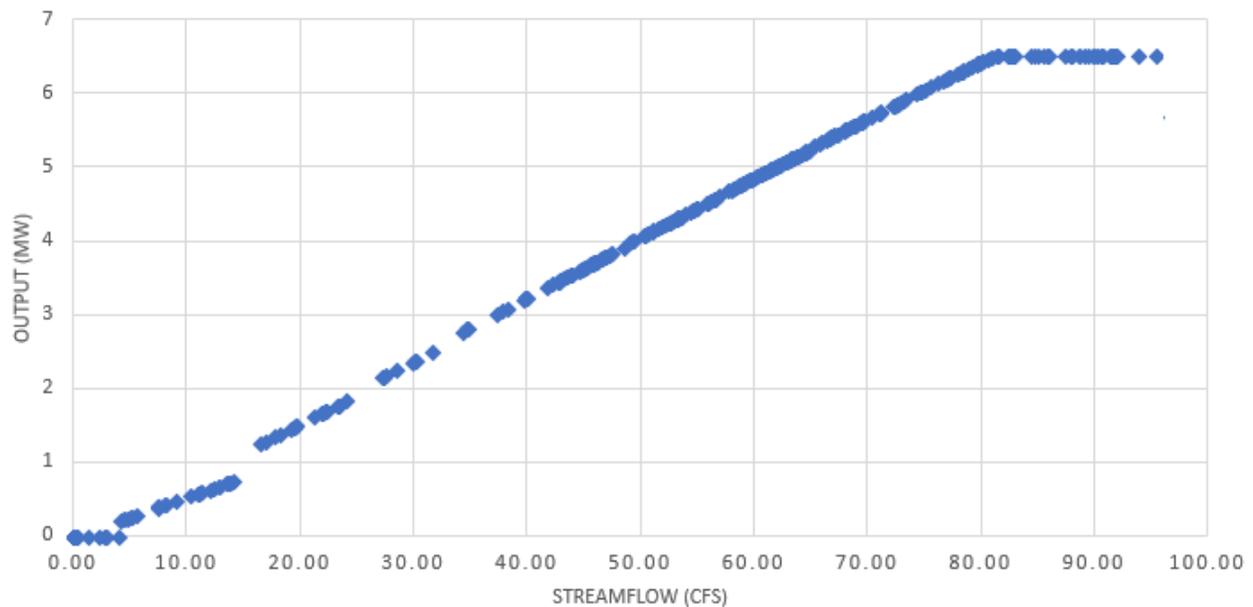


Figure 2.2: Generation Curve for Hancock Creek

2.4 Youngs Creek

Unlike the Calligan Creek and Hancock Creek projects, Youngs Creek is primarily sourced from snowpack in the Cascades. As a result, Youngs Creek has a flashier hydrograph, resulting in a shorter precipitation lag time, higher peak discharge, and faster drainage rate. The hydroelectric plant operates by bypassing streamflow through a 51 inch diameter penstock, and discharged through a single 7MW Pelton generator, providing a gross head of 930 feet. The generator output function $P_o(Q_{eff})$ is outlined below, and the generator curve is visualized in Figure 2.3.

$$P_o(Q_{eff}) = \begin{cases} 0 & \text{for } Q_{eff} < 6.33 \\ 0.0514Q_{eff} + 0.0012 & \text{for } 6.33 \leq Q_{eff} < 25.33 \\ 0.0661Q_{eff} - 0.0334 & \text{for } 25.33 \leq Q_{eff} < 118.67 \\ 7.700 & \text{for } Q_{eff} \geq 118.67 \end{cases} \quad (2.3)$$

2.5 Hydrology Summary

Although the ROR projects are located in close vicinity of each other, there are some key hydrological differences for each project. The key differences are summarized in Table 2.3.

| | Calligan Creek | Hancock Creek | Youngs Creek |
|--------------------------|----------------|---------------|--------------|
| Gross Head (ft) | 1032 | 1113 | 930 |
| Generation Capacity (MW) | 6 | 6 | 7 |
| IFR | Constant | Seasonal | Seasonal |
| Water Source | Lake | Lake | Snowmelt |

Table 2.3: Key Hydrological Differences

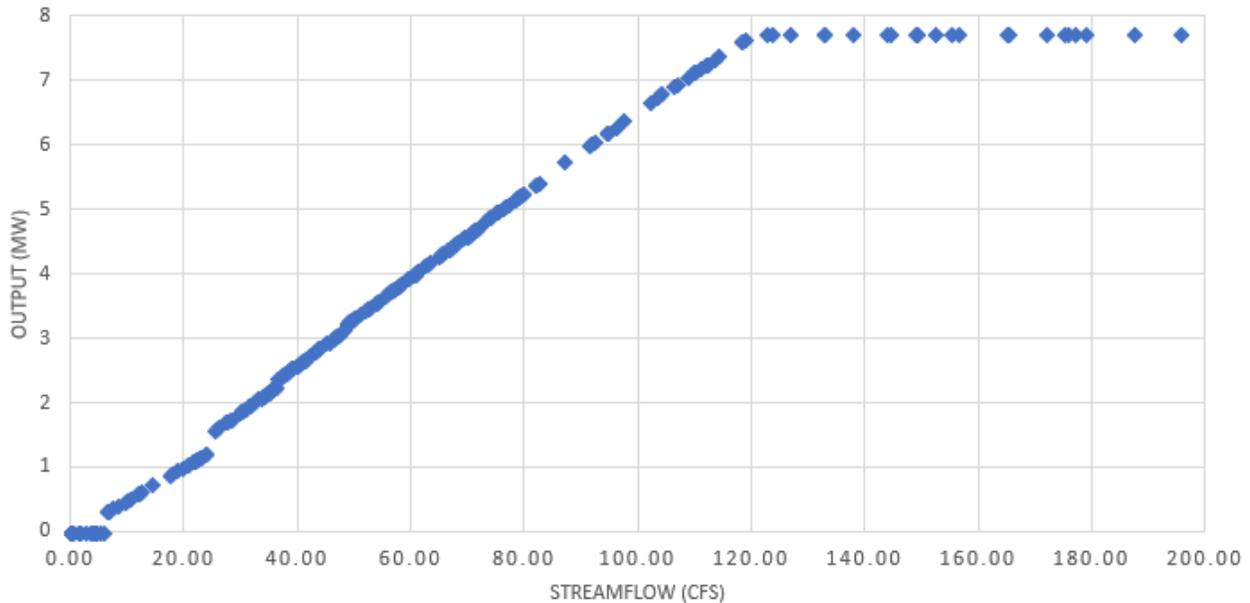


Figure 2.3: Generation Curve for Youngs Creek

2.6 Precipitation Data

Although approximately 50 years worth of daily streamflow data was collected by the intake sites, the amount of precipitation data spans a small fraction of the time frame. Although SnoPUD has installed weather stations at each of the intake sites capable of recording and collecting a multitude of meteorological data including precipitation and temperature, there were a few technical setbacks with the data collection process. Firstly, the meteorological data collection was began in 2010, which provides a shorter time span compared with the streamflow data. Secondly, there were large gaps in the precipitation data due to unexpected technical issues and loss of connection to the hub.

Therefore, precipitation data was collected from the Alpine Meadows National Oceanic Atmospheric Administration (NOAA) weather station (N47°46'48" W121°42'0"). This site was chosen due to its proximity to the intake sites, and its location on the same side of the Cascade Range, albeit situated 1000ft above the intake sites. As visualized in Figure

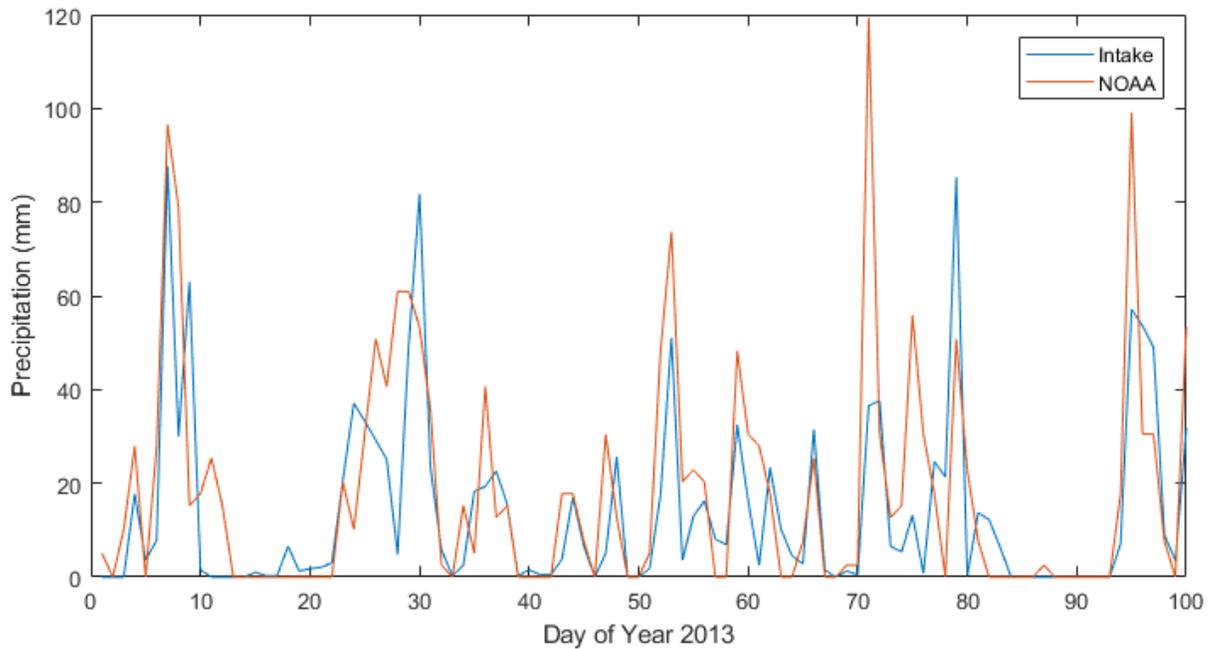


Figure 2.4: 2013 Precipitation Data At Hancock Creek Intake (Blue) and NOAA Alpine Meadows Station (Red)

2.4, the precipitation levels recorded at the intake sites follow the NOAA data relatively well. Therefore, utilizing the NOAA precipitation data instead of the intermittent intake precipitation data will be sufficient for the modelling process.

Chapter 3

MODEL DESCRIPTION

3.1 Hidden Markov Model Overview

3.1.1 Markov Chains

The use of a discrete Markov chain is based on the probabilistic nature of the graph representation of the system, wherein a sequence of observations can be summarized and classified by its statistical characteristics. The discrete-time Markov chain can be described as the sequence of states

$$Q = q_1 q_2 \dots q_T$$

where each event is from a set of N possible states in the model

$$q_t \in S = \{s_1 s_2 \dots s_N\}$$

wherein each state s_i represents the different discrete observations that can be made in the system¹.

Using this state sequence Q , a transition probability matrix A can be defined, where each element a_{ij} represents the probability of transitioning from state i to state j

$$a_{ij} = P[q_{t+1} = s_j | q_t = s_i] \quad \forall 1 \leq i, j \leq N$$

As well, under the Markov Assumption, it is established that

¹For example, a set of states S that can be used to describe weather in a local region can be $S = \{\text{sunny, cloudy, rainy}\}$

$$0 \leq a_{ij}$$

and

$$\sum_{j=1}^N a_{ij} = 1$$

which satisfies the primary statistical characteristics of the Markov Chain.

3.1.2 Hidden Markov Models

While a Markov chain is useful for computing probabilities for a sequence of events that can be directly observed, Hidden Markov processes consider the hidden observations and behavior that can occur in the system². The Hidden Markov Model (HMM) expands on the Discrete Markov Chain by relating 2 sets of states: the set of N hidden states

$$S = \{s_1, s_2, \dots, s_N\}$$

and set of M observation states

$$V = \{v_1, v_2, \dots, v_M\}$$

The HMM is inherently a doubly stochastic process, which consists of an underlying sequence which is hidden from the observer

$$Q = q_1 q_2 \dots q_T$$

where $q_t \in S$, and an observable sequence

$$O = o_1 o_2 \dots o_T$$

²For a great example involving the prediction of missing weather data with an assiduously recorded ice cream log, see [18]

where $o_t \in V$. As the HMM is an expansion of the Markov Chain, the HMM $\lambda = (A, B)$ is governed by 2 matrices: the transition probability matrix A and an emission probability matrix B .

