DISSERTATION

A VALUE-FUNCTION BASED METHOD FOR INCORPORATING ENSEMBLE FORECASTS IN REAL-TIME OPTIMAL RESERVOIR OPERATIONS

Submitted by

Matthew E. Peacock

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: John W. Labadie

Jorge Ramirez
Chuck Anderson
Lynn Johnson

ABSTRACT

A VALUE-FUNCTION BASED METHOD FOR INCORPORATING ENSEMBLE FORECASTS IN REAL-TIME OPTIMAL RESERVOIR OPERATIONS

Increasing stress on water resource systems has led to a desire to seek methods of improving the performance of reservoir operations. Water managers face many challenges including changes in demand, variable hydrological input and new environmental pressures. These issues have led to an interest in using ensemble streamflow forecasts to improve the performance of a reservoir system. The currently available methods for using ensemble forecasts encounter difficulties as the resolution of the analysis increases in order to accurately model a real-world system. One of the difficulties is due to the "curse of dimensionality" as computing time exponentially increases when the discretization of the state and action spaces becomes finer or when more state or action variables are considered. Another difficulty is the problem of delayed rewards. When the time step of the analysis becomes shorter than the travel time due to routing, rewards may not be realized in the same time step as the action which caused them. Current methods such as dynamic programming or scenario-tree based methods are not able to handle delayed rewards. This research presents a value function-based method which separates the problem into two subproblems: computing the state-value function in the no-forecast condition, and finding optimal sequences of decisions given the ensemble forecast with the state-value function providing information about the value at any state at the end of the forecast horizon. A continuous action deep reinforcement learning algorithm is used to overcome the problems of dimensionality and delayed rewards, and a particle swarm method is used to find optimal decisions during the forecast horizon. The method is applied to a case study in the Russian River basin and compared to an idealized operating rule. The results show that the reinforcement learning process is able to generate policies that skillfully operate the reservoir without forecasts. When forecasts are used, the method is able to produce non-dominated performance measures. When the water stress to the system is increased by removing a transbasin diversion, the method outperforms the idealized operations.

ACKNOWLEDGEMENTS

I would like to extend many thanks to my advisor, Dr. John Labadie, for his guidance throughout this project and for inspiring my interest in the reservoir operations problem. I am very fortunate that he was willing to accept me as an advisee and have benefited in many ways from his support.

The members of my dissertation committee have been most gracious with their time and expertise. Thank you to Dr. Chuck Anderson for providing me with a foundation to explore machine learning, to Dr. Lynn Johnson for sharing his in-depth knowledge of the Russian River and the ongoing research in the basin, and a great big thank you to Dr. Jorge Ramirez for somehow making me believe that I could do things I didn't know I could.

My fellow graduate students of Lab C213 helped me in so many ways by providing motivation, inspiration and comedic relief. It was a great opportunity for me to share a few years with such talented people. Best of luck to everyone in the future.

This work was partially supported by a grant from the Sonoma County Water Agency and the National Weather Service. I am very appreciative for the opportunity to study reservoir operations in the Russian River basin and hope that this research will help advance the mission of the FIRO program.

And finally, thank you to my parents, sisters, brothers and friends in the mountains. Without them, this would not have been possible.

DEDICATION

To my parents, for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

Reservoir operation and management is currently subject to a rapidly changing environment of shifting objectives and constraints. Factors such as non-stationary hydrological inputs, aging infrastructure, modified regulations and new objectives are requiring that reservoir operations seek strategies to meet conflicting goals. Utilizing forecast information to improve operational decision-making is an area that is showing promise to improve performance. Optimization methods have been successfully applied to enhancing reservoir performance in the past (Labadie, 2004) and research has been performed to incorporate forecasts into the optimization analysis. One specific area of research is in the use of probabilistic forecasts in the form of ensembles of possible scenarios. This has involved primarily dynamic programming (DP) (Dias, Pereira, & Kelman, 1985; Tejada-Guibert, Johnson, & Stedinger, 1993; Faber & Stedinger, 2001; Kim, Eum, Lee, & Ko, 2007; Eum, Kim, & Palmer, 2011; Anvari, Mousavi, & Morid, 2014) and scenario-tree methods (Boucher, Tremblay, Delorme, Perreault, & Anctil, 2012; Raso, Giesen, Stive, Schwanenberg, & Overloop, 2013; Schwanenberg et al., 2014; Schwanenberg et al., 2015; Etkin et al., 2015) in water resources, and (Dupaová, Gröwe-Kuska, & Römisch, 2003; Gröwe-Kuska, Heitsch, & Romisch, 2003) in power management, see also (Côté & Leconte, 2016) which provides a comparison of the two approaches. Recently, (Nayak, Herman, & Steinschneider, 2018) used a policy optimization method to develop decision-tree based rules which can be used with ensemble forecasts.

1.2 Statement of the Problem

Seeking increased performance from reservoir operations brings with it the need to study the problem at increasingly finer resolution. This compounds the already daunting curse of dimensionality for both DP and scenario-tree methods as more and more discrete possible states and actions are allowed. This problem is present as well when increasing the resolution of the probability mass functions that describe the random variables as the number of possible transitions, and thus the number of conditional probabilities which must be computed, will exponentially increase. With the

generally limited amount of data that is available in water resources problems, these probabilities can be hard to define accurately. This problem continues when the scope is expanded to include multiple reservoirs. Beyond the problem of tractability is the limit of current solution methods when the discrete time step becomes shorter than the travel time due to routing. When this occurs, the problem of *delayed rewards* is present as a cost or benefit is not realized in the same time step as the action which produced it. Backwards solution methods such as DP cannot take delayed rewards into account.

Additionally, decision-making in any context likely involves consideration of multiple perspectives and preferences which are processed in some fashion to select a single action. This is particularly a challenge in water resource problems as the set of preferred actions from the stakeholders often contains both "store" and "release" which is unfortunately not possible. Many currently available methods require that multi-objective problems be reduced to a single objective problem through the use of a weighted combination of value or utility functions. This may confuse the analysis as the selection of a value function for a single objective is already an abstraction of a mental concept of preference into a mathematical function. Combining two or more of these functions through weighting adds another layer of abstraction to the process and risks losing meaningfulness.

Furthermore, reservoir operations in general would appear to rely on information only available at the time a release decision is made. Rules and schedules are made but the decision is left to an operator who exercises judgment. No level of precision in the computational analysis provides enough certainty that the system can become autonomous. And it would seem that the reliance on current information is necessarily true when using forecasts, as in practice, forecasts are made in an ongoing manner and updated at some interval. There must be some new information making its way into the process, otherwise, any forecast could have been made the day before. The output of many current methods is a policy for decision selection which does not provide information on the consequence of deviating from that policy.

1.3 Research Questions

The main questions addressed in this research are:

- How can we use knowledge of the state-value function of a system to make decisions given an ensemble forecast?
- How can we use reinforcement learning to overcome the obstacles of tractability and delayed rewards that are often present when we try to calculate the state-value function?
- Can we compute this state-value function for objectives individually? And if so, how can we then use multiple state-value functions to make operational decisions given an ensemble forecast?

1.4 Purpose of the Study

The purpose of this study is to address the research questions by separating the problem of operational decision-making using forecasts into two subproblems: decision-making in the absence of a forecast, and decision-making within the forecast horizon. The purpose is also to seek realism in both the decision framework and the numerical model. Realism of the decision framework in this sense refers to limiting the complexity of the abstraction from the "value" that a decision-maker associates with an outcome, and the mathematical model of that value that is used in the optimization. The analysis seeks to avoid, to the extent possible, assuming knowledge about the value of an outcome beyond the minimum. Realism, in reference to the model, is providing analysis at the resolution which is appropriate for the operational objectives and constraints of the problem. At the minimum, the resolution must be able to consider the effect of delayed rewards. Additionally, the research is intended to produce general methods which can be applied to a range of reservoir operation problems.

1.5 Need for the Study

The World Commission on Dams (WCD) was created in 1997 with the task of performing a global evaluation of the effectiveness of large dams and to provide recommendations for future decision-making regarding development and management of large storage projects. In its report published in 2000 (World Commission on Dams, 2001), the WCD found that many large dams did not meet the physical and economic targets which motivated their development while having overall more negative than positive impacts on ecosystems. It found as well that conflicts over dams

had increased in the preceding 20 years due to the historical lack of consideration of the social and environmental impacts that accompany these large development projects. One of the strategic priorities laid out in the commission's recommendations is to address the management of existing dams and one of the policy principles to enact as part of this priority is to optimize the operation of these existing large reservoirs.

The case study presented in this research is an example of a collection of stakeholders which is looking to new methods of reservoir operation to meet its current objectives, some of which were not present at the time the dam was constructed. Coyote Valley Dam which created Lake Mendocino was originally constructed for flood control. Agriculture, primarily the growing of grapes for wine production, has become a major economic driver in the basin. At the same time, there are currently three species of fish listed as endangered or threatened in the Russian River. The Sonoma County Water Agency (SCWA) manages the majority of the conservation storage in Lake Mendocino. Facing the risk of shortages, the SCWA has been required to file a Temporary Urgency Change Petition with the California State Water Board (SWB) five times since 2006 to get reprieve from the environmental flow requirements of Decision 1610 from the SWB and the Biological Opinion of the National Marine Fisheries Service (Sonoma County Water Agency, 2015).

The region experiences wet seasons characterized by infrequent short-duration, high-intensity storm events during the winter and extended periods during the summer in which little if any precipitation occurs. The lack of summer precipitation or the benefit of snowmelt runoff make storage decisions in the wet season even more critical. Absent a method for anticipating future precipitation, operators must continue to maintain storage capacity for possible flood events through March, facing the possibility that no precipitation will occur later to provide supply.

The use of forecasts is one possible method for increasing the performance of the reservoir operations. This reservoir system presents multiple complexities which limit the applicability of current methods. The research presented in this study looks to overcome these challenges by developing a method for including ensemble forecasts in the decision-making process which can be applied to a larger set of reservoir operation problems.

1.6 Contribution of This Research

This research makes the following contributions:

1. A novel approach to using ensemble forecasts in reservoir operations which is based on the state-value functions of the system and which considers objectives individually throughout the process.

2. The first application of reinforcement learning which is continuous in both the state and action dimensions to the reservoir operation problem.