Similar to the Markov chain model, A describes the transition between the hidden states in the sequence

$$a_{ij} = P[q_{t+1} = s_j | q_t = s_i] \quad \forall 1 \leq i, j \leq N$$

The emission matrix B on the other hand describes the probability that of a particular observation state given the hidden state at time index t

$$b_j(k) = P[v_{k,t} | q_t = s_j] \quad 1 \leq j \leq N, 1 \leq k \leq M$$

and with matrix characteristics

$$0 \leq b_j(k)$$

and

$$\sum_{k=1}^M b_j(k) = 1$$

A visual representation of the state transitions is shown in Figure 3.1.

Using the HMM model, the three main types of problems of interest are[19]:

- **Likelihood Problem:** Given an HMM $\lambda = (A, B)$ and an observation sequence O , determine the likelihood $P(O|\lambda)$
- **Decoding Problem:** Given an observation sequence O and an HMM $\lambda = (A, B)$, discover the best hidden state sequence Q
- **Learning Problem:** Given an observation sequence O and the set of states in the HMM, learn the HMM parameters A and B

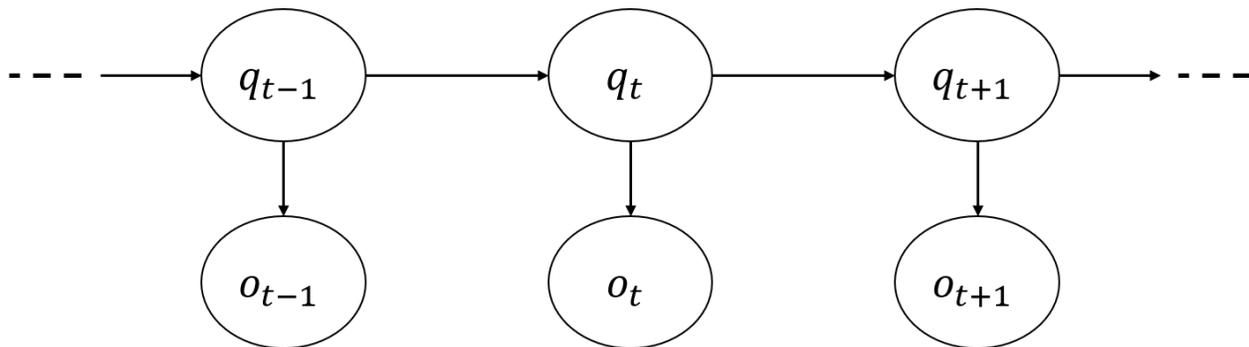


Figure 3.1: Visual Representation HMM State Transitions

For the purposes of this project, methods to solve the second and third problems will be outlined in the subsequent sections.

3.1.3 Decoding Problem

For any model which contains hidden variables, the decoding task is defined as the task of determining the sequence of hidden variables which is responsible for the sequence of observations. In other words, the task of the decoder is, given an HMM $\lambda = (A, B)$ and a sequence of observations $O = o_1 o_2 \dots o_T$, determine the most probable sequence of states $Q = q_1 q_2 \dots q_T$.

The most common decoding algorithm being implemented is the Viterbi algorithm [20], which uses dynamic programming trellis. At iteration t , the Viterbi algorithm determines the most likely trellis cell $v_t(j)$ to determine the probability that $q_t = S_j$ based on the state path estimated for q_1, q_2, \dots, q_{t-1} . This sequence can be formally expressed as

$$v_t(j) = \max_{q_0, q_1, \dots, q_{t-1}} P(q_0, q_1, \dots, q_{t-1}, o_1, o_2, \dots, o_t, q_t = j | \lambda)$$

Under the Markov assumption, this expression can be reduced to a simplified recursive function

$$v_t(j) = \max_{i=1}^N v_{t-1}(i) a_{ij} b_j(o_t)$$

where $v_{t-1}(i)$ represents the Viterbi path probability from the previous time step, a_{ij} represents the transition probability from state q_i to state q_j , and $b_j(o_t)$ represents the state observation likelihood of the observation o_t given the current state S_j .

3.1.4 Learning Problem

In solving the Learning Problem, the goal of the algorithm is to, given an observation sequence O and the set of possible states in the HMM, the HMM parameters A and B should be estimated. In this case, it is assumed that the vocabulary of the hidden states S is defined beforehand.

The standard algorithm for HMM training is the Baum-Welch algorithm [21], which uses an iterative algorithm to determine the best estimates for the elements \hat{a}_{ij} and $\hat{b}_j(k)$.

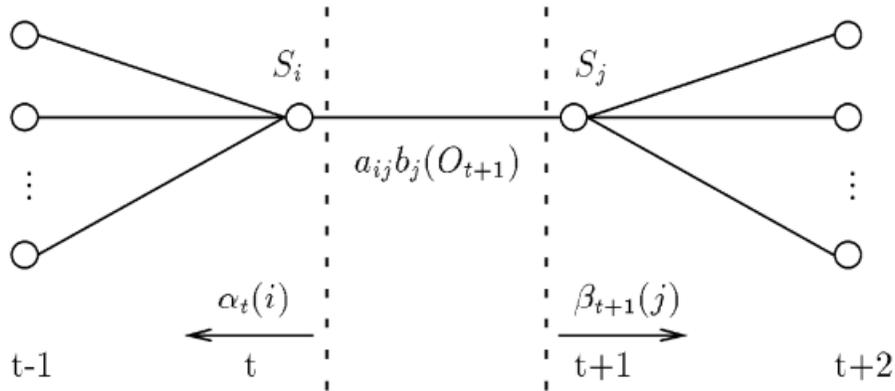


Figure 3.2: Visual Representation of the Joint Event [3]

In solving for the parameter estimation, it is important to examine the joint event between states S_i and S_j as illustrated in Figure 3.2. During iteration t , the joint event that connects the two states, $a_{ij} b_j(O_{t+1})$ is governed by the forward variable

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = S_i | \lambda)$$

which expresses the probability of the partial observation sequence of $o_1 o_2 \dots o_t$ given state $q_t = S_i$ and model λ , and the backward variable

$$\beta_t(i) = P(o_{t+1} o_{t+2} \dots o_T | q_t = S_i, \lambda)$$

which represents the probability of the partial observation sequence $o_{t+1} o_{t+2} \dots o_T$ state $q_t = S_i$ given model λ .

Using the forward and backward variables, the behavior of the joint event $\xi_t(i, j)$ that connects states S_i at iteration t and S_j at iteration $t + 1$ can be expressed as

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)}$$

As well, the state variable $\gamma_t(i)$ which describes the probability of being in state S_i at time t can be represented as

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) = \sum_{j=1}^N \xi_t(i, j)$$

By combining the joint events and state variables, the Baum-Welch algorithm estimates the state transition and emission probabilities:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad 1 \leq i \leq N, 1 \leq j \leq N$$

$$\hat{b}_j(k) = \frac{\sum_{t=1}^T \sum_{o_t=v_k} \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad 1 \leq j \leq N, 1 \leq k \leq M$$

While the Baum-Welch algorithm outlines an iterative approach to calculating the HMM parameters, the convergence the the algorithm is not guaranteed. For example, if a sudden torrential downpour is recorded during the summer months, the Baum-Welch algorithm treats this anomalous event as a common occurrence, which skews the resulting λ . Therefore,

it is also important to initialize the algorithm with a set of initial state probabilities and define any state transitions that are physically impossible.

3.2 Model Construction and Implementation

3.2.1 Model Definitions

As outlined in the previous section, the role of the HMM is to relate the set of observation states V with the set of hidden states S . The observation states V is defined using the precipitation data recorded at the Alpine Meadows NOAA station. Each observation state v_i was based primarily on the precipitation amount (mm). As well, the number of consecutive days with 0mm precipitation was considered to account for the time delay associated with the rainfall-runoff model. The V states are defined in Table 3.1.

| | | | | | | | | | | |
|-------------------------------|---|---|---|----------|-----|------|-------|-------|-------|-----------|
| State v_i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Precipitation (mm) | 0 | 0 | 0 | 0 | 0-5 | 5-10 | 10-20 | 20-30 | 30-50 | ≥ 50 |
| Consec. Days with 0mm Precip. | 1 | 2 | 3 | ≥ 4 | | | | | | |

Table 3.1: Observation State V Definitions

As an example, a histogram of the observation states for Calligan Creek is visualized in Figure 3.3. Due to the relatively high count for $V_i = 4$, it is indicative of a long dry period during the summer months.

The set of hidden states S for this research project is defined to be the difference of the daily potential generating capacity (MW) between 2 consecutive days. In order to comply with the planning department at SnoPUD, the generating capacity values are rounded down to the nearest integer using the floor function

$$\hat{x} = \lfloor x \rfloor$$

Each state can be formally represented as

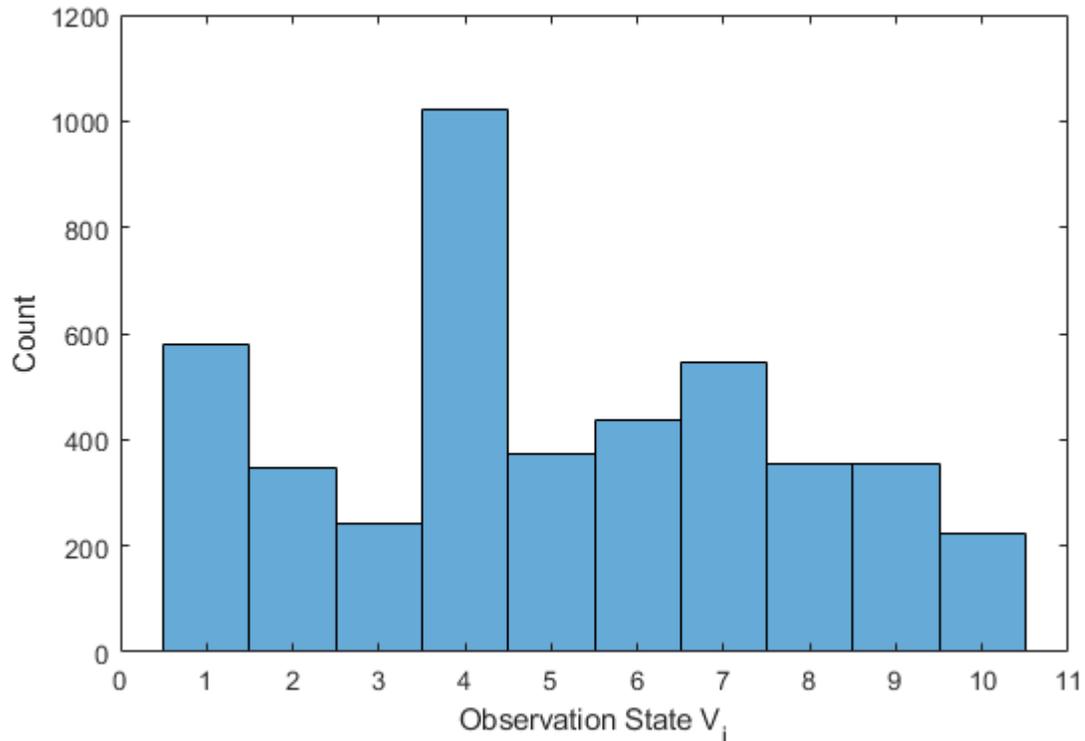


Figure 3.3: Observation State V_i Distribution for Calligan Creek

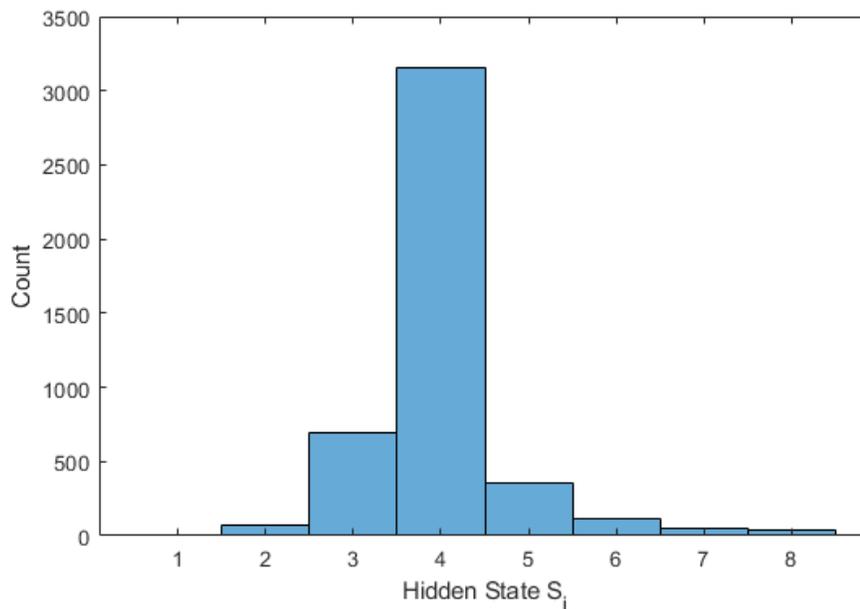
$$S_i = \Delta[P_o] + C = [P_{o,t}] - [P_{o,t-1}] + C$$

where C represents the zero-offset. The S definitions for each ROR station are outlined in Table 3.2.