1.7 Organization of This Dissertation

This dissertation begins with the introductory chapter in which the problem is described and the contributions are presented. Chapter 2 contains a review of the literature related to reservoir operations which use ensemble forecasts. Chapter 3 presents the methods used in this research including background information in reinforcement learning and particle swarm optimization as well as the process for generating synthetic data for use in the reinforcement learning process. Chapter 4 introduces the Russian River basin which is used as a case study for application of the methods developed in this research. Chapter 5 presents the results of this research and begins with a quantitative analysis of the forecast performance. Next, the results of the continuous action deep reinforcement learning method used to calculate state-value functions for multiple objectives are shown followed by the results from the simulation using the ensemble forecasts. Chapter 6 contains the conclusions of this research and recommendations for future research.

Chapter 2

Review of the Literature

2.1 Introduction

A review of the literature currently available relating to this research was conducted. The review focused on applications of optimization-based methods to reservoir operations which incorporate ensemble forecasts into the decision-making process. A review of the research from the field of computer science which produced the reinforcement learning algorithm used in this research is also included.

2.2 Optimization With Forecasts

Application of optimization methods which include ensemble forecasts can generally be categorized as dynamic programming (DP) or scenario-tree based methods, and both are discussed in the following sections.

2.2.1 Dynamic Programming Applications with Forecasts

The method of sampling dynamic programming (SADP), which forms the backbone of methods that will later incorporate forecasts, dates to at least 1974. A reference to a study by (Araujo & Terry, 1974) is made in (Dias, Pereira, & Kelman, 1985) as research that had earlier used the SADP method. The paper by (Araujo & Terry, 1974) is in Portuguese and its citation is available, though the paper itself was not to be found on the internet. The study by (Dias, Pereira, & Kelman, 1985) uses SADP to generate operating rules for three reservoirs in Brazil which are operated for flood control and hydropower production. The SADP method in this case was selected as the optimization method over traditional stochastic dynamic programming (SDP) because the current study wanted to examine daily operating policies. Traditional SDP methods typically include a previous inflow as a state variable. The authors point out that this assumes that the inflow process is autoregressive. When examining daily operation, considering the previous time step's inflow may not be sufficient to describe the autoregressive behavior of the inflow process. Including additional state variables for previous time steps would lead to the curse of dimensionality and the problem of

calculating conditional probabilities. The SADP method avoids these difficulties by considering an ensemble of streamflows as possible inflow scenarios and we are left with a set of deterministic DP problems which are easier to solve. The scenarios are sequences of historical inflow values for the same calendar period, in this case the flood season. Each scenario is considered equally likely, and the optimization is performed for each scenario before the mean reward is calculated.

Later research by (Kelman, Stedinger, Cooper, Hsu, & Yuan, 1990) used historical streamflow scenarios with probabilities conditioned on a streamflow forecast for the following stage, a method given the name sampling stochastic dynamic programming (SSDP). In this formulation, a deterministic streamflow forecast is considered as a state variable. The process requires conditional probabilities for the scenarios based on the forecast in the current stage as well as the conditional probability of the next period's forecast based on the current period's flow and forecast. These are developed using Bayes theorem and regressions of historical data. The objective at each stage maximizes the expected value over the set of scenarios of the benefit from the release plus the expected value of the future benefit conditioned on the stage, forecast and scenario. In this way, the process is only performing the expectation over the scenarios after it has calculated the objective function as if the scenario is maintained into the future. The optimization process was also divided into two stages, with the optimal decision as the one optimizing the objective function with consideration of the transition probabilities between scenarios. The future value function, which is conditioned on the scenario, is updated based on each scenario occurring. This study was able to nearly match a deterministic optimization solution. Though the authors did comment that the additional value provided by the deterministic forecast depended on the objective function and the system configuration.

The study by (Kelman, Stedinger, Cooper, Hsu, & Yuan, 1990) also commented on the advantages present in using a sampling type method over traditional SDP. The authors noted that the scenarios contain information on the distribution of the flows for the current period and also contain information on the season-to-season persistence of streamflows. Also, the error that can be introduced when discretizing the probability distributions is avoided.

(Faber & Stedinger, 2001) developed an updated version of SSDP by replacing the historical streamflow scenarios with forecasted scenarios. The benefit of using forecasted scenarios comes from the forecasts being conditioned on the current state of the system. The ensembles, which contain sequences of streamflows, also maintain spatial and temporal correlation. Historical data

was used to develop the conditional probabilities with the water content of the snowpack used in place of a current period forecast. In a case study of the Williams Fork Reservoir in Colorado, this method was applied with the goals of meeting minimum storage targets and maximizing hydropower production. A single season time period was examined from February to September as this is a snowmelt driven system. An end of season target storage is defined with penalties being assigned to deviations below this target. The SSDP with historical traces and SSDP with ESP performed similarly with both outperforming the actual operations of the reservoir. When the management scenario was adjusted to represent a more realistic situation, the SSDP with ESP outperformed the SSDP with historical traces leading to the conclusion that the method is "better able to assess the trade-off between immediate benefits and the expected value of water in the future."

(Kim, Eum, Lee, & Ko, 2007) applied SSDP to the operation of a multi-reservoir system in the Geum River basin in South Korea during the drawdown period preceding the monsoon season with the goals of maintaining in-stream flows, meeting target storages, and maximizing hydropower production. Forecasts extending to the final stage of the operation period allowed the study by (Faber & Stedinger, 2001) to solve the SSDP through a backward calculation. The forecasts for the Geum River basin extended only through a single one-month time-step. Historical scenarios were used to solve the backward SSDP problem, with the ensemble forecasts used for the single-stage re-optimization. The use of the ensemble forecasts improved performance over exclusively using historical scenarios. The authors noted that the accuracy of the forecast will have a large effect on the performance of the method.

The collection of methods discussed above was compared in a study by (Anvari, Mousavi, & Morid, 2014). A case study of the Zayandeh-Rud reservoir system in Iran with the objectives of satisfying hydropower, municipal, agricultural and environmental demands was used for this analysis. The authors concluded that the SSDP method with forecasts generated using an artificial neural network outperformed standard SDP and SSDP with historical scenarios. Though in this study, the benefit of using the reoptimization technique of (Tejada-Guibert, Johnson, & Stedinger, 1993) was not confirmed.

2.2.2 Methods Using Scenario Trees

Research focused on the use of scenario trees has been primarily concerned with the methods for generating the scenario trees from ensemble forecasts which are then used in multi-stage stochastic programming (MSP) (Schwanenberg et al., 2015; Raso, Giesen, Stive, Schwanenberg, & Overloop, 2013) or model predictive control (Ficchì et al., 2016) optimization methods. (Gröwe-Kuska, Heitsch, & Romisch, 2003) proposed the use of a successive reduction method which bundles scenarios and assigns the accumulated probability of the eliminated scenarios to those remaining. (Raso, Giesen, Stive, Schwanenberg, & Overloop, 2013) proposed Information Flow Modelling (IFM) as a method of generating scenario trees using an assumed uniform prior probability distribution over the scenarios followed by a Bayesian approach which determines the points at which scenarios diverge sufficiently to be considered parts of different branches in the tree. (Schwanenberg et al., 2015) developed a method which uses precipitation dependent distance metrics for determining the branching points, thereby allowing consideration of the lag time between precipitation and streamflow. (Fan et al., 2016) modified this method to maintain a constant flow weighted probability of the original scenarios. Other methods exist and many are discussed in (Dupaová, Consigli, & Wallace, 2000) and (Gröwe-Kuska, Heitsch, & Romisch, 2003).

(Raso, Giesen, Stive, Schwanenberg, & Overloop, 2013) points out that the major limitation of MSP is intractability. This is due to MSP's own curse of dimensionality, in this case arising from the number of scenarios considered and the number of stages over which the optimization takes place. Thus the scenario tree method requires the use of scenario reduction techniques to reduce the number of scenarios in an ensemble while limiting the loss of information. Such techniques have been proposed by (Dupaová, Gröwe-Kuska, & Römisch, 2003; Gröwe-Kuska, Heitsch, & Romisch, 2003; Fan et al., 2016).

The MSP method of solution requires a cost function representing the value of the final state within the objective function. In the literature, the development of this function has not been the focus of detailed discussion. This generally seems to take the form of penalties on deviation from an end of period target storage (Boucher, Tremblay, Delorme, Perreault, & Anctil, 2012; Etkin et al., 2015; Ficchì et al., 2016) or a time-varying target storage or guide curve (Schwanenberg et al., 2014; Schwanenberg et al., 2015; Fan et al., 2016). (Côté & Leconte, 2016) used a state-value function, referred to as a "water-value function" derived from stochastic dynamic programming. It

is suggested in (Fan et al., 2016) that this is developed with regard to a deviation from a guide curve or bounds on the storage.

(Côté & Leconte, 2016) performed a comparison of four methods for optimizing reservoir operations. Current operation of the Rio Tinto Alcan system in Québec, Canada involves solution of a deterministic optimization of each member of an ensemble forecast, the resulting decision being based on the mean of the solutions. A case study compares current operations with SDP, SSDP and a scenario-tree based method. Results showed that the methods that incorporated scenarios outperformed SDP and deterministic methods in terms of increased energy production. The deterministic method consistently maintained higher reservoir levels resulting in more spills. The effect of underdispersion was also evaluated for these methods. The performance was reduced when underdispersion was artificially introduced into the ensembles for all methods, however, the relative performance between methods remained the same. The authors propose that the methods that use the scenarios are able to capture the spatial and temporal correlation of the streamflows while SDP is not.

2.3 Machine Learning

Reinforcement learning (RL) is a category of machine learning in which an agent is able to select actions, in pursuit of a goal, which affect a real or simulated environment and during this process learn the benefit of these actions based on a reward signal observed (Sutton & Barto, 2018). An early application of RL in water resources was performed by (Bouchart & Chkam, 1998) which used reinforcement learning to select pumping rates at three stations which managed levels in the five reservoirs making up the Burncrooks reservoir system in Scotland. The reinforcement learning process trained a set of neurons connected in a cascading manner to select from a set of four possible pumping rates. The reward was a binary signal representing success or failure to meet any of four criteria related to water supply demand or reservoir maximum or minimum storage levels. A controller was created for each of the three pumping stations. This study was able to show the ability to learn optimal decisions given only limited information on the state of the entire system as each controller was provided with only the information on the storage in adjacent reservoirs. The study concluded that the agents were able to learn the proper actions to take to avoid failure, though the quantitative data relating to the results is lacking.