Likewise, the histogram for the S distribution for Calligan Creek (Figure 3.4) shows a noticeable peak at state 4, which also corresponds to long periods during the summer season where there is insufficient streamflow for generation.

Therefore, the original model problem can be restated as (as originally structured by Eisner [18]): Given a sequence of precipitation observations at the Alpine Meadows Station V , forecast the hidden sequence of S which corresponds to the available generation output

| State S_i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------------------|-----------|----|----|----|---|---|---|----------|----------|
| Calligan Creek (MW) | ≤ -3 | -2 | -1 | 0 | 1 | 2 | 3 | ≥ 4 | |
| Hancock Creek (MW) | ≤ -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | ≥ 4 |
| Youngs Creek (MW) | ≤ -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | ≥ 4 |

Table 3.2: Hidden State S DefinitionsFigure 3.4: Hidden State S_i Distribution for Calligan Creek

for the 3 run-of-the-river hydroelectric systems.

3.2.2 Model Construction

The historical streamflow and precipitation data were partitioned into two sets: a construction set and a validation set. Table 3.3 outlines the time period for each creek that were used for each set.

| | Construction Period | Validation Period |
|----------------|----------------------|--------------------|
| Calligan Creek | 10/1/1994-12/31/2006 | 1/1/2007-9/29/2008 |
| Hancock Creek | 10/1/1994-12/31/2006 | 1/1/2007-9/29/2008 |
| Youngs Creek | 10/1/1994-3/31/2000 | 4/1/2000-3/31/2001 |

Table 3.3: Date Periods Used to Construct and Validate the Hidden Markov Models

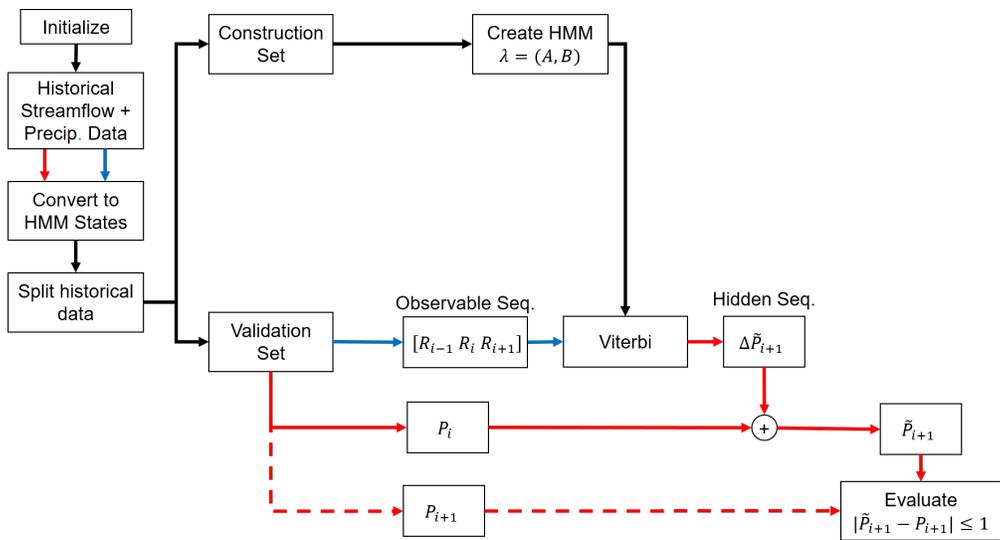


Figure 3.5: Visualized Pseudocode Flowchart

Using the construction set, the HMM matrices $\lambda = (A, B)$ were constructed using 3 different methodologies: Baum-Welch model, aggregate model, and seasonal models. A generalized flowchart is visualized in Figure 3.5

Model 1: Baum-Welch Model

Using this methodology, the Baum-Welch algorithm was used to produce the best estimates for the model. It should be noted that, because no filtering was done for the construction set to remove any anomalous data, convergence was not guaranteed. The steps taken to

construct the Baum-Welch model are as follows:

1. Calculate the effective streamflow Q_{eff} that can be utilized for generation
2. Generate the set of hidden states Q using the generator curve $P_o(Q_{eff})$ and Table 3.2
3. Generate the set of observable states O using Table 3.1
4. Construct the best estimate HMM (A, B) by running the Baum-Welch algorithm until either convergence or reaches timeout

Model 2: Aggregate Model

In the aggregate model, a normalized histogram of O and Q during the construction period was utilized. For the transition matrix \hat{A} , each element represents the count of all the transitions between indexes t and $t+1$. Thus, each of the elements represents the conditional probability

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$$

By analyzing through all of the dates in the construction period, the model $\lambda = (A, B)$ is constructed as

$$a_{ij} = \frac{\sum_{t=1}^{T-1} x_{ij,t}}{\sum_{i=1}^N \sum_{t=1}^{T-1} x_{ij,t}}$$

$$b_{ij} = \frac{\sum_{t=1}^T y_{ij,t}}{\sum_{i=1}^N \sum_{t=1}^T y_{ij,t}}$$

where T represents the number of dates in the construction period, and x and y represents the count, where

$$x_{ij,t} = \begin{cases} 1 & \text{for } q_{t+1} = s_j | q_t = s_i \\ 0 & \text{otherwise} \end{cases}$$

and

$$y_{ij,t} = \begin{cases} 1 & \text{for } o_t = v_j | o_t = s_i \\ 0 & \text{otherwise} \end{cases}$$

Similar to the steps for constructing the Baum-Welch model, the aggregate model is constructed as follows:

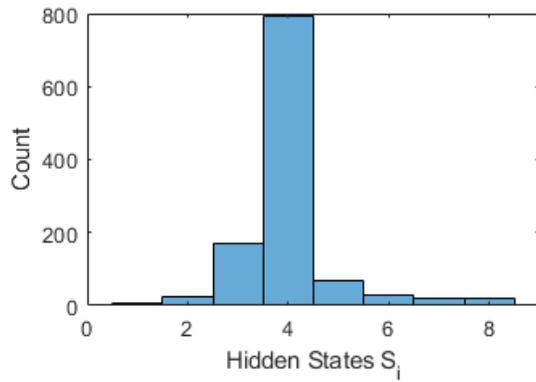
1. Calculate the effective streamflow Q_{eff} that can be utilized for generation
2. Generate the set of hidden states Q using the generator curve $P_o(Q_{eff})$ and Table 3.2
3. Generate the set of observable states O using Table 3.1
4. Construct (A, B) by building row-normalized histograms

Model 3: Seasonal Model

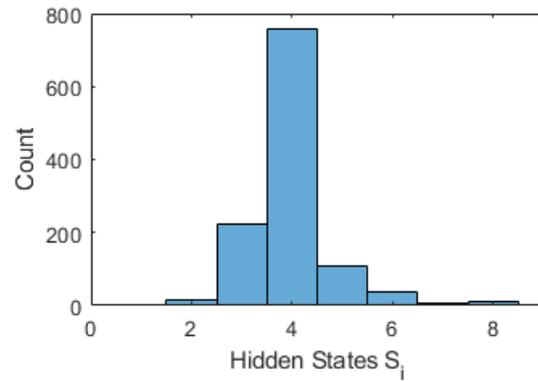
Due to the effect of seasonal variability for each of the hydrological systems, it is also useful to construct 4 different aggregate models based on the seasons. By separating the data, each seasonal effect can be isolated from each other; the effects of the dry period during the summer can now be separated from the winter months. The seasons were defined as per Table 3.4. Although the hidden states were separated into its respective seasons (Figure 3.6), it should be noted that the state corresponding to $\Delta[P_o] = 0\text{MW}$ still comprise the majority of the hidden states.

The steps for building the seasonal model are:

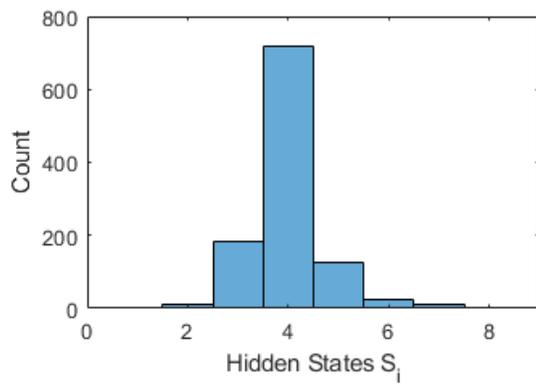
1. Calculate the effective streamflow Q_{eff} that can be utilized for generation
2. Generate the set of hidden states Q using the generator curve $P_o(Q_{eff})$ and Table 3.2
3. Generate the set of observable states O using Table 3.1
4. Separate the states O and Q based on the season each data falls on



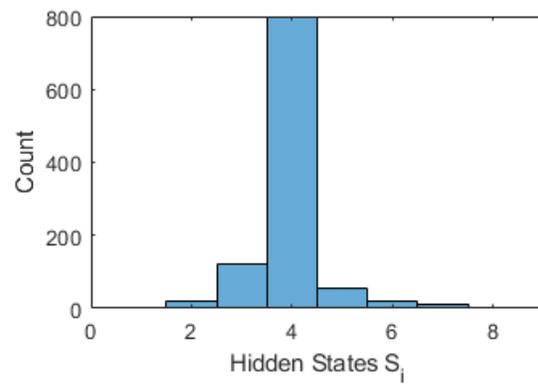
(a) Winter



(b) Spring



(c) Summer



(d) Autumn

Figure 3.6: Seasonal Hidden State S_i Distributions for Calligan Creek

| Season | Months | | |
|--------|--------|-----|-----|
| Winter | Dec | Jan | Feb |
| Spring | Mar | Apr | May |
| Summer | Jun | Jul | Aug |
| Autumn | Sep | Oct | Nov |

Table 3.4: Season Definitions for the Seasonal Model

5. Construct 4 different models $(A_1, B_1), (A_2, B_2), (A_3, B_3), (A_4, B_4)$ using the row-normalized histogram approach outlined in the aggregate model

3.2.3 Model Validation

Using the validation set, the accuracy of the model forecast is compared with the historical values. The validation steps are outlined as follows:

1. Acquire the generation capacity $P_{o,i}$
2. Acquire the last 2 days of rainfall data $\begin{bmatrix} R_{i-1} & R_i \end{bmatrix}$ to create the observable sequence
3. Append the rainfall forecast for the following day \tilde{R}_{i+1}
4. Run the Viterbi algorithm using the model $\lambda = (A, B)$ to get the most probable hidden state sequence, \tilde{s}_{i+1}
5. Using the last element from the Viterbi algorithm, compute the forecast generation capacity $P_{o,i+1} = P_{o,i} + \Delta P_{o,i+1}$
6. Calculate the accuracy value based on $\pm 0\text{MW}$ and $\pm 1\text{MW}$ tolerance levels

Chapter 4

MODEL TESTING AND RESULTS

4.1 *Model Results*

For each the methodologies outlined in the previous chapter, the results of the validation algorithm was compared with a benchmark analysis model. The accuracy of the models were calculated by evaluating if the forecast values were within the tolerance bound of the historical values.