Most of the methods that have been applied recently to reservoir operations are categorized as temporal-difference (TD) methods. These methods are related to both Monte Carlo methods and DP methods in that they do not require a model of the environment and are able to develop estimates using bootstrapping (Sutton & Barto, 2018). The TD(λ) method and the Q-learning method are two TD methods that have been proven to converge to the optimum under certain conditions (Tsitsiklis & Roy, 1997; Watkins & Dayan, 1992).

Temporal difference learning methods focus on solving problems which are structured similar to Markov Decision Processes (MDP) though without an explicit model of the environment. The Q-learning method (Watkins, 1989) seeks to estimate the Q-function, $Q(s, a)$, which returns the expected long-term rewards for a given state, s, and action, a. Operating policies can be developed by selecting the action that provides the maximum expected value for a given state. With Q-learning, the underlying stochastic structure is learned through the use of long time series of repeated epochs of measured streamflow, which results in optimal coordinated operating policies being learned by the algorithm. The forward-looking recursion process of Q-learning also helps to alleviate the curse of dimensionality since it performs a depth-first search process, rather than the breadth-first process traditionally used in dynamic programming and Markov decision processes.

(Lee & Labadie, 2007) presented reinforcement learning as a method for solving the challenging problem of stochastic optimization of multi-reservoir systems with forecasting. Seasonal forecasting is included with this Q-learning model, which is applied to the Geum River basin in South Korea, with consideration of multiple objectives of hydropower generation, water supply, flood control, and in-stream flow requirements. The results showed that Q-learning was able to produce policies which performed better than those developed through implicit stochastic dynamic programming and SSDP.

Q-learning and a related temporal difference learning method, SARSA, were applied by (Rieker & Labadie, 2012) to the selection of a seasonal policy to manage release of stored water to benefit downstream water quality targets. The goal was to select a yearly release strategy from four alternatives based on beginning-of-season storage and forecasted peak storage, the latter being derived from analysis of basin snowpack. The was applied to the single reservoir Truckee river system in California and Nevada. The study considered both Q-learning and SARSA methods and variants of these methods which consider eligibility traces which store information on the states

that the algorithm has visited in order to update multiple preceding states rather than only the immediately preceding state. The results showed that a SARSA implementation with eligibility traces performed the best. Addition of the eligibility trace did not improve performance of the Q-learning method. An attempt was made to use an ANN to approximate the Q-function, though this did not improve the performance.

(Castelletti, Corani, Rizzolli, Soncinie-Sessa, & Weber, 2002) developed a method called Q-learning-planning in which an extra variable representing exogenous hydrologic conditions of the basin is included in the state. The tabular form of the Q-learning method is applied. The method also used an exhaustive approach to updating the Q-function, performing updates to each value of the Q-function at each state for every state transition. This was applied in a multi-objective case study of operations at Lake Como, Italy using the previous time period's flow as the hydrologic condition included in the state variable. The method was compared to an SDP solution which described the hydrologic inflow variable with a first-order auto-regressive model. The results showed that the method was able to successfully develop a Pareto front which dominated that developed by SDP.

The tabular form of the Q-learning algorithm can encounter long computing times when the discretization of the state is small. One remedy is to develop an approximation for the Q-function which can interpolate between discrete states. The study by (Rieker & Labadie, 2012) used an ANN as an approximator with moderate results. (Castelletti, Galelli, Restelli, & Soncini-Sessa, 2010) used a tree-based regression method, which is a non-parametric approximator. This approach was applied to the operation of Lake Como in Italy and showed improved performance over SDP. During the learning process, the regression model was trained in a batch mode in which a set of data points containing tuples of the (initial state, action, reward, next state) were collected using a Q-learning method.

A recent and rather famous application of Q-learning was developed by the Google Brain team for playing Atari games (Mnih et al., 2015). This study used a Q-learning method (deep Q-learning) which utilized deep convolutional neural network as a function approximator (deep Q-network). The authors note that divergence can be a problem when using a neural network for function approximation and applied two techniques for avoiding this situation. First, the batch learning method was modified to randomly sample experiences from the agent's memory (referred to as

experience replay) when training the ANN. The second step was to maintain two separate ANNs, one to be used during the learning process and the other which would be trained using experience replay. The trained ANN would periodically be copied to the learning ANN. The goal of this step was to reduce correlation of the action-values experienced and the target values computed. One of the goals of the study was to be able to design a single deep Q-network that could learn to play a wide variety of games. The deep Q-network was able to outperform previous algorithms on all of the games tested and performed at the level of a human professional in 75% of the games. One interesting observation is while the deep Q-network was able to excel at games that required a wide variety of strategies, the most challenging games were those that required temporally extended planning strategies.

Deep Q-learning is only applicable in situations where the action is selected from a finite set of choices. This method suffers from the curse of dimensionality when the number of possible actions increases. (Lillicrap et al., 2015) developed the Deep Deterministic Policy Gradients (DDPG) method which can be applied to continuous action spaces. This work uses the deterministic policy gradient method of (Silver et al., 2014) along with the stabilizing features of Deep Q-learning (Mnih et al., 2015) to develop an agent which learned to perform several physics-based tasks which require continuous control. DDPG uses an actor-critic method which estimates the policy function (actor), $\pi(s)$, as well as the action-value function (critic). The critic is updated based on gradient descent of the loss function of the temporal-difference errors as in Deep Q-learning. Updates to the policy function are based on the gradients within the $Q(s, a)$ function. Incremental updating is used in DDPG as well which requires an additional target ANN for both functions. The study showed that the method could be applied to high-dimensional problems with experiments including environments with as many as 37 state and 12 action dimensions. Other methods have been proposed for continuous control. (Mnih et al., 2016) proposed the asynchronous advantage actor-critic (A3C) method which estimates the state-value function, $V(s)$ instead of the action-value function. A3C performed well when using either the physical state of the system or pixels of an image as input.

2.4 Risk Analysis

Performance with respect to some objectives can be defined in quantities such as the amount of power produced, or the volume of supply delivered. Other objectives focus on the risk associated

with either desired or unwanted events. Because one of the objectives in the case study concerns flood control operations which judges the performance of the reservoir operations in binary terms of success or failure, a review of currently available methods for analyzing risk was performed.

Risk analysis is an important area of focus in civil engineering which requires study in many areas of engineering as well as decision-making, though a formalized process for such analysis does not yet exist, with even the concept of risk analysis lacking a singular definition (Faber & Stewart, 2003). A common definition of risk is the product of the probability of occurrence of an event and the consequence of that event (Faber & Stewart, 2003). To assess the consequences of an event, the hazardous events must first be identified, and techniques have been developed to do so such as Preliminary Hazard Analysis and Failure Mode and Effect Analysis (Faber & Stewart, 2003).

Recognizing that some of these events may be conditioned on multiple components of an engineering system, logic tree analysis focuses on studying the contribution of individual components to an overall risk (Faber & Stewart, 2003). Given an identified hazardous event, the consequence of that event then must be determined. This being quantified in terms of not only economic losses, but societal, environmental and human losses as well. To calculate the risk, the probability of this event is needed. This can be obtained by using classical reliability theory which considers large sets of similar components in independent loading events. The probability of failure being calculated based on the frequency of failure. Components in the civil engineering field are not easily analyzed in this manner as many systems are unique in design and loading situation. Probabilistic modeling must be applied to estimate the failure probability. This would include modeling the stochastic nature of the loading to the system. The determination of the acceptable level of risk allowed involves a subjective process that must consider economic and societal factors. Political matters may arise as the groups accepting higher risks may not be the same as those that will receive the benefits (Faber & Stewart, 2003).

Multi-attribute utility theory (Keeney, 1973; Keeney & Wood, 1977) attempts to translate a decision-maker's risk preferences into a utility function which can then be used when performing further analyses such as multi-criteria decision analysis. (Hashimoto, Stedinger, & Loucks, 1982) point out the difficulties that may arise when developing a utility function for a single decision-maker, and which would be compounded for multiple decision-makers. (Hashimoto, Stedinger, & Loucks, 1982) introduced three risk related performance measures: reliability (likeliness of failure), resiliency

(time to recovery) and vulnerability (severity of consequences). Computing these values involves performing simulation of the reservoir system forced by a selected synthetic or historical data set and counting the number of stages in which a criterion is violated and recording the magnitude of the violation. The performance measures are then calculated from this data.

Variations on these measures have been suggested such as time-based reliability, volume-based reliability, and steady-state reliability. (McMahon, Adeloye, & Zhou, 2006) performed a thorough analysis of these measures as well as sustainability and a drought risk index. The drought risk index being a score based on a weighted combination of reliability, resiliency and vulnerability. These three measures are related and conflicting. A decrease in the number of failures can lead to an increase in the length or severity of the deviation from the target. (Moy, Cohon, & ReVelle, 1986) performed an analysis of the trade-offs between risk measures that included a maximum shortfall measure. This approach relies on a pre-defined operating plan which can be simulated.

Early work in the study of the development of operating plans that considered risk explicitly includes the method known as chance constraint programming which resulted from research performed to determine heating oil production rates. In the investigation by (Charnes, Cooper, & Symonds, 1958), the objective is to minimize the expected value of the cost (stated to be equivalent to maximizing the expected value of the profit) subject to the expected demand and storage limitations. The study acknowledges that the demand for heating oil is a random variable due to its relationship with the weather, thus constraints on meeting customer demand cannot be met with certainty. Given an acceptable level of failure, a deterministic demand constraint is replaced with a probabilistic constraint. That is, the probability that the demand will be satisfied must be greater than the acceptable level. This explicitly acknowledges that failure to meet constraints is certain at some level, given the presence of random variables in the problem formulation. With a known distribution for the random variable, a probabilistic constraint can be replaced with a certainty equivalent constraint, given suitable properties of the form of the constraint function. There are many possible forms for these functions though the linear decision rule was one that had the most applicability. This work was followed by another study of the chance constrained method, (Charnes & Cooper, 1963), which introduced the “P model” and the “V model” The “P model” maximized the probability that the objective function would be greater than a certain value, and the “V model” minimized the mean squared error about a value of interest.

The interest in using the chance constrained method in water resource applications seems to have taken off with the work of (Revelle, Joeres, & Kirby, 1969) which used the linear decision rule to develop both the release rules as well as the design storage capacity of a reservoir system. (Revelle, Joeres, & Kirby, 1969) notes that the linear decision rule is not necessarily the only or best form of the rule, but it has the advantage of leading to linear programming problems which are easily solved. Another advantage identified is that "most importantly the chance-constrained formulation comes squarely to grips with the impossibility of absolutely ensuring the specific performance of a reservoir fed by random inputs."