4.1.1 *Benchmark Model*

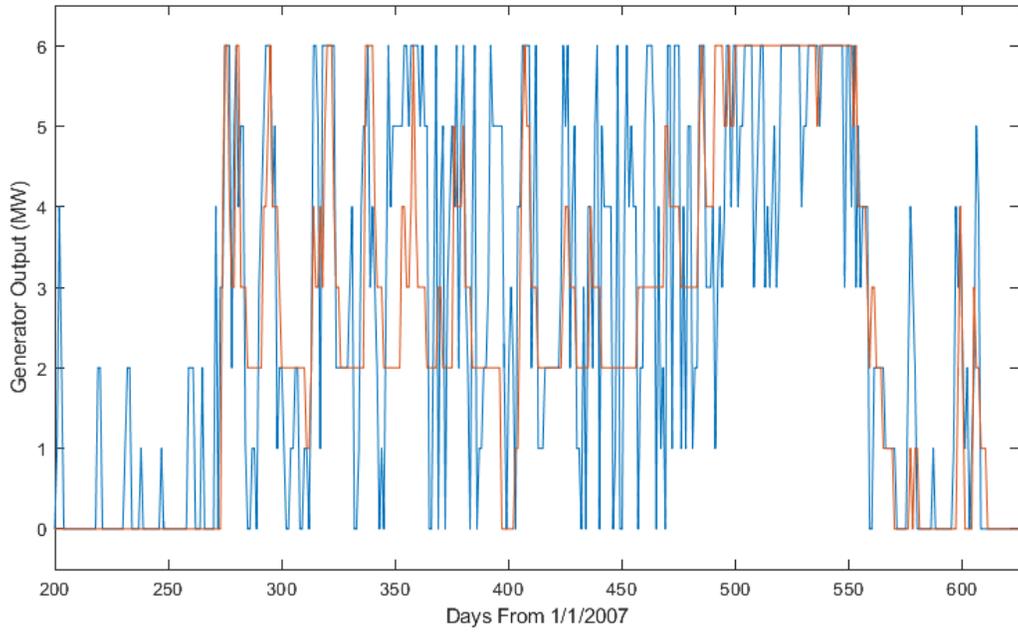
The benchmark case was chosen to be the persistence model, where the recorded value for index t is used as the forecast value for the following time index $t + 1$:

$$\hat{p}_{t+1} = p_t$$

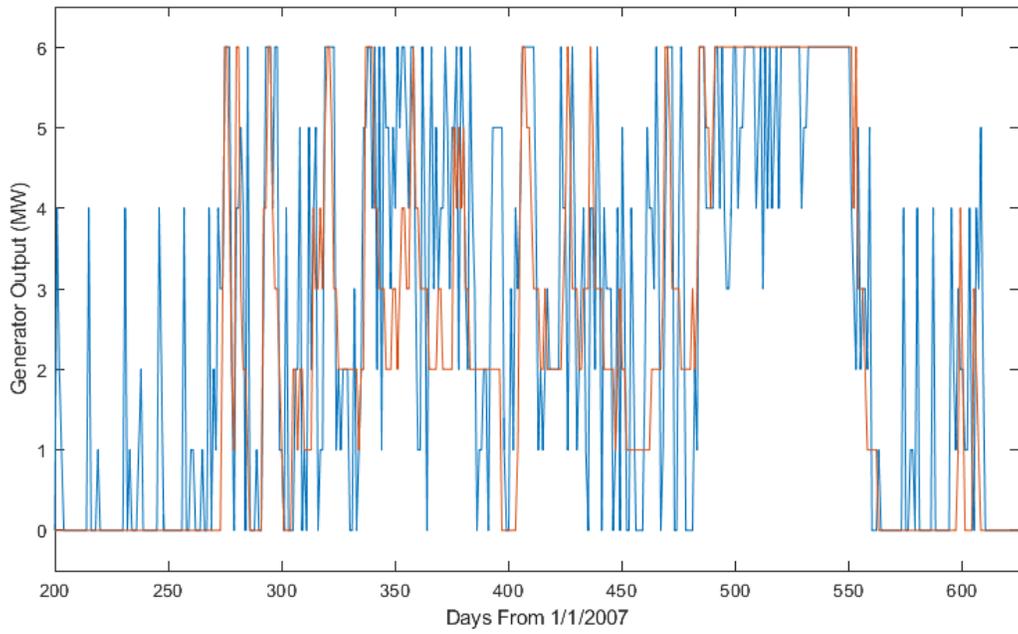
This methodology is commonly used in forecasting short-term forecasting for other types of variable renewable energy sources, particularly in wind generation where the persistence model has shown to be relatively effective in the short time ranges[22]. Hence, any forecast models that is developed should be tested against this classical benchmark of persistence method to check how much it can improve over the persistence derived forecasts[23].

4.1.2 *Baum-Welch Model*

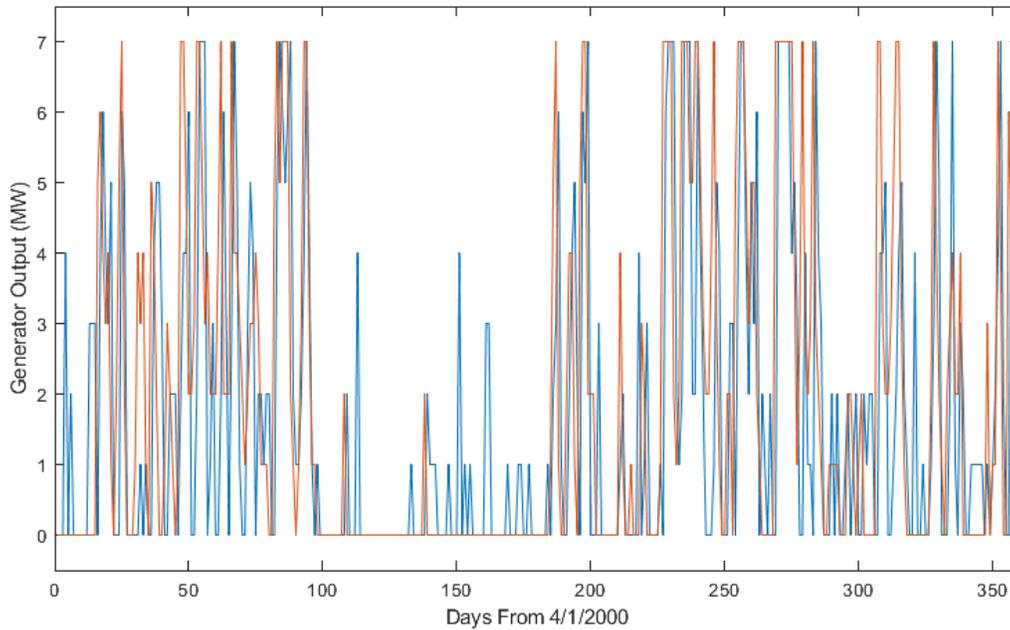
Using the Baum-Welch model, it is visually evident that the forecast values do not match the historical values very well. Even during the summer season (days 200-240), the model makes nonzero generation capacity predictions despite the low levels of precipitation.



(a) Calligan Creek



(b) Hancock Creek



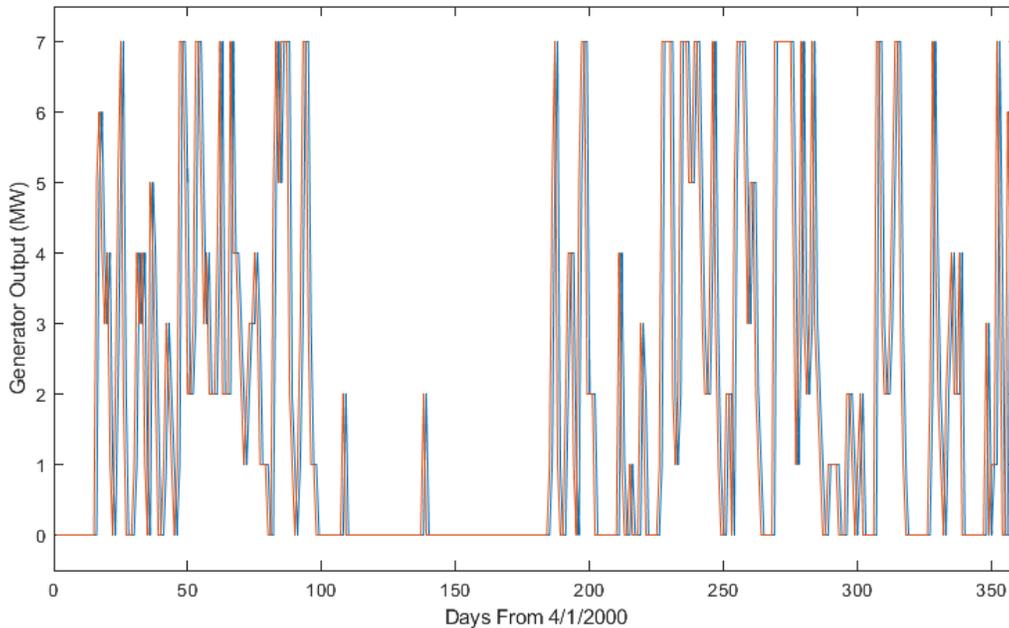
(c) Youngs Creek

Figure 4.1: Baum-Welch Model Forecast Results (blue) Compared With Actual Historical Values (red)

The forecast results are indicative of the inherent drawbacks with the Baum-Welch model. The performance of the model is dependent on knowing the ideal distribution for each system beforehand, which is not always guaranteed. In addition, initializing the algorithm by defining state transitions which are physically impossible is required to assist the convergence of the algorithm.

4.1.3 Aggregate Model

Using the aggregate model, it is evident that, on first inspection, the plots match up perfectly. However, upon closer inspection, the forecast values are actually offset by exactly 1 day. This behavior implies that the aggregate model behaves identical to the persistence model.



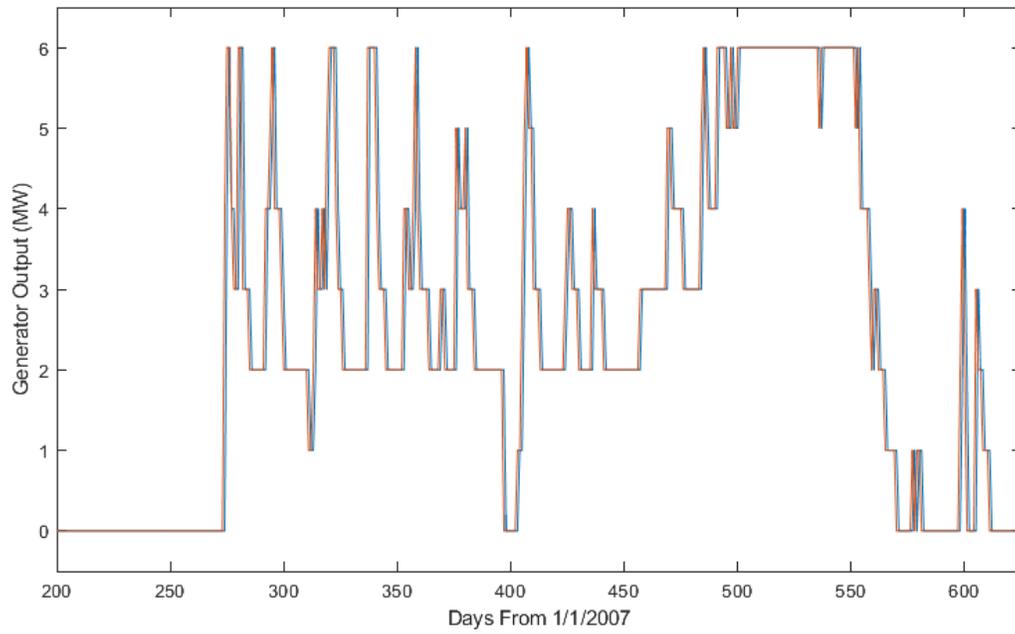
(c) Youngs Creek

Figure 4.2: Aggregate Model Forecast Results (blue) Compared With Actual Historical Values (red)

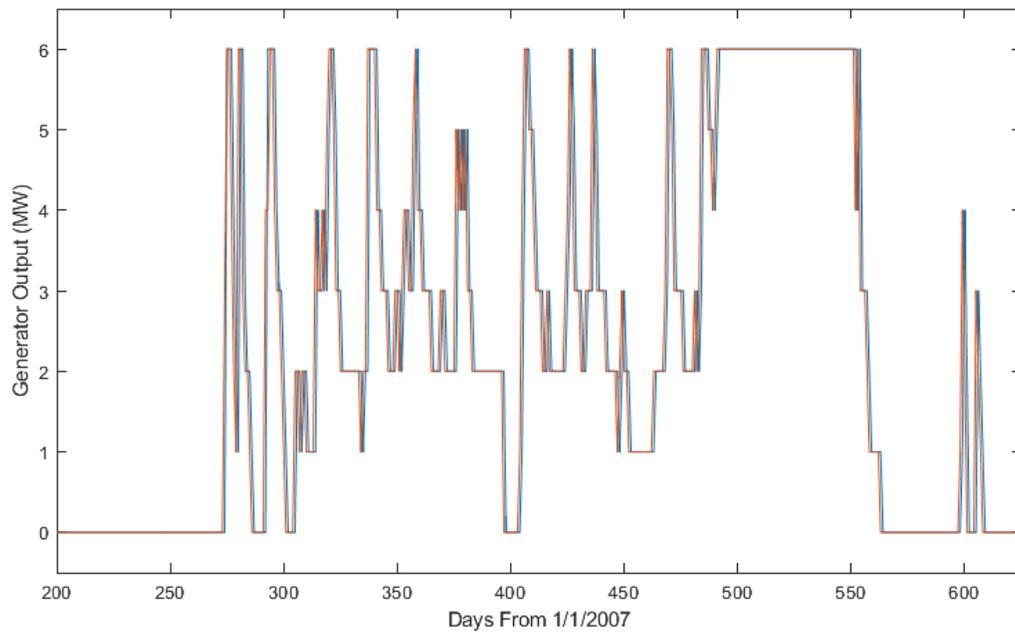
This behavior in the aggregate model is primarily due to the unbalanced hidden state distribution. For example, the state distribution for Calligan Creek has a large peak at state 4. This state corresponds to $\Delta P_o = 0$ MW, which is identical to the definition for the persistence model. Therefore, when the Viterbi algorithm is run, the trellis end up favoring the state with the largest probability which is coincidentally same as the condition for the persistence model.

4.1.4 Seasonal Model

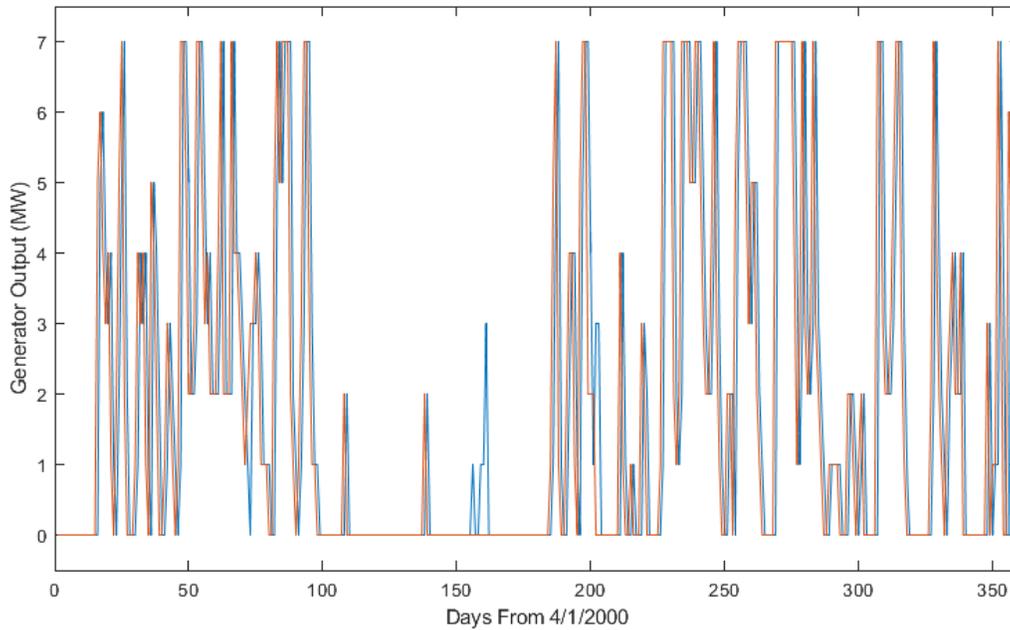
By utilizing the four different HMMs, the seasonal models were able to capture the seasonal effects, resulting in a slight increase in accuracy.



(a) Calligan Creek



(b) Hancock Creek



(c) Youngs Creek Creek

Figure 4.3: Seasonal Model Forecast Results (blue) Compared With Actual Historical Values (red)

Although the improvement in accuracy using the seasonal model is difficult to spot visually, the slight improvement is apparent in Table 4.1.

4.1.5 Hierarchical Model

Based on the concept of Hierarchical Hidden Markov Models (HHMMs)[24], the output of the Viterbi algorithm is modified from a single variable output to a probability distribution for all the possible states. For example, during an iteration of the Viterbi algorithm performed at Calligan Creek, the correct state transition (highlighted in red) has the second largest probability value (p-value). In the current form of the Viterbi algorithm, the forecast returns the state corresponding to the largest p-value, which is attributed to the initial hidden state

S_i distribution for Calligan Creek. Therefore, a method of extracting the correct state transition given the probability distribution is be utilized.

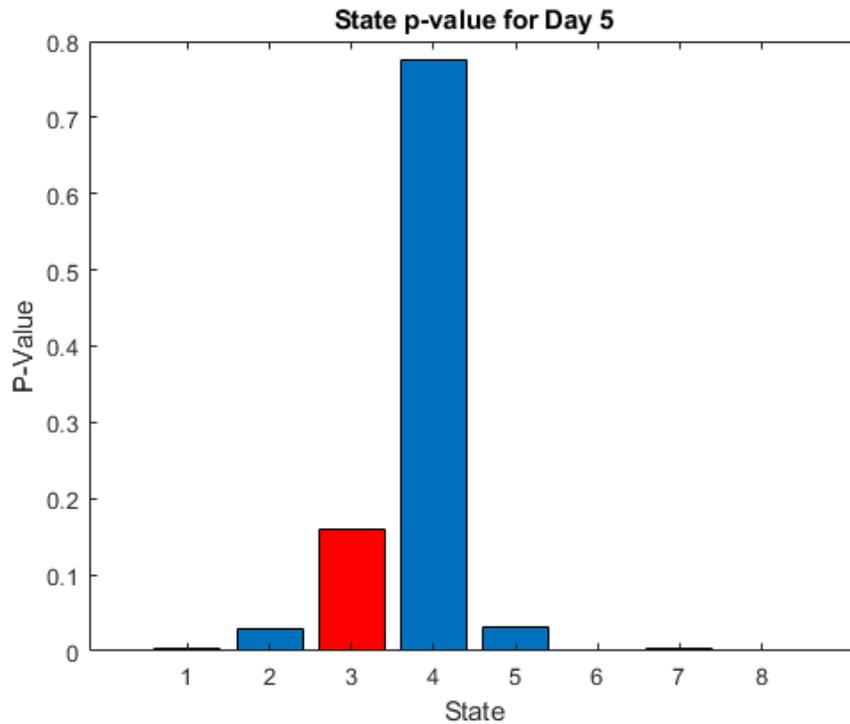


Figure 4.4: Probability Distribution Output of the Viterbi Algorithm

Using this new output format, a threshold test was implemented using the states with the 2 largest p-values (s_1 and s_2 respectively). Let us define the difference between the 2 largest p-values as \bar{p}

$$\bar{p} = p_1 - p_2$$

where p_1 and p_2 represent the p-values for the respective states s_1 and s_2 . Using this, the next state transition S_{i+1} can be chosen as

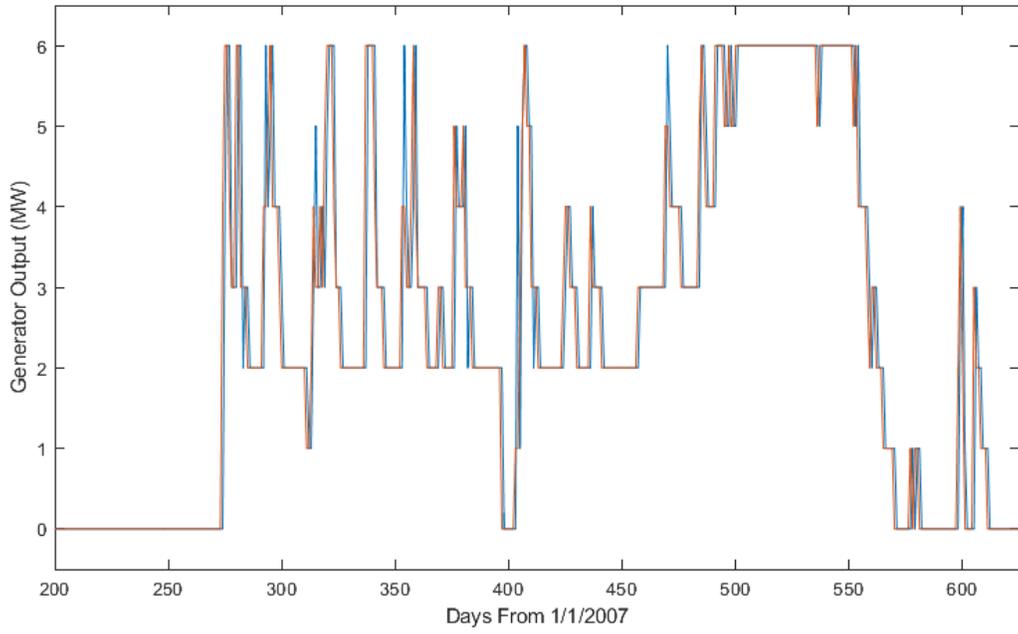
$$S_{i+1} = \begin{cases} s_1 & \text{if } \bar{p} \geq \eta \\ s_2 & \text{if } \bar{p} < \eta \end{cases}$$

where η represents the threshold value, which is calculated by running the aggregate model over h days in the construction period¹. By computing the weighted average of the historical \bar{p} values and comparing the results of running the Viterbi model during this period, η is computed as

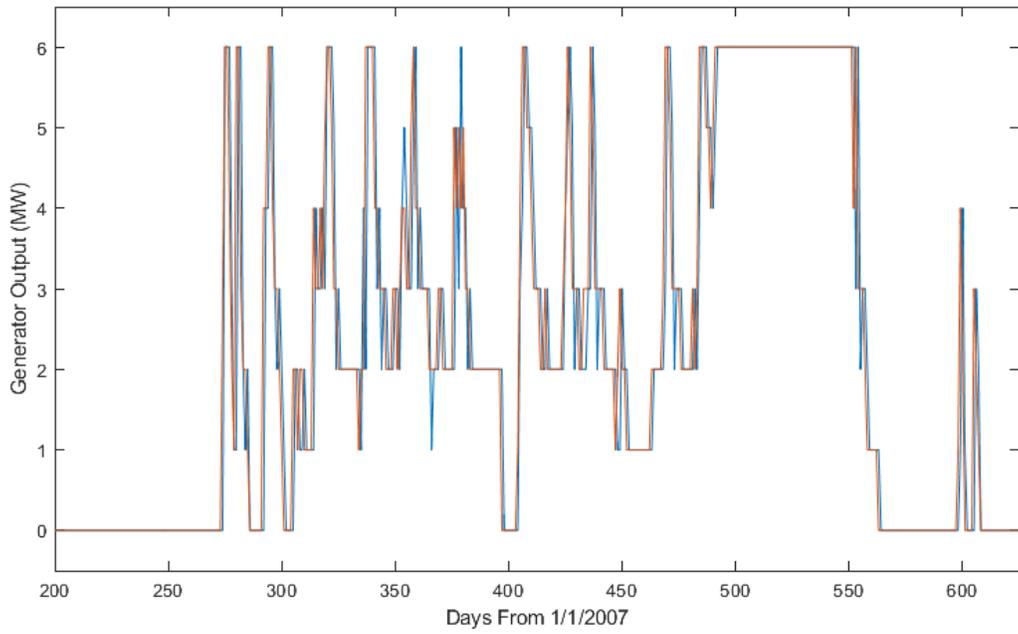
$$\eta = 1 - \frac{1}{h} \sum_{i=1}^h (1 - \bar{p}_i)$$

Based on this modification of the aggregate model, the resultant plots in Figure 4.5 show that the forecast values perform equally well, if not better than the persistence model. The usage of the threshold test dissuades the model from choosing the persistence solution, resulting in more nonzero state transitions.