Many other authors have investigated chance constrained programming. (Eisel, 1972) applied chance constrained programming to a similar reservoir design problem using a different form of the linear decision rule. The author noted the difficulties that can arise when determining deterministic equivalent constraints. This required some restrictions on parameter ranges to regions where linear approximations were valid. (LeClerc & Marks, 1973) applied chance constrained programming to an eight-reservoir system in Québec, Canada. The author concluded that the method was acceptable as a screening process though it was noted that the disadvantages of the method included that chance constrained programming does not consider the magnitude of the failure and considers all random variables uncorrelated in space and time. Other studies include (Askew, Yeh, & Hall, 1971), (Loucks & Dorfman, 1975), (Houck, 1979), (Houck & Datta, 1981), and (Joeres, Seus, & Engelmann, 1981).

Reliability programming (Colorni & Fronza, 1976; Simonovic & Mariño, 1980) is an extension of chance constrained programming which inserts a loss term into the objective function in order to optimize the risk level that the system is to achieve. In addition to known distributions of the random variables, this method also requires assumptions of convexity of the benefit or loss functions for the objectives. (Defourny, Ernst, & Wehenkel, 2008) presents a discussion of a risk-aware method for analyzing sequential decision processes. This method involves replacement of the expectation functional in an MDP with a parameterized functional. The parameter is then varied to provide results that favor more conservative decision-making. The author notes that the approach is limited in the number of functionals that can be used in this process. This method does require the knowledge of the model of the system. Also, the computation of the actual risk is performed after the generation of the policy.

In their survey of multi-objective sequential decision-making, (Roijers, Vamplew, Whiteson, & Dazeley, 2013) note that some authors have argued that consideration of risk necessarily leads to a multi-objective problem, in that the decision-maker must choose between the amount of reward and the probability of achieving that reward. This issue has come up in the study of shortest stochastic paths (SSP) in the field of artificial intelligence and machine learning which has some relevance to the proposed research. The SSP problem considers a true MDP (probability distributions for state transitions are known) and seeks to determine a path to a goal state. In the reservoir operation problem, the goal state may be any state if the time horizon is long enough, that is, achieving any storage at a distant time step could be considered a success in that the operations did not fail in the meantime. The objective can be formulated in terms of maximizing the probability of reaching the goal state, achieving the goal state with the lowest cost, or as a multi-objective problem in which case the two objectives may be scalarized, or one objective can be optimized conditioned on a minimum performance level of the other (Bryce, Cushing, & Kambhampati, 2007). (Bryce, Cushing, & Kambhampati, 2007) present a method for approaching this problem as a multi-objective optimization problem where non-dominated solutions are collected along branches of the path during a multi-objective dynamic programming analysis. A decision is made when the non-dominated solutions are identified for all of the branches leading to the starting state. This study also presents methods for performing efficient search of the state-space.

2.5 Considerations When Using Forecasts

There are several issues that must be considered when developing an optimization model that includes forecasting, such as the forecast horizon and the accuracy of the forecast. (You & Cai, 2008) define decision horizon and forecast horizon as: "the decision horizon may be defined as the initial periods in which decisions are not affected by forecast data beyond a certain period defined as the forecast horizon." (Takeuchi & Sivaarthitkul, 1995) performed a hypothetical analysis of the effect of forecast accuracy and forecast horizon on the performance of reservoir operations. The results found that an increase in the size of the reservoir compared to the mean annual flow reduced the effects of forecast error. For the larger reservoirs, use of nearly any forecast improved performance. Smaller reservoirs required significantly higher forecast accuracy to achieve performance comparable to using the long-term mean. It was suggested that this is due to the inability of a smaller system

to recover from the loss of benefit related to releases made based on forecasted precipitation that is not realized. The larger reservoirs also benefited from an increase in the forecast horizon, with smaller reservoirs showing a decrease in performance as the horizon increased past the short-term. (Maurer & Lettenmaier, 2004) observed the sensitivity of smaller reservoirs to forecast skill with the added observation that smaller reservoirs also receive more benefit from including forecasts.

(Simonovic & Burn, 1989) developed a method for determining an optimal forecast horizon using a multi-objective compromise programming model which considered the trade-off between forecast horizon and forecast accuracy. This process revealed that a fixed forecast horizon was adequate for periods of generally consistent conditions but a variable horizon was beneficial during transition periods such as seasonal changes.

(You & Cai, 2008) performed a theoretical analysis of the interaction of marginal utility with factors such as forecast horizon, water stress, inflow uncertainty and reservoir size. The study revealed a complex interaction between these factors with a higher marginal utility leading to a shorter forecast horizon. The marginal utility in turn being affected by other factors including inflow uncertainty, discount rate, stationarity of inflows, and water stress.

2.6 Forecast Performance Measures

The performance of an ensemble forecast can be measured in multiple ways and several studies have been performed to evaluate ensemble forecasting systems. (Alfieri et al., 2014) studied the performance of the European Flood Awareness System, (Roulin & Vannitsem, 2005) studied the Ensemble Prediction System of the European Centre for Medium-Range Weather Forecasts, (Fan, Collischonn, Meller, & Botelho, 2014) studied ensemble forecasts of inflows to the Três Marias hydroelectric power plant in Brazil and (Renner, Werner, Rademacher, & Sprokkereef, 2009) studied ensemble forecasts for flow along the Rhine River. These studies focus on the Brier Skill Score, the Ranked Probability Skill Score (or Continuous Ranked Probability Skill Score), reliability diagrams and rank histograms as performance metrics. (Renner, Werner, Rademacher, & Sprokkereef, 2009) observed that, as may be expected, skill decreased with extended lead time, however it was notable that this rate of decrease was accelerated for smaller basins.

The BSS and (C)RPSS require benchmark reference forecasts and (Pappenberger et al., 2015) performed an analysis of the use of different benchmark forecasts for comparison when determining

the added benefit that comes from a forecast. This is performed for only the threshold continuous ranked probability score ($CRPS^t$) performance measure. Results showed that benchmarks derived from meteorological analysis were harder to beat than those derived from hydrological data.

2.7 Summary

The search of the literature did not reveal any studies which had performed the analysis presented in this dissertation. Reinforcement learning has been applied in the past with (Lee & Labadie, 2007), (Rieker & Labadie, 2012), (Castelletti, Galelli, Restelli, & Soncini-Sessa, 2010) and (Pianosi, Castelletti, & Restelli, 2013) perhaps the most similar. This research differs in that the focus is the estimation of the state-value functions to be used along with the ensemble forecasts to ultimately make release decisions. This research is also distinguished by the use of continuous state and action spaces.

Chapter 3

Methods

This section will describe the methods used in the current research. As mentioned above, the problem of decision-making using forecast information separates into two sub-problems: determination of the state-value function in the no-forecast condition and decision-making during the forecast horizon. Reinforcement learning techniques are used to estimate the state-value function, and a particle swarm optimization (PSO) is utilized to determine release decisions given a forecast and considering multiple objectives. A case study is shown which applies this method to the Russian River basin in northern California which involves three criteria: flood management, water supply, and environmental flows.

3.1 Reinforcement Learning

The characteristics of this problem lead to the use of reinforcement learning as a means of estimating the state-value function. Reinforcement learning is a category of machine learning algorithms in which an intelligent agent is able to learn beneficial actions based on interaction with an environment. The intelligent agent is given a goal to optimize and the ability to take actions which, along with possible random input, result in changes to the environment. The agent is also able to observe a reward signal from the environment informing the agent of the value of the action taken and a state signal returning the resulting state of the environment.

3.1.1 Markov Decision Processes

The basis for the formulation of a reinforcement learning problem is the Markov Decision Process (MDP). In a finite MDP, the problem consists of a discretized state, $S_t \in \mathcal{S}$, a set of actions that can be taken in a given state, $A_t \in \mathcal{A}(s)$ and the rewards that can be realized $R_t \in \mathcal{R}$. The reward received and the resulting state of the system are related to the initial state and selected action through the system dynamics equation:

$$p(s', r|s, a) \doteq Pr\{S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a\}. \tag{3.1}$$

This equation gives the probability of finding the system in state s' and receiving a reward r after starting in state s and taking action a.

The agent is given a goal which, in this case, is to maximize the rewards obtained over a series of (possibly infinite) time steps. This is referred to as the return:

$$G_t = R_{t+1} + R_{t+2} + \cdots + R_T. \tag{3.2}$$

To avoid infinite returns and to represent the diminished value attributed to rewards that are received in the future, the concept of *discounted* returns is introduced and the return becomes

$$G_t = \sum_{\tau=0}^{T} \gamma^\tau R_{t+\tau+1}, \tag{3.3}$$

where $\gamma \in [0, 1]$ and T is allowed to increase to infinity. The objective then is to determine the set of actions which will maximize some function of the return, this function commonly being the expected value of the return:

$$J = \mathbb{E}[G_t | S_t = s]. \tag{3.4}$$

3.1.2 Value Functions and Optimal Policies

Methods that attempt to solve MDPs such as dynamic programming typically consider two value functions. The *state-value* function represents the expected value of the return based on a known state conditioned on selecting actions according to a policy π:

$$v_\pi(s) \doteq \mathbb{E}_\pi[G_t | S_t = s]. \tag{3.5}$$

In general, the policy is allowed to be stochastic in which case $\pi(a|s)$ is a mapping from a state to the probability of selecting action a. In the deterministic case, the policy maps from a state to an action and the notation is changed to use $\mu(s)$ to note the difference. The *action-value* function describes the expected reward for a given state conditioned further upon taking a particular action and following the policy π at all subsequent stages:

$$q_\pi \doteq \mathbb{E}_\pi[G_t | S_t = s, A_t = a]. \tag{3.6}$$

The solution to this problem takes advantage of the recursive property of the value functions known as the Bellman equations (Bellman, 1957). The Bellman equation for the state-value function is

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(s')|S_t = s]. \tag{3.7}$$

The agent has found the optimal policy π^* when it has found a policy such that

$$v^*(s) = \max_\pi v_\pi(s), \tag{3.8}$$

for all possible states. For an optimal policy, the action-value function shares this property:

$$q^*(s, a) = \max_\pi q_\pi(s, a). \tag{3.9}$$

The state-value and action-value functions are related by

$$v(s) = \max_a q(s, a). \tag{3.10}$$

Combining this notion of optimality with the recursive property of the value functions allows the agent to consider the problem as a greedy optimization of the return over the next stage plus the value of the resulting state, which is described by the Bellman *optimality* equations:

$$v^*(s) = \max_a \mathbb{E}[R_{t+1} + \gamma v^*(S_{t+1})|S_t = s, A_t = a], \tag{3.11}$$

$$q^*(s, a) = \mathbb{E}[R_{t+1} + \gamma \max_{a'} q^*(S_{t+1}, a')|S_t = s, A_t = a], \tag{3.12}$$

$$= \mathbb{E}[R_{t+1} + \gamma v^*(s')]. \tag{3.13}$$

With the state dynamics equation, 3.1, the optimal value functions, 3.11, 3.12 can be computed using dynamic programming which assumes an initial state value and iteratively solves the Bellman optimality equations in a backwards manner, updating the current estimate at each iteration. This continues until an acceptable level of convergence is obtained.