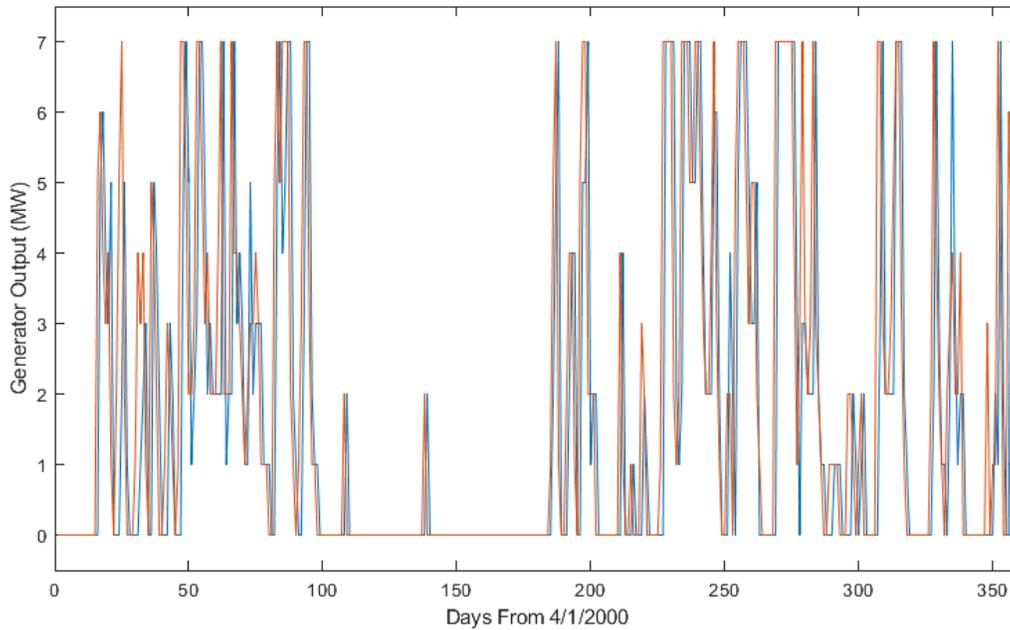
¹The threshold computation period 1/1/2002-12/31/2004.



(a) Calligan Creek



(b) Hancock Creek



(c) Youngs Creek Creek

Figure 4.5: Hierarchical Model Forecast Results (blue) Compared With Actual Historical Values (red)

Theoretical Accuracy Rates

While the threshold test that was implemented for the hierarchical model is limited to the first 2 largest p-values, it brings up the question of the theoretical accuracy rate: if the hierarchical model methodology was able to incorporate up to the n -th largest p-value in the probability distribution output from the Viterbi algorithm and be able to choose the correct state transitions, what is the theoretical accuracy that can be achieved using this model?

Based on Figure 4.6, the improvement in the accuracy starts to diminish as n approaches 3^2 . As well, since these theoretical accuracy rates evaluate the $\pm 0\text{MW}$ tolerance, it

²For the case when $n = 1$, the theoretical maximum is identical to the persistence model as the largest p-value will always correspond to the state where $\Delta P = 0$.

is inferred that the accuracy rate for the $\pm 1\text{MW}$ tolerance would be much higher. Therefore, it is sufficient to use $n = 2$ and achieve a potentially high accuracy rate, provided a more sophisticated threshold test methodology which would reliably return the correct state transitions.

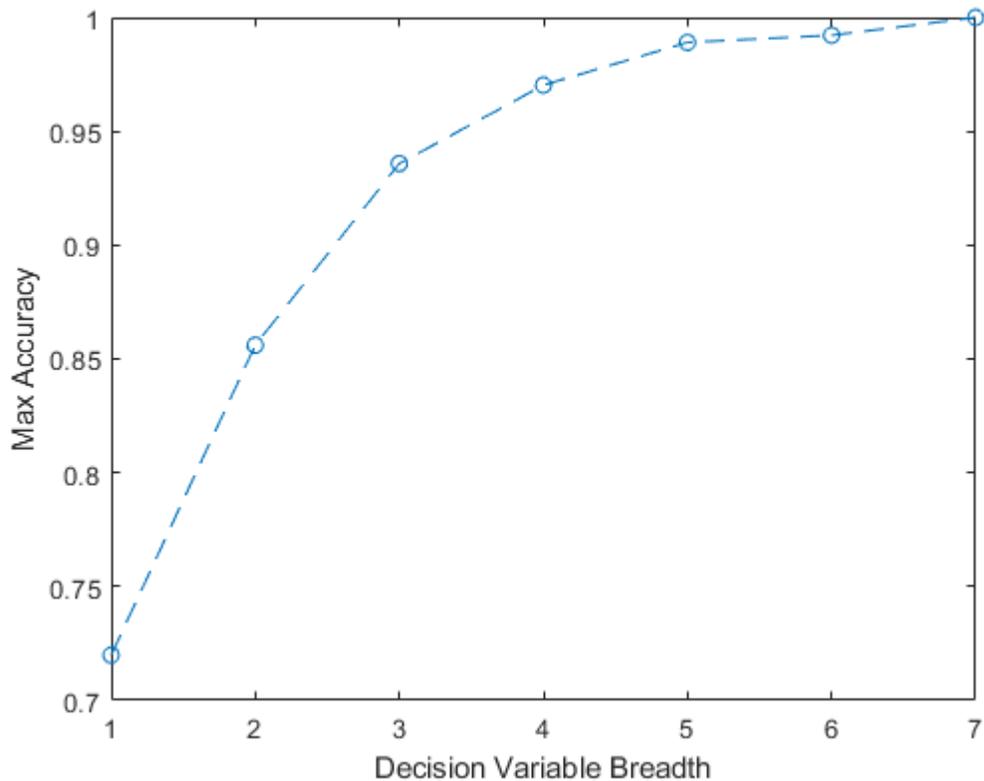


Figure 4.6: Maximum Model Accuracy Based On Decision Variable Breadth, Calligan Creek

4.1.6 Summary of Model Results

As summarized in Table 4.2, utilizing both seasonal and hierarchical models acted as a data filter for the HMM construction, resulting in slight improvements over the benchmark model. By incorporating both seasonality effects and methodologies to minimize the effect of the uneven hidden state distribution, the validation accuracy was able to outperform the

| | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ |
|----------------|---------|---------|---------|---------|
| Calligan Creek | 0.720 | 0.856 | 0.936 | 0.970 |
| Hancock Creek | 0.633 | 0.830 | 0.915 | 0.962 |
| Youngs Creek | 0.518 | 0.649 | 0.696 | 0.795 |

Table 4.1: Theoretical Maximum Accuracy Using Hierarchical HMM with Increasing n

persistence model. Additionally, the improvements in accuracy is dependent on the type of hydrological system the models were used on, as the evident in the greatest increase for Youngs Creek, which was identified as the flashiest system.

| Model Type | Calligan Creek | | Hancock Creek | | Youngs Creek | |
|--------------|----------------|------------|---------------|------------|--------------|------------|
| | ± 0 MW | ± 1 MW | ± 0 MW | ± 1 MW | ± 0 MW | ± 1 MW |
| Persistence | 0.720 | 0.928 | 0.693 | 0.920 | 0.540 | 0.712 |
| Baum-Welch | 0.382 | 0.613 | 0.438 | 0.662 | 0.408 | 0.627 |
| Aggregate | 0.720 | 0.928 | 0.693 | 0.920 | 0.540 | 0.712 |
| Seasonal | 0.721 | 0.930 | 0.695 | 0.922 | 0.521 | 0.701 |
| Hierarchical | 0.730 | 0.930 | 0.673 | 0.923 | 0.553 | 0.712 |

Table 4.2: Cumulative Accuracy Results for Running All 4 Models over the 3 ROR Systems

4.2 Discussion of Results

4.2.1 Financial Advantages

In order to quantify the improvements in the forecast accuracy, it is useful to analyze this in terms of savings in operational cost. While the forecast values are accurate within the tolerance level of ± 1 MW, the cost of balancing this offset will need to be considered from an operational standpoint.

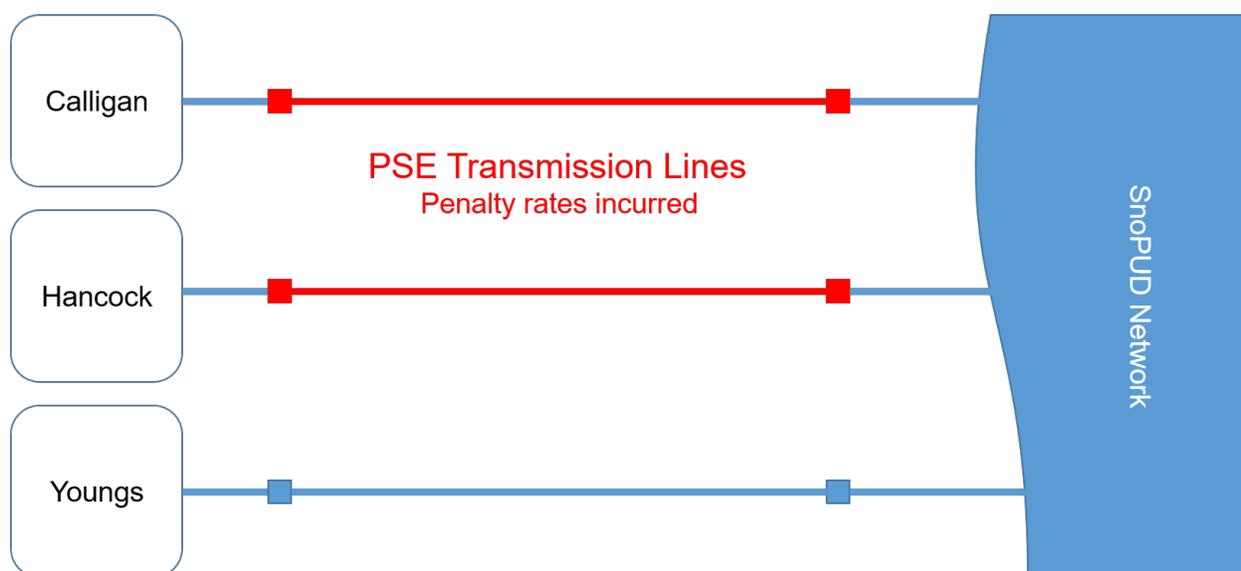


Figure 4.7: Transmission Overview

Youngs Creek

The Youngs Creek plant is connected to the SnoPUD network through a transmission line which is owned and operated by SnoPUD. As a result, the cost of balancing the mismatched generation is dependent on the marginal price within the network. Therefore, the implications of having a more accurate forecast would yield to operational savings, given that the marginal cost of operating the Youngs Creek plant is lower than that of the next cheapest generator that is available.

Calligan and Hancock Creek

Unlike Youngs Creek, Calligan Creek and Hancock Creek plants are connected to the SnoPUD network via transmission lines owned and operated by Puget Sound Energy (PSE). Therefore, if there is more power that is sent across those lines than initially planned for, the cost of balancing the mismatched generation is the total sum of both the marginal cost in the SnoPUD network, and the excess penalty cost incurred by PSE. Because this excess

penalty is set by the real time market price of electricity in the network, this penalty cost can be much greater than the marginal cost.

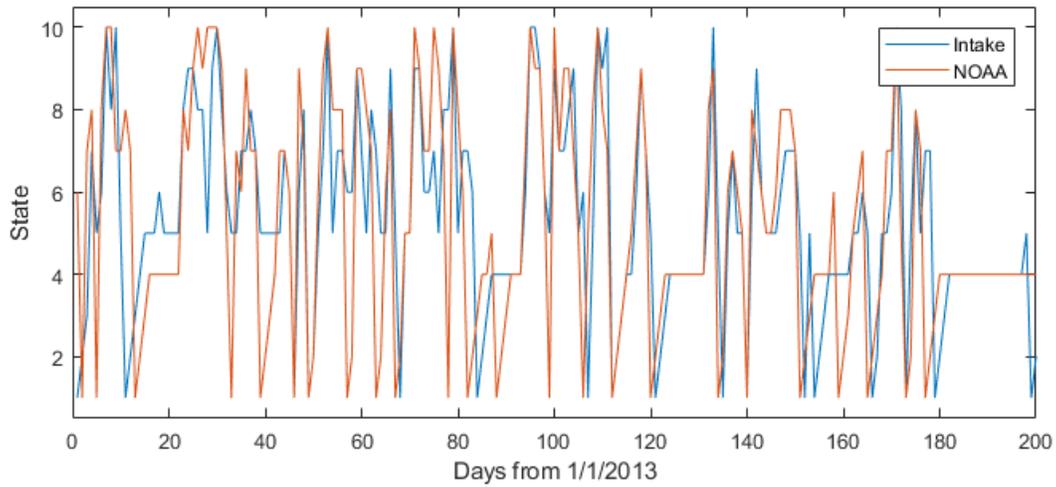
4.2.2 Precipitation Data Error

The main source of error for the data used for the modeling process stems from the utilization of NOAA sites instead of precipitation data recorded directly at the intake sites. Since the NOAA station is located approximately 15 miles north and 1000 feet above the intake sites, there are some deviations in precipitation measurements which must be considered. Figure 4.8 visualizes the observation state error Δo_i between the NOAA data and the Hancock Creek intake data during 2013. The deviations in the states can be generalized by the meteorological differences between the 2 sites.

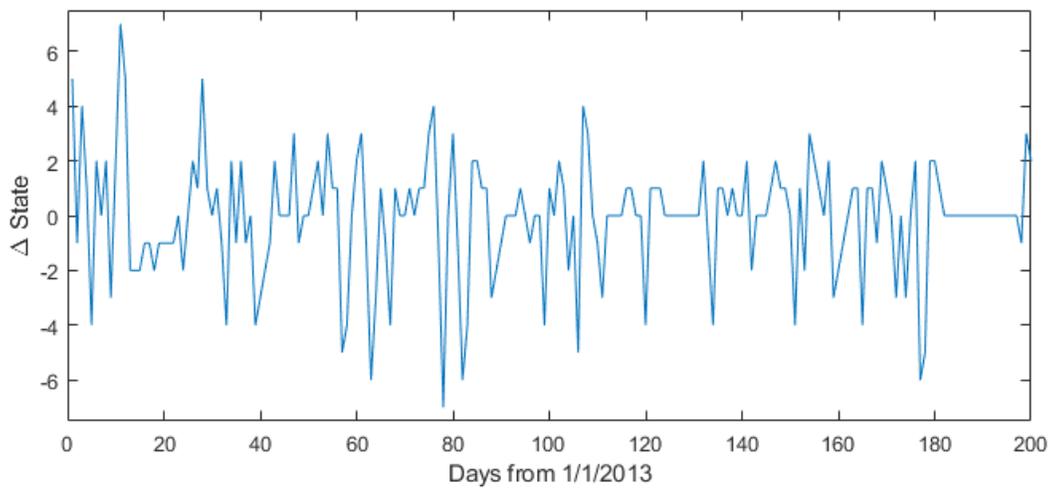
During dry periods in the summer months, deviations in the precipitation data does not cause large differences in the observation states. Conversely, during periods of heavy precipitation, there is some state deviation between the 2 sites as it is likely that both sites are experiencing the same rainfall. However, there are periods when one site experiences a heavier rainfall compared to the other site. However, the majority of the state deviation stem from the utilization of a hard 0mm precipitation level for defining states 1-4. Therefore, it is important to match the accuracy of the weather stations to better record the precipitation data.

The other biggest source of error is the lack of temperature data during the model construction. The difference in altitude between the 2 sites translates to a temperature difference of 3°F, which can translate to a difference between rainfall and snowfall. Since snowfall leads to accumulation rather than contributing to discharge, the lack of categorization in the precipitation data can lead to misleading forecasts.

Therefore, while the use of the NOAA data is valid for a majority of the days in the validation period, there are some inherent limitations due remoteness from the intake sites.



(a) Observation States

(b) Observation State Error $o_{i,NOAA} - o_{i,Intake}$ Figure 4.8: Observation Sequence o_i Error Between NOAA and Intake Sites

Chapter 5

CONCLUSION

The application of statistical learning models for the operation of run-of-the-river hydroelectric generation is a growing field with numerous potential applications based on both previous research work and from the Hidden Markov Model approach proposed in this research project. Through the implementation of HMMs, this research project has demonstrated accurate day-ahead generation forecasts with better accuracy compared to the persistence model approach.

Using various algorithms and filtering methods to create different types of HMMs, it was shown that the hierarchical HMM approach yielded the greatest improvement in the accuracy rate, especially for Youngs Creek. This was primarily due to the management of the uneven hidden state distribution, which created a bias in the Viterbi algorithm. Although this improvement is relatively small ($< 1.5\%$), the ability to achieve this improvement using only one set of observation states shows a promising sign for greater improvements with the addition of additional observations.

In addition to the choice of the construction algorithm, the effectiveness of HMMs is dependent on the type of hydrograph that they are utilized on. Within the scope of the research project, it was demonstrated that the flashier hydrographs such as Youngs Creek saw the greatest improvement in the accuracy rate compared to the steadier systems at Calligan and Hancock Creeks. This was primarily due to the relatively low accuracy rates for the persistence model for Youngs compared to Hancock and Calligan, where the persistence model sets a high benchmark. Despite the lumping of the additional complexities in the hydrograph system such as the effects of groundwater leaching and evapotranspiration, the black-box approach of HMMs is effective in capturing the effect of precipitation and the

variations in the available amount of generation for each system.

5.1 Future Work

This project paves the way for potential future work in the area of run-of-the-river hydroelectricity operation, including:

- Implementation of a wider range of meteorological data to introduce more observational states, including temperature and snow water equivalence
- Determining a more sophisticated threshold test to improve the accuracy of the hierarchical HMM
- Extending the forecast period to test the effectiveness of forecasting over a longer period of time (ie. week-ahead forecasts)
- Constructing a tool to translate remote precipitation data at the NOAA weather station to forecast the precipitation data at the intake sites

BIBLIOGRAPHY

- [1] Snohomish County Public Utility District. Power supply, 2016.
- [2] U.S. Department of Energy. 2017 hydropower market report, April 2018.
- [3] X.Ku, M.Parizeau, and R.Plamondon. Training hidden markov models with multiple observations - a combinatorial method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):371–377, 2000.
- [4] B.Obama. The irreversible momentum of clean energy. *Science*, 2017.
- [5] Snohomish County Public Utility District. Renewable power at the pud, 2017.
- [6] Snohomish County Public Utility District. Quick facts, 2017.
- [7] FM3 Research. Washington voter views of wild salmon and snake river dams, 2018.
- [8] B.R.Deemer, J.A.Harrison, S.Li, J.J.Beaulieu, T.DelSontro, N.Barros, J.F.Bezerra-Neto, S.M.Powers, M.A.dos Santos, and J.A.Vonk. Greenhouse gas emissions from reservoir water surfaces: A new global synthesis. *Bioscience*, 66(11):949–964, 2016.
- [9] J.D.Salas and J.T.B.Obeysekera. Arma model identification of hydrologic time series. *Water Resources Research*, 18(4):1011–1021, 1982.
- [10] J.D.Salas, D.C.Boes, and R.A.Smith. Estimation of arma models with seasonal parameters. *Water Resources Research*, 18(4):1006–1010, 1982.
- [11] İ.Can, F.Tosunoğlu, and E.Kahya. Daily streamflow modelling using autoregressive moving average and artificial neural networks models: Case study of Çoruh basin, turkey. *Water and Environment*, 26(4):567–576, 2012.
- [12] A.Shabri and Suhartono. Streamflow forecasting using least-squares support vector machines. *Hydrological Sciences Journal*, 57(7):1275–1293, 2012.
- [13] J.I.Pérez-Díaz and J.Fraile-Ardanuy. Neural networks for optimal operation of a run-of-river adjustable speed hydro power plant with axial-flow propeller turbine. In *2008 16th Mediterranean Conference On Control and Automation*, pages 309–314, June 2008.

- [14] T.S. Kokkonen and A.J. Jakeman. A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resources Research*, 37(9):2345–2352, 2001.
- [15] N.H. Augustin, L. Beevers, and W.T. Sloan. Predicting river flows for future climates using an autoregressive multinomial logit model. *Water Resources Research*, 44(7):W07403, 2008.
- [16] D. Pender, S. Patidar, G. Pender, and H. Haynes. Stochastic simulation of daily streamflow sequences using a hidden markov model. *Hydrology Research*, 47(1):75–88, 2016.
- [17] State of Washington Department of Ecology. Protecting instream flows, 2018.
- [18] J. Eisner. An interactive spreadsheet for teaching the forward-backward algorithm. In *ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, pages 10–18, July 2002.
- [19] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [20] G.D. Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973.
- [21] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.
- [22] W. Chang. A literature review of wind forecasting methods. *Journal of Power and Energy Engineering*, 2(4):161–168, 2014.
- [23] S.S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium 2010*, pages 1–8, Sept 2010.
- [24] S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, 32(1):41–62, 1998.
- [25] Snohomish County Public Utility District. Federal energy regulatory commission order issuing original license - calligan creek, 2015.
- [26] Snohomish County Public Utility District. Federal energy regulatory commission order issuing original license - hancock creek, 2015.

- [27] Snohomish County Public Utility District. Federal energy regulatory commission order issuing original license - youngs creek, 1992.

Appendix A

CALLIGAN CREEK PROJECT OVERVIEW

In the following pages, a brochure which summarizes the details of the Calligan Creek hydroelectric project is outlined courtesy of Jessica Spahr from SnoPUD. A more comprehensive FERC report can be seen in [25].