Implicit in the formulation of the problem is that the agent's current observation of the state contains all information necessary to describe the agent's environment. That is, the current state

must contain all of the information necessary to provide reliable probabilities to the state dynamics equation. The situation of having less than complete information about the state of the system is called a *partially observable state* in which case the problem becomes a Partially Observable MDP (POMDP). Much research has been conducted into the solution of POMDPs (Lovejoy, 1991).

A realistic analysis of a water resources problem may require consideration of the state as partially observable as many factors contribute to the transition from one state to the next, such as past precipitation and soil moisture, measurements of which are not always readily available. Along with the problem of partially observable states, is the problem of originally relying on probability distributions to describe the state dynamics equation. With more than a few states or random variables, or with short time steps, the transition probabilities for a water resource system can be inaccurate and unable to represent the temporal and spatial correlations of the random variables. Additionally, when a system is analyzed at fine temporal resolution, routing effects may lead to rewards being realized several time steps after the decision was made and this is known as *delayed rewards*. A probabilistic model is not able to handle this type of problem.

3.1.3 Temporal Difference Learning

Reinforcement learning is closely related to dynamic programming. One distinction is the use of a trial and error (Monte Carlo) approach (Sutton & Barto, 2018). Learning in this manner does not require knowledge of a probabilistic model of the environment and instead relies on interaction with the environment and reinforcing information provided through a reward. Methods which do not use probabilistic models are referred to as *model-free* methods. Another feature of some reinforcement learning methods is *temporal-difference* learning, which is used in the very successful method of Q-learning (Watkins & Dayan, 1992). In this method, the agent stores an estimate, $Q(s, a)$, of the $q(s, a)$ function and updates the estimate based on iteratively interacting with the environment and receiving a reward at each time step. The agent selects actions according to a policy that insures exploration such as the ϵ-greedy method, where for a value of $\epsilon \in [0, 1]$ the action $A_t = \text{argmax}_a Q(S_t, a)$ is selected with probability 1-ϵ, and a random action is selected with probability ϵ. The estimate is updated according to

$$Q(S, A) \leftarrow Q(S, A) + \alpha[R + \gamma \max_a \{Q(S', a)\} - Q(S, A)]. \tag{3.14}$$

The update increments the current estimate, $Q(S, A)$, in the direction of a new estimate by a step size α. The temporal difference error (TD-error):

$$\delta_t = R + \gamma \max_a Q(S', a) - Q(S, A), \tag{3.15}$$

is the difference between the previous estimate, $Q(S, A)$, and the new estimate made up of the single step reward and the best possible value of Q at the next state, $R_t + \gamma \max_a Q(S', a)$. Because this new estimate is based on the possibility of taking the action that maximizes Q, as opposed to the action determined by a fixed policy π, the method is referred to as an *off-policy* method.

For discrete action and state spaces, the Q function becomes a look up table and this is referred to as a *tabular* method. Q-learning has been shown to converge to the optimal state-value function in the tabular case (Watkins & Dayan, 1992). The convergence proof requires that all states be visited an infinite number of times which is not practical for implementation and methods such as ϵ-greedy action selection are used to ensure exploration of the state-space.

3.1.4 Regression with ANN

It will become advantageous to use an artificial neural network (ANN) to approximate our value functions and a brief discussion of ANNs is included here. An artificial neural network (ANN) can be used for both classification and regression. This work looks to ANNs to act as approximators for the policy, state-value and action-value functions. There are many possible configurations to consider when using ANNs for either classification or regression. This work will consider only fully-connected, feed-forward networks. This configuration is then a series of matrix calculations:

$$\mathbf{z}^{[l]} = \mathbf{W}^{[l]} \cdot \mathbf{a}^{[l-1]} + \mathbf{b}^{[l]}, \tag{3.16}$$

$$\mathbf{a}^{[l]} = g^{[l]}(\mathbf{z}^{[l]}), \tag{3.17}$$

which can be visualized as a network of interconnected nodes arranged into layers as shown in Figure 3.1.

Representing the state vector as a vertical vector of size $[N_0 \times 1]$, where N_0 is the number of state dimensions, the state vector becomes the input to the network by setting $\mathbf{a}^{[0]} = \mathbf{x}_t$. One set of these equations constitute one "layer" in the neural network. The final layer is the output of

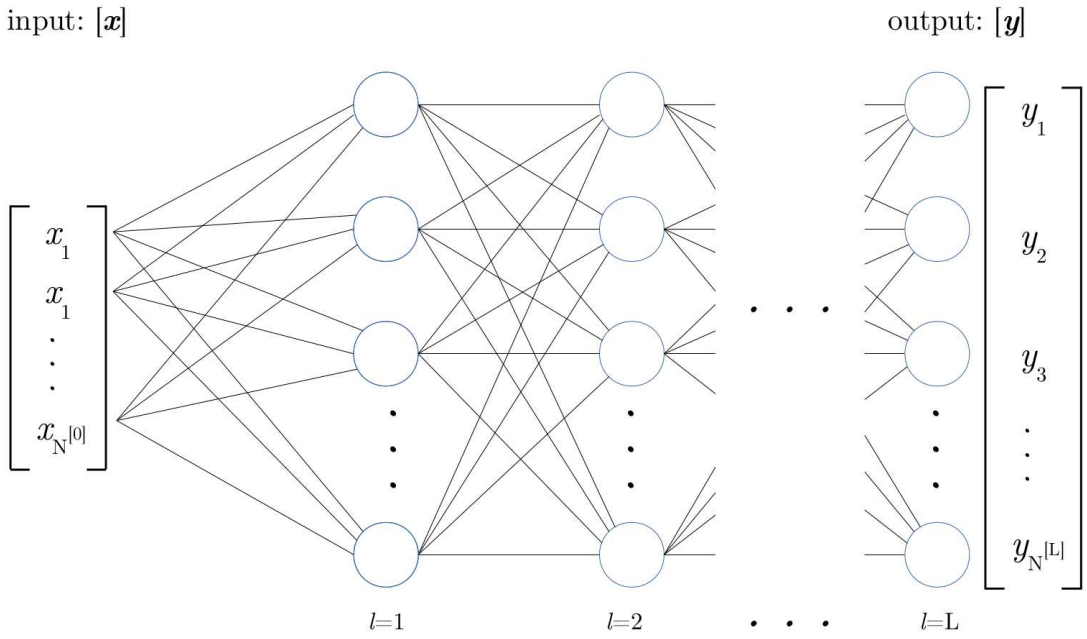

Figure 3.1: Schematic depiction of an artificial neural network. This configuration is a fully-connected, feed-forward network. Fully-connected means that every node is connected to every node in the adjacent layers. In a feed-forward network, the calculations only proceed in one direction from input to output through each layer.

the network and the prior layers are referred to as hidden layers. The set of weights, $\mathbf{W}^{[l]}$ is of dimension $[N^{[l]} \times N^{[l-1]}]$ where $N^{[l]}$ is the number of "nodes" in the layer. For the first layer ($l = 1$), $N^{[0]}$ is the number of dimensions of the state variable. The bias, $\mathbf{b}^{[l]}$, is of size $[N^{[l]} \times 1]$.

The functions $g^{[l]}$ are *activation* functions that map the transformed input to a single value output for each node. The output of the final layer is then $\mathbf{y}_t = \mathbf{z}^{[L]}$ and by setting the number of nodes in the final layer to one, the output is a vector with a single value. The set of all weights and biases $[\mathbf{W}^{[1]}, \mathbf{b}^{[1]}, \ldots, \mathbf{W}^{[L]}, \mathbf{b}^{[L]}]$ are referred to collectively as the parameters θ, and it is these values that are adjusted during the training process.

The activation functions can take many forms with the selection depending on the application. For this study, the hidden layers use a leaky rectified linear unit (LReLU) (also known as the parameterized rectified linear unit)(He, Zhang, Ren, & Sun, 2015). The LReLU was selected because it is easy to differentiate, does not suffer from the problem of vanishing gradients, and the leaky feature helps prevent "dying" nodes. The activation for the single node in the output layer is a

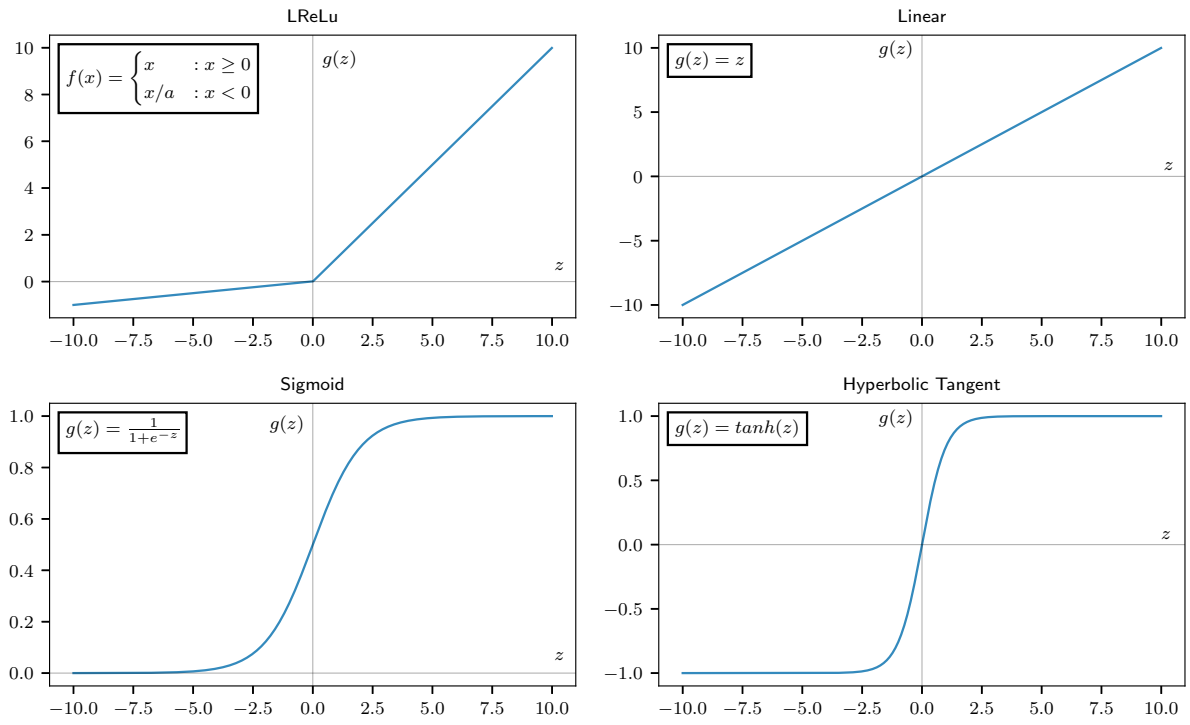

Figure 3.2: Four common activation functions. The LReLu activation function is used in all hidden layers with a value of $a = 100$. The linear function is used for the network output node. The sigmoid and hyperbolic tangent functions were used initially for the output of the action policy network to ensure the bounds of the max/min release constraints. This was not successful and it is believed that this was due to the property of these functions which have derivatives close to zero for large positive or negative values of z causing the learning process to become "stuck" at these values.

linear function that simply passes along the matrix transformation (Equation 3.16). This allows the output to take on any real number.

3.1.5 Continuous Action Methods

The tabular method suffers from the curse of dimensionality as the number of state and action variables increase and with finer discretization of the state and action spaces. It is then desirable to use a function approximator to serve as the estimate of the Q function. Artificial neural networks are considered to be universal function approximators, though the use of non-linear function approximators has been shown to be unstable (Tsitsiklis & Roy, 1997). (Mnih et al., 2015) overcame this instability using two techniques, a replay buffer and incremental updates, which reduce the effect of the correlation between samples. This method is only applicable to discretized actions and suffers from the same curse of dimensionality as the action space increases. Also, with the Q-learning update there is an embedded optimization step which can be time consuming. (Lillicrap

et al., 2015) combined the deterministic policy gradient method of (Silver et al., 2014) with the advances from deep q-learning to develop the deep deterministic policy gradient (DDPG) method which is able to utilize a continuous action space.

The DDPG method uses an actor-critic approach in which an estimate of the policy (actor) is maintained and updated based on gradients along the action axis within the estimate of the Q function (critic). The notation μ will be used now for deterministic policies and the approximation of the action-value function for policy μ given a parameter set θ^Q is denoted as $Q(s_t, a_t|\theta^Q)$. The critic, in this sense, is the Q function which is evaluating a policy μ.

Given a set of observed state transitions and the corresponding rewards, the parameters θ^Q are adjusted by minimizing the loss function:

$$L(\theta^Q) = \mathbb{E}[(y_t - Q(s_t, a_t|\theta^Q))^2], \tag{3.18}$$

where:

$$y_t = r_{t+1} + \gamma Q(s_{t+1}, \mu(s_{t+1})|\theta^Q), \tag{3.19}$$

is the new estimated value for the action-value function based on the new observation and the action-value at the resulting state with the current parameter set θ^Q and conditioned on selecting the action according to policy μ at the next state. The policy function is updated based on the approximation of the policy gradient from (Silver et al., 2014):

$$\nabla_{\theta^\mu} J \approx \mathbb{E}[\nabla_{\theta^\mu} Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t|\theta^\mu)}] \tag{3.20}$$

$$= \mathbb{E}[\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s=s_t}]. \tag{3.21}$$

Experience Replay

Experience replay is one of two techniques used by (Mnih et al., 2015) to stabilize the reinforcement learning process when ANNs are used to approximate the value functions. This method creates a buffer, B, of size B_{max}, which stores memories of the agent's experiences as it takes actions, and the subsequent state transitions that occur as a result of these actions. A single memory is then the tuple (S, A, R, S'). Mini-batches of memories are randomly selected from the buffer and used when

training the ANN. When the buffer is full, the oldest memories are discarded and replaced by the most recent.

Incremental Updates

The other technique used by (Mnih et al., 2015) to stabilize the learning process is an incremental update process in which a target ANN is used to compute the temporal difference which is used to train the primary ANN. This reduces the correlation between the temporal difference and the policy that selected the action which resulted in a particular reward. This can be implemented as a complete replacement which occurs after a given number of updates to the primary ANN or as a partial update in the form:

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta', \tag{3.22}$$

where $\tau \in (0, 1)$.

The implementation of DDPG for the present application is shown in Algorithm 1. Sections 3.1.6 and 3.1.7 define the state and action spaces and the boundary conditions that are specific to the case study. Additional refinements to the DDPG algorithm that were needed for this research are described in Sections 3.1.8-3.1.11.

3.1.6 State and Action Spaces

The state variable describes properties of the system under consideration and in a reservoir system typically contains at least the storage of one or more reservoirs. This research considers a single reservoir system, and the first property of the state vector is the storage in the reservoir at the start of the time step. The storage is normalized, with the limits $[0, 1]$ corresponding to the storage levels which are designated as *terminal* or *boundary* states. Terminal and boundary states are discussed in more detail in Section 3.1.7. For now, let c'_t be the unnormalized storage in the reservoir. The first value in the state vector is then

$$c_t = \frac{c'_t - c_{min}}{c_{max} - c_{min}}, \tag{3.23}$$

with c_{min} and c_{max} representing the storage at the terminal or boundary states which constrain the minimum and maximum allowable storage.

Algorithm 1 DDPG for the Single Reservoir Problem

For maximum n-steps N, and maximum episode length E
Initialize replay buffer, B, of maximum size B_{max}
Initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^{\mu)}$ with weights θ^Q and θ^μ
Initialize target networks Q' and μ' with weights $\theta^{Q'} = \theta^Q$ and $\theta^{\mu'} = \theta^\mu$
Initialize indicator variable i
loop
    Randomly initialize state variable $\mathbf{S}_{i-g}$
    Warm up routed flows in the network for g time steps
    Uniformly resample storage, c^*, update $\mathbf{S}_i \leftarrow [c^*, d_{1,i}, d_{2,i}]$
    Initialize n-step counter $n = 1$, set $t = i$
    for $j = t, t + E$ **do**
        With probability ϵ, select random action A_j, otherwise $A_j = \mu(\mathbf{S}_j)$
        Retrieve random variable input $\mathbf{W}_j$ from synthetic data generator
        Observe ($\mathbf{S}'$,R_{j+1}) from state transition $T(\mathbf{S}_j, A_j, \mathbf{W}_j)$
        if $n = N$ **or** $\mathbf{S}'$ is terminal **then**
            Calculate the n-step rewards $R_{nstep} = \sum_{k=-(n-1)}^{0} R_{j+1+k}$
            if $\mathbf{S}'$ is terminal **then**
                Add $(\mathbf{S}_{j-n+1}, A_{j-n+1}, R_{nstep}, \mathbf{S}', n)$ to B, p times ▷ Prioritized replay
                Uniformly resample storage, c^*, update $\mathbf{S}' \leftarrow [c^*, d'_1, d'_2]$
            else
                Add $(\mathbf{S}_{j-n+1}, A_{j-n+1}, R_{nstep}, \mathbf{S}', n)$ to B
            end if
            $n = 1$
        else
            $n = n + 1$
        end if
        Sample minibatch $(\mathbf{S}_k, A_k, R_{nstep,k}, \mathbf{S}'_k, n_k)$ of size M from B
        if $\mathbf{S}'_k$ is terminal **then**
            Get return $G_{BC,k}$ from boundary conditions
            $y_k = (R_{nstep,k} + \gamma^n G_{BC,k})/p$ ▷ Bias correction
        else
            $y_k = R_{nstep} + \gamma^n Q'(\mathbf{S}_k, \mu'(\mathbf{S}_k))$
        end if
        Update critic by performing gradient descent on L (Eq. 3.18)
        Update actor using the policy gradient (Eq. 3.21), apply inverted gradients to $\nabla_a Q(\cdot)$
        Update target networks
        Update $\mathbf{S}_{j+1} = \mathbf{S}'$
        Increment $i = i + 1$
    end for
end loop

This study also considers the calendar day as a state variable similar to (Castelletti, Galelli, Restelli, & Soncini-Sessa, 2010) but with a variation to make the temporal dimension periodic. This is done by representing the calendar day as a Cartesian transformation of the calendar day represented in polar coordinates on a circle with radius 1, and angle described by an indicator variable. These are subsequently normalized so that all values fall into the range $[0, 1]$. That is, each calendar day is assigned an indicator $i \in [0, 364]$ with January 1 having a value $i = 0$ (note that zero-based indexing is used for compatibility with the implementation in code). The state variable then represents the day as

$$\mathbf{d}_i = [d_{1i}, d_{2i}] = \left[\frac{\cos\left(\frac{2\pi i}{365}\right) + 1}{2}, \frac{\sin\left(\frac{2\pi i}{365}\right) + 1}{2} \right]. \tag{3.24}$$

The state transition of the indicator is according to the rule

$$i_{t+1} = \begin{cases} 0 & \text{if } i_t = 364 \\ i_t + 1 & \text{otherwise.} \end{cases} \tag{3.25}$$

Throughout the study, leap days are omitted. The representation of the temporal dimension of the state space in polar coordinates creates a cylindrical state space with only the storage limits as boundaries, the first and last day of the year are adjacent. This ensures that the function approximations are continuous at the step from December 31 to January 1. Thus, the state variable is a three-dimensional vector $\mathbf{s}_t = [c_t, d_{1t}, d_{2t}]$.

The action that the agent may take is the daily volume of release from the reservoir. This again is normalized to the range [0,1] corresponding to the minimum and maximum releases based on the physical or operational constraints of the system. This is a single reservoir system and thus the action variable is one-dimensional, $\mathbf{a}_t = [a_t]$. The input to the policy function, $\pi(\mathbf{s})$, is the three-dimensional state variable. The input to the action-value function, $Q(\mathbf{s}, \mathbf{a})$, is a 4-dimensional concatenation of the state and action variables:

$$\mathbf{x}_t = [\mathbf{s}_t, \mathbf{a}_t] = [c_t, d_{1t}, d_{2t}, a_t]. \tag{3.26}$$

3.1.7 Boundary Conditions

The situation of a system that potentially has both episodic and continuing tasks has been dealt with by formulating a unifying notation (Sutton & Barto, 2018) in which all episodes are modeled as continuous by considering a terminal state as an absorbing state which has the property of only transitioning to itself and returning a fixed reward . In (Sutton & Barto, 2018), this reward is a value of zero which considers the agent reaching this state never to leave and subsequently never to again receive a reward.

In this research, a terminal state is used for states which are considered failures according to the particular objective under consideration. In the case of flood control, this is any state above the maximum allowable storage level. In the case of the environmental flow and water supply objectives, the storage below the lower storage limit is a terminal state. As a verification step, the agent is given an objective which is a combination of the flood control and environmental flow objectives. For this objective, both the upper and lower storage limits are terminal states.

For the flood control objective, the reward structure differs from the environmental flow and water supply objectives in that the agent receives a reward of zero for any state transition that does not exceed the maximum storage limit and a penalty of -1 otherwise. Because the terminal state is an absorbing state which restricts all subsequent state transitions to return to the terminal state, according to this model, the agent also receives a penalty of -1 at all subsequent state transitions. To implement this, the state-value function for all terminal states is given a value of

$$V(s : s \in \mathcal{S}^{-}) = \sum_{t=o}^{\infty} \gamma^{t}(-1) = \frac{-1}{1-\gamma}, \tag{3.27}$$

so that an episode which reaches a terminal state after n steps has a return of

$$G_t = \gamma^0 0 + \gamma^1 0 + \cdots + \gamma^{n-1} 0 + \gamma^n \sum_{\tau=0}^{\infty} \gamma^{\tau}(-1) = \gamma^n \left(\frac{-1}{1-\gamma} \right). \tag{3.28}$$

Because an objective could reach a physical limit of the system which is not also considered a failure, the concept of a *boundary state* is introduced. Consider the objective of flood control, with the agent receiving a reward of 0 for any state transition that does not result in storage exceeding the maximum desired level and a reward of -1 for any state transition that does exceed the maximum.

It is possible that the agent would empty the reservoir and, because of the fixed minimum release, the storage could become negative. The storage has encountered the physical limits of the reservoir system, but it is not a failure according to this objective. The reservoir is empty, but the maximum storage limit has been met, and we have no need to terminate this episode. In the physical system, the reservoir would reach a lower limit and no additional water would be released until inflow to the reservoir increases the storage level again.

A *boundary state* is used to model this limit. A boundary state has the property of intercepting a state transition that would cross the boundary and instead the final storage is set to the level of the boundary state. In contrast to a terminal state, a boundary state allows the next state transition to leave the boundary state if the state transition results in a final storage on the acceptable side of the boundary. The actual implementation in the code terminates any episode that reaches a boundary state and the return for the episode is the discounted sum of the n-step rewards and the value of the state-value function at the boundary state.

3.1.8 n-step Temporal Difference Learning

In a simple application of temporal difference learning, the TD-error is computed using a single step reward:

$$\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t), \tag{3.29}$$

where $R_{t+1} + \gamma V(S_{t+1})$ is the new estimate of the value function at state S_t, and $V(S_t)$ is the current estimate. The speed of the learning process can be increased by allowing the agent to collect consecutive rewards to use in computing the new estimate of the value function (Sutton & Barto, 2018):

$$\delta_t = (R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{n-1} R_{t+n} + \gamma^n V(s_{t+n})) - V(s_t). \tag{3.30}$$

This can be applied to any temporal difference learning algorithm such as Q-learning, SARSA, or TD(λ). When this is the case, the algorithm name is given the preface "n-step" (i.e. n-step Q-learning)

3.1.9 Inverted Gradients

The action that the agent is allowed to take is the release made from the reservoir. This naturally has physical limits on the values that it can take. More restrictive limits may also be imposed according to operational constraints. ANNs used for regression, when the output is bounded, often use a sigmoidal activation function which bounds the output between $[0, 1]$ or the hyperbolic tangent (tanh) function which bounds the output to a value of $[-1, 1]$ (see Figure 3.2). This value then can be scaled to the bounds of the release.

One of the drawbacks of using the sigmoid or hyperbolic tangent is the problem of *vanishing gradients*. The learning process relies on the partial derivatives of the output, y_t, with respect to the parameters θ. If the current approximation is high or low using one of these functions, the partial derivatives can be extremely small which can lead to slow learning. In the extreme case, the partial derivatives can even be below the smallest possible value which can represented by a float variable in the CPU. One solution to this problem is to use the method of *inverted gradients* (Hausknecht & Stone, 2015) which modifies the computed gradient depending on the current value of the parameter relative to the bounds according to the rule:

$$\nabla_y = \nabla_y \cdot \begin{cases} (y_{max} - y)/(y_{max} - y_{min}), & \text{if } \nabla_y \text{ suggests increasing } y \\ (y - y_{min})/(y_{max} - y_{min}), & \text{otherwise.} \end{cases} \tag{3.31}$$

This has an advantage over using a "squashing" function such as the hyperbolic tangent or the sigmoid as it will limit updates to the parameter if the gradient will increase or decrease the output beyond the bounds, and will quickly move away from the bound if the gradient suggests this. This prevents the weights and biases from becoming "stuck" at the boundaries (Hausknecht & Stone, 2015).

3.1.10 Prioritized Replay

The frequency with which the agent obtains a reward can be rare in some cases. In the case study used for this research, the nature of the inflow to Lake Mendocino is that of large but infrequent storms. As a result, the agent was able to meet the maximum storage limit a majority of the time with any policy. The buffer of stored memories is then dominated by state transitions which do not

result in failure. Training of the Q function uses mini-batches of memories to stochastically update the parameters θ^Q based on minimizing a loss function of the squared errors between the current estimate of the Q function and the new estimate based on the observations. The reward that the agent is given is zero for all cases, except when failure occurs. In which case, the reward is a unit penalty, and the return for the entire episode is

$$G_t = \sum_{\tau=0}^{N} \gamma^\tau R_{t+\tau+1} = 0 + 0 + \cdots + \gamma^N \left(\frac{-1}{1-\gamma} \right), \tag{3.32}$$

where N is the number of steps in the episode. This can be become a very small value as an episode is made up of many successes before failure occurs. Also, for stability, these updates only increment the parameters of θ^Q by a limited step size in the direction of the gradient. Because of this, when a mini-batch does contain a failure, the adjustment to the parameters can be truncated and without sufficiently frequent samples representing a failure, the influence of those rare samples which suggest large changes in the estimate of the Q function can vanish.

These rare events are of great concern when considering flood control operations. To address this issue, a form of *prioritized replay* is implemented in the learning process. Prioritized replay assigns a priority to the importance of a memory and then uses that priority to increase the frequency with which the memory is resampled from the buffer. The implementation here was inspired by the work of (Narasimhan, Kulkarni, & Barzilay, 2015) which saved memories into two sets based on a binary outcome. Batches for training then were made up of a fixed proportion of samples from each set. The present application inverts this process by populating the buffer with p, ($p \in \mathbb{Z} : p \geq 1$), records of the same memory when failure occurs, while placing a memory that does not contain a failure into the buffer once. Resampling mini-batches from the replay buffer is then done uniformly. (Schaul, Quan, Antonoglou, & Silver, 2015) implemented prioritized replay using the TD-error to prioritize resampling. That study also addressed the issue of bias generated from changing the distribution of the training samples. The DDPG algorithm and other RL algorithms estimate the expected value of the return, and thus require that the training samples represent the same distribution as the returns received during the learning process (Schaul, Quan, Antonoglou, & Silver, 2015). (Schaul, Quan, Antonoglou, & Silver, 2015) proposed a method referred to as *annealing the bias* to reduce this effect. In the current application, a simplified version of this method is used in which the target

estimate of the Q function (Eq. 3.19) is scaled by $1/p$. This simplification is possible because of the binary nature of the reward.

3.1.11 Maximum Episode Length

In this study, the agent is considered to be in an episodic environment with the length of the episode being a random variable. An episode is considered to end when the agent reaches a terminal state. The starting state is chosen randomly from a uniform distribution, and the agent was able to quickly learn actions that would avoid the boundaries for many action steps, leading to very long episode lengths. This led to the majority of the samples representing a small portion of the state space. In order to provide more diversity to the sample set, the episodes were limited to a maximum length.

This was implemented by carrying out the learning process in "sessions", where a session consisted of an initial series of state-transitions to warm-up the routed flows, followed by a fixed number of state-transitions with actions selected from the current policy. If the agent encountered a terminal or boundary state, the episode was ended and a new one was begun with the agent randomly assigned a value for the storage component of the state vector. The temporal component was retained and advanced one day. When the maximum number of transitions was reached, the episode was ended and the session concluded. A new session would then begin with all agents randomly assigned a storage value, the temporal component again being retained.

3.2 Ensemble Based Decisions

3.2.1 Separation of the Problem

Because this is a sequential decision process, we can examine the problem of decision-making using ensemble forecasts as a separable problem of rewards accumulated during the forecast horizon plus the value of the state at the end of the forecast horizon. That is, consider the return as the (possibly infinite) sum of a series of rewards

$$G_t = \gamma^0 R_t + \gamma^1 R_{t+1} + \gamma^2 R_{t+2} \cdots , \tag{3.33}$$

which is equal to the sum of the first reward and the series of following rewards:

$$G_t = \gamma^0 R_t + \gamma^1 \sum_{\tau=1}^{T} \gamma^{\tau-1} R_\tau. \tag{3.34}$$

We can extend this separation for the number of stages in the forecast horizon, H:

$$G_t = \gamma^0 R_t + \gamma^1 R_{t+1} + \cdots + \gamma^H R_H + \gamma^{H+1} \sum_{\tau=H+1}^{T} \gamma^{\tau-H-1} R_\tau. \tag{3.35}$$

We end up with a series of rewards of length equal to the forecast horizon plus the discounted series of rewards at the next stage past the forecast horizon:

$$\sum_{\tau=H+1}^{T} \gamma^{\tau-H-1} R_\tau = G_{H+1}. \tag{3.36}$$

The problem can then be separated to consider the time steps during and after the forecast horizon with two different methods. The value functions:

$$V(s) = \mathbb{E}[G_t | S_t = s], \tag{3.37}$$

$$Q(s, a) = \mathbb{E}[G_t | S_t = s, A_t = a], \tag{3.38}$$

can be obtained through reinforcement learning and provide the expected value of the return, G_t, for any state in a no-forecast condition, which is the value of the return after the forecast horizon (Equation 3.36). The total return is then the sum of this value and the rewards obtained during the forecast horizon. An optimization technique can then be applied to find the sequence of decisions which optimize this return.

3.2.2 Particle Swarm Optimization

The process of selecting a decision when presented with an ensemble forecast is not predetermined by the problem. The information provided by an ensemble must be interpreted in some fashion in order to be useful. One common assumption is to assume a uniform distribution of probabilities so that for an ensemble of J scenarios, each is assumed to have a $1/J$ probability of occurrence as in (Faber & Stedinger, 2001). The uniform distribution can be misleading as any scenario has near zero probability of occurring. Other methods attempt to fit other forms of probability distributions to

the scenarios. Creating a distribution from the ensemble, though, removes the temporal correlation that is contained in a sequence of flows, and the spatial correlation among a set of ensembles. It was desired, then, to find a method which could keep the ensembles intact and avoid assuming a probability distribution.

Considering a scenario as a deterministic forecast, the problem becomes one of determining the optimal sequence of decisions for each scenario. This is the sequential decision process from Equation 3.35 with the final term, Equation 3.36, provided as a function of the state of the system computed through the reinforcement learning application. This is then a deterministic problem, though it cannot be computed using the familiar backward solution process of dynamic programming because the stream routing effects lead to delayed rewards.

In order to compute the optimal decisions, the particle swarm optimization method (PSO) (Kennedy & Eberhart, 1995) was used with the inclusion of the inertia feature proposed by (Shi & Eberhart, 1998). PSO is an heuristic optimization algorithm which searches using a population of particles which move through the decision space and are endowed with both individual and social cognition. A particle holds a memory of the location and objective value (fitness) of the best performing location it has visited. The particle also has access to the position and fitness of the best performing location that the swarm as a whole has visited. The particle's velocity is adjusted based on the distance from these two points and the particle's own inertia, which acts to maintain the current velocity. Randomness is introduced into the velocity update to encourage exploration.

For this application there is an optimization problem to be solved for each scenario. If we label the scenarios $s = 1 \dots S$, and the individual particles $p = 1 \dots P$, each with a set of decisions, one for each day in the forecast horizon $h = 1 \dots H$, then we have a set of S arrays of size $[P \times H]$. Each array represents one swarm and contains the current position of the particle, $X^{[s]}$. Each particle has a velocity in the decision space with the velocities of all the particles in a swarm held in the array $V^{[s]}$. The best fitness value a particle has experienced is stored in the set of vertical vectors $F^{[s]}_{Pbest}$ of size $[P \times 1]$ with the locations at which this was achieved held in the array formed from the position vectors, $X^{[s]}_{Pbest}$ of size $[P \times H]$. The best fitness value that each swarm has experienced is held in a vertical vector F_{Sbest} of size $[S \times 1]$ and the corresponding positions in the array X_{Sbest} of size $[S \times H]$.

The particles are given random initial positions sampled uniformly from the decision space. The decision is the normalized release from the reservoir and so each dimension is constrained to the limit of $[0, 1]$. The arrays X_{Sbest} and $X^{[s]}_{Pbest}$ are also initialized randomly and the vectors of F_{Sbest} and $F^{[s]}_{Pbest}$ are given negative values below the lowest possible fitness to ensure that the first update will replace the positions in X_{Sbest} and $X^{[s]}_{Pbest}$.

The velocity of each particle is updated based on its own inertia plus a randomized influence from both its own best location and the swarm's best location:

$$V_i^{[s]} \leftarrow wV_i^{[s]} + \text{rand}_1()c_1(X^{[s]}_{Pbest,i} - X_i) + \text{rand}_2()c_2(X^{[s]}_{Sbest} - X_i), \tag{3.39}$$

with $v_i^{[s]}$ identifying the velocity of particle i in swarm s, w being a parameter representing the particle's own inertia, c_1 and c_2 are parameters determining the influence of the particle's best remembered position and the swarm's best position respectively. The functions $\text{rand}_1()$ and $\text{rand}_2()$ return random values uniformly sampled from the range $[0, 1]$ at each iteration. The positions are incremented by the new velocities:

$$X_i^{[s]} \leftarrow X_i^{[s]} + V_i^{[s]}. \tag{3.40}$$

The fitness of each particle is evaluated and the memories $X^{[s]}_{Pbest}$, X_{Sbest}, $F^{[s]}_{Pbest}$, and F_{Sbest} are replaced if the new location provides improvement.

Each swarm is solving a different optimization problem but the problems are closely related in that many of the scenarios share similarities throughout the forecast horizon. This property allows us to add the additional step of "communication" between storms where the position of the particle with the lowest fitness is updated to the position of the best particle from another swarm selected at random.

3.3 Modified-Puls Routing

The modified-puls routing method is based on a level-pool storage model with flow out of the reach based on the storage in the reach. The method relies on the development of a storage-discharge relationship which is particular to the stream reach. The method considers the mass balance equation:

$$I - O = \frac{dS}{dt}, \tag{3.41}$$

where I is the inflow to the reach, O is the flow out of the reach, and S is the storage in the reach. This can be represented in discrete form as

$$\frac{1}{2}(I_1 + I_2)\Delta t - \frac{1}{2}(O_1 + O_2)\Delta t = S_2 - S_1. \tag{3.42}$$

This is then further rearranged to obtain

$$\left(\frac{2S_2}{\Delta t} + O_2\right) = (I_1 + I_2) + \left(\frac{2S_1}{\Delta t} - O_1\right). \tag{3.43}$$

The storage-discharge relationship of the reach $S_t = f(O_t)$ is modified in this procedure to consider instead the relationship $\frac{2S_t}{\Delta t} + O_t = f(O_t)$. This function and its inverse $O_t = f'(\frac{2S_t}{\Delta t} + O_t)$ are then used iteratively to solve for outflow from the reach. This is done by assuming an initial state of $O_{t=0} = O_0$, $S_{t=0} = S_0$), and, given the series of inflows I_0, I_1:

1. Set $t = 0$
2. Find $\left(\frac{2S_t}{\Delta t} + O_t\right) = f(O_t)$
3. Calculate $\left(\frac{2S_t}{\Delta t} - O_t\right) = \left(\frac{2S_t}{\Delta t} + O_t\right) - 2O_t$
4. Using equation 3.43 calculate $\left(\frac{2S_{t+1}}{\Delta t} + O_{t+1}\right) = (I_{t+1} + I_t) + \left(\frac{2S_t}{\Delta t} - O_t\right)$
5. Use the inverse of f to find the outflow $O_{t+1} = f'\left(\frac{2S_{t+1}}{\Delta t} + O_{t+1}\right)$
6. Increment $t \leftarrow t + 1$ and continue.

The requirement for performing step 2 is particular to the implementation in code for the current project as only the streamflow, O_{t+1}, and not the storage information, S_{t+1}, is passed to the next iteration.

3.4 Synthetic Data

Reinforcement learning depends on the ability of the agent to interact with the environment and observe a state transition as a result of the actions of the agent as well as realizations of the

random variables for a particular timestep. This process must be repeated over a simulated time period much longer than the available data set. This requires that a method of generating synthetic data must be used to provide the realizations of the random variables.

The number of iterations in the learning process is not something that is known beforehand and so some method of continuous generation of synthetic data is needed. The data for each node k is available from 1950 - 2010. The forecasts are available from 1985-2010, so we want to be able to use 1985-2010 as our test data, this means that the years 1950-1984 will make up the training data. The sequence of time steps in this data set is given an index beginning at $i = 0$ on Jan 1, 1950 and incrementing sequentially through December 31, 1984 ($i = 9124$, leap days are excluded), so that the vector $\mathbf{w}_i = [w_{i,1}, \ldots, w_{i,k}, w_{i,k+1}, \ldots w_{i,k+l}]$ represents the inflow and loss at each node in the network.

A baseline counter in the range $j \in [0, 364]$ is established with property

$$b_{t+1} = \begin{cases} 0 & \text{if } b_t = 364 \\ b_t + 1 & \text{otherwise.} \end{cases} \tag{3.44}$$

and this increment occurs at every state transition. A random variable q is introduced with a uniform distribution $P(q) = 1/Q$ where $Q = 35$, the total number of calendar years in the training data. Another random variable r with a normal distribution:

$$P(r) = \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}, \tag{3.45}$$

is defined which will provide a temporal offset. The input vector for step t in the learning process is then selected according to Algorithm 2.

Algorithm 2 Synthetic Data Generator

Select parameters μ_1, σ_1^2, $\mu_2 = 0$, σ_2^2, ϵ
For data series $\mathbf{W}$ containing a sequence of T vectors, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2 \ldots, \mathbf{w}_T]$
of the random input, and corresponding to Q years:
initialize $t = 0$, $b_0 = 0$, $q_0 \sim \mathcal{U}\{0, Q\}$, $r_0 = 0$
loop
    $t \leftarrow t + 1$
    $b_t \leftarrow \begin{cases} 0 & \text{if } b_{t-1} = 364 \\ b_{t-1} + 1 & \text{otherwise} \end{cases}$
    sample $x \sim \mathcal{U}(0, 1), q \sim \mathcal{U}\{0, Q\}, r \sim \mathcal{N}(\mu_1, \sigma_1^2), \mathbf{p} \sim \mathcal{N}(\mu_2, \sigma_2^2), \dim \mathbf{p} = \dim \mathbf{w}_i$
    if $x < \epsilon$ **then**
        $q_t \leftarrow q$
        $r_t \leftarrow r$
    else
        $q_t \leftarrow q_{t-1}$
        $r_t \leftarrow r_{t-1}$
    end if
    $i' = b_t + r_t + 365 q_t$
    $i_t = i' \mod T$
    $\mathbf{w}_{synth,t} \leftarrow \mathbf{w}_i \odot \mathbf{p}$
end loop

Chapter 4

Case Study

4.1 Overview

The case study used to show the application of the operations method developed in this research is the Upper Russian River Basin in northern California. Lake Mendocino is a multi-purpose reservoir created by the construction of Coyote Valley Dam in 1959. This reservoir is located on the east fork of the Russian River in the northern part of the watershed and regulates the flow