QUICK FACTS

Location:

- 9 miles northeast of the city of North Bend, in King County
- Runs between Calligan Lake (2,226 feet msl) and the confluence with the North Fork Snoqualmie River (1,100 feet msl)
- Within the 89,500-acres of timber forest lands owned by Campbell Global Forest Management

Generation:

- Nameplate capacity – 6.0 megawatts
- 20,700 megawatt-hours annually on average
- Run-of-the-river operation
- Typically not operating in July-September

Facilities:

- Impoundment – 1.04-acre-ft, 0.26 surface acres, no active storage
- Weir – 45' x 8' spillway, also includes sluice gate and trash racks
- Fish screens – self-cleaning, approach velocities of 0.4 ft per second or less
- Fishway – pool and weir fish passage, passes minimum instream flows
- Penstock – 1.2-miles, 41-45" diameter pipe, conveying up to 88 cubic feet per second of water
- Powerhouse – 48' x 60' x 41' building with one 6-MW Pelton turbine
- Tailrace – 135' riprap-lined channel, with 2' high fish exclusion barrier
- Switchyard – step-up transformer
- Transmission line – 2.5-mile long, 34.5kV buried line
- Roads – two short access roads – one to intake, one to powerhouse

Construction:

- Estimated cost – \$24 million
- Started September 2015, estimated finish late 2017
- Estimated start of operation after commissioning – early 2018

Licensing:

- Started due diligence investigation in 2009
- Filed Pre-Application Document in September 2011 and Final License Application in August 2013, using FERC's Traditional Licensing Process
- Received License on June 23, 2015, with 50-year term (expires May 2065)

Environment:

- Fish – above anadromous fish barrier (Snoqualmie Falls), no ESA species or Essential Fish Habitat in area; mostly cutthroat, brook, and rainbow trout, stickleback, whitefish, and sculpin
- Animals – no ESA species, typical of forest management areas
- Plants – no rare plants, some noxious weeds and wetland, typical of forest management areas
- Recreation – provided by Campbell Global with permit, hiking, hunting, primitive camping
- Cultural – no National Register eligible properties, general area historically used for timber, mining, and hops farming, and Snoqualmie tribe use

Mitigation:

- Minimum instream flows – 2 cfs downstream of weir (year round), 15 cfs (5/15–9/14) and 6 cfs (9/15–5/14) downstream of spring-fed area
- Downramping rates – 1-2" per hour based on time of year and day/night
- Flow monitoring – stream gage monitoring of instream flows and downramping rates in two locations
- Trout monitoring and adaptive management – snorkel survey up to 5 yrs to compare pre- and post-operation flow adequacy, increase minimum instream flows if catastrophic decline in trout population reported
- Flushing sediments – flush once/yr when flows above 80 cfs, for 6 hrs min.
- Water quality monitoring – monitoring water temperature (6 locations) and turbidity (1 location)
- Project placement – relocated penstock to avoid wetlands, buried penstock to allow animal crossing and t-line to avoid bird collisions
- Land mitigation and management – 10.53 acres for preservation, replanting/seeding penstock corridor
- Weed management – control weeds on lands within project boundary
- Recreation – allow access to penstock corridor for those with Hancock Forest Management permit
- Cultural – implementation of the Unanticipated Discovery Plan if needed, allow tribal access to penstock corridor
- Aesthetics – use natural colors, directional and time-phased lighting, and vegetative screening at facilities

Appendix B

HANCOCK CREEK PROJECT OVERVIEW

In the following pages, a brochure which summarizes the details of the Hancock Creek hydroelectric project is outlined courtesy of Jessica Spahr from SnoPUD. A more comprehensive FERC report can be seen in [26].



QUICK FACTS

Location:

- 7 miles northeast of the city of North Bend, in King County
- Runs between Lake Hancock (2,172 feet msl) and the confluence with the North Fork Snoqualmie River (1,043 feet msl)
- Within the 89,500-acres of timber forest lands owned by Campbell Global Forest Management

Generation:

- Nameplate capacity – 6.0 megawatts
- 22,100 megawatt-hours annually on average
- Run-of-the-river operation
- Typically not operating in July-September

Facilities:

- Impoundment – 0.65-acre-ft, 0.18 surface acres, no active storage
- Weir – 46' x 6' spillway, also includes sluice gate and trash racks
- Fish screens – self-cleaning, approach velocities of 0.4 ft per second or less
- Fishway – pool and weir, passes minimum instream flows
- Penstock – 1.5-miles, 39-44" diameter pipe, conveying up to 81 cubic feet per second of water
- Powerhouse – 48' x 60' x 40' building with one 6-MW Pelton turbine
- Tailrace – 100' riprap-lined channel, with 2' high fish exclusion barrier
- Switchyard – step-up transformer
- Transmission line – 0.3-mile long, 34.5kV buried line
- Roads – two short access roads – one to intake, one to powerhouse

Construction:

- Estimated cost – \$28 million
- Started April 2016, completed early 2018
- Start of operation February 16, 2018

Licensing:

- Started due diligence investigation in 2009
- Filed Pre-Application Document in September 2011 and Final License Application in August 2013, using FERC's Traditional Licensing Process
- Received License on June 19, 2015, with 50-year term (expires May 2065)

Environment:

- Fish – above anadromous fish barrier (Snoqualmie Falls), no Endangered Species Act (ESA) listed species or Essential Fish Habitat in area; mostly rainbow, cutthroat, and brook
- Animals – no ESA species, typical of forest management areas
- Plants – no rare plants (except water lobelia is a State threatened plant), some noxious weeds and wetlands, typical of forest management areas
- Recreation – provided by Campbell Global with permit, hiking, hunting, primitive camping
- Cultural – no National Register eligible properties, general area historically used for timber, mining, and hops farming, and Snoqualmie tribe use

Mitigation:

- Minimum instream flows – 5 cfs (10/16-6/15) and 20 cfs (6/16-10/15) downstream of weir
- Downramping rates – 1-2" per hour based on time of year and day/night
- Flow monitoring – stream gage monitoring of instream flows and downramping rates in two locations
- Trout monitoring and adaptive management – snorkel survey up to 5 yrs to compare pre- and post-operation flow adequacy, increase minimum instream flows if catastrophic decline in trout population reported
- Flushing sediments – flush once/yr when flows above 100 cfs, for 6 hrs min.
- Water quality monitoring – monitoring water temperature (4 locations) and turbidity (1 location)
- Project placement – relocated penstock to avoid wetlands, buried penstock to allow animal crossing and t-line to avoid bird collisions
- Land mitigation and management – 4.08 acres for preservation, replanting/seeding penstock corridor
- Weed management – control weeds on lands within project boundary
- Recreation – allow access to penstock corridor for those with Hancock Forest Management permit
- Cultural – implementation of the Unanticipated Discovery Plan if needed, allow tribal access to penstock corridor
- Aesthetics – use natural colors, directional and time-phased lighting, and vegetative screening at facilities

Appendix C

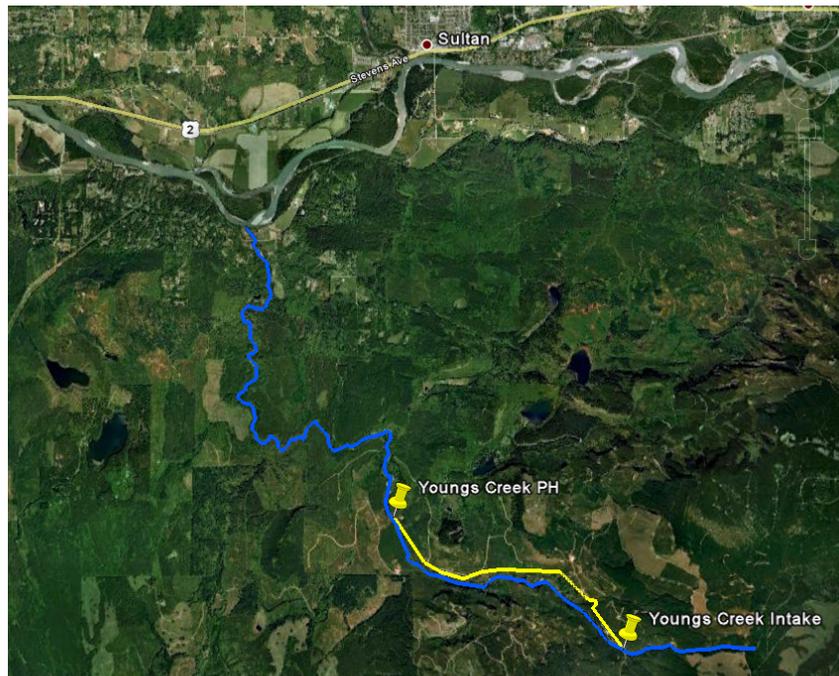
YOUNGS CREEK PROJECT OVERVIEW

In the following pages, a brochure which summarizes the details of the Youngs Creek hydroelectric project is outlined courtesy of Jessica Spahr from SnoPUD. A more comprehensive FERC report can be seen in [27].

YOUNGS CREEK HYDRO

renewable ♪ reliable ♪ environmentally responsible
S N O H O M I S H C O U N T Y P U D

The Youngs Creek Hydroelectric Project is a new run-of-the-river hydroelectric facility with an installed capacity of 7.5 megawatts located approximately 4.5 miles south of the city of Sultan in Snohomish County, Washington. The Project was licensed with the Federal Energy Regulatory Commission in May 1992; however, due to a lack of progress on construction the FERC was going to terminate the license in the late 2000s. The District acquired the Project assets in 2008, and immediately began engineering and permitting activities for the construction of the Project. The Project is the first run-of-the-river project (not using existing facilities) in Washington State to be constructed in over 20 years; and the first to be built in Snohomish County in the last 30 years.



Environment

The Project is located on Youngs Creek, a tributary to Elwell Creek which is a tributary to the Skykomish River. Project features begin 1.4 miles upstream of a permanent adult migration barrier to anadromous fish at river mile (RM) 1.0 on Youngs Creek. The Project



is situated on approximately 25.3 acres of District-owned, previously-classified timberland on the west slope of the Cascade Mountain Range in the Skykomish River Basin. Nearly the entire drainage has been logged at least once within the last 75 years, leaving timber stands of various ages and plant species associations. No ESA-listed species are in the area.

Construction

The District initiated construction of the Project in February 2010 and expects to be completed with construction in October 2011. Testing of the turbines commenced in summer 2011. Formal commissioning and operation of the Project is to occur in October 2011. The Project will have cost approximately \$28,000,000 to purchase, engineer, license, permit and construct.

Facilities

Intake

A diversion weir and intake structure is at RM 5.0 on Youngs Creek (elevation 1,530 feet above mean sea level). The weir is 12.0 feet high and 65 feet long. The total pool behind the weir is approximately 9,150 square feet (0.21 acres). The water intake structure consists of a concrete structure with a self-cleaning trash rack, fish screens, and closure gate built to Washington Department of Fish and Wildlife fish protection standards.

Penstock

The buried penstock runs along the right bank along with the intake and powerhouse structures, so no crossing of Youngs Creek occurs. A 14,300-foot long, steel penstock with an initial diameter of 51 inches transitioning to 48 inches, is routed down existing roads that have been cleared of trees since 1994.

Powerhouse

The powerhouse is located at RM 2.4, upstream of a complete natural barrier to anadromous fish migration located at RM 1.0. The powerhouse is an approximately 46-foot by 48-foot wide concrete structure, set-back from the ordinary high water mark of the river approximately 40-feet. Discharge flows would be returned to the creek in a 12 -foot wide by riprap lined channel. An outdoor switchyard is located next to the powerhouse to contain the main power transformer and other electrical equipment. The powerhouse contains a Pelton turbine/generator with an installed capacity of 7.5 MW. That is enough to power over 5600 homes annually based on nameplate capacity.

Transmission Line



The new transmission line follows the access and other existing roads for 7.8 miles to the Snohomish County PUD substation at Sultan. Portions of the transmission line are buried along the existing roads; others are overhead on existing poles or new poles that would contain an existing distribution line and the new transmission line.

Mitigation, Protection and Enhancement Measures (PM&Es)

Multiple management plans are in place to monitor and mitigate the fishery, terrestrial (avian, vegetation), and water resources of the Project area.

Ribbon Cutting

A ribbon cutting will take place at the Project powerhouse around October 2011 to commemorate the commissioning of the Project. (The exact date is currently being selected in consultation the District's General Manager's and Commissioners' schedules). This is the first new constructed run-of-the-river project in many years for Washington State and the first generation facility built and owned by the District since 1984 when the Jackson Hydro Project was complete. Lunch and program will occur from 11:30-1:00 with tours of the powerhouse and intake sites from 1:00-2:00. Further details will follow once the date is selected for this event.

Contact

For further information on the Ribbon Cutting Event or the Youngs Creek Project, contact Dawn Presler in the Generation Resources Department at 425-783-1709 or DJPresler@snopud.com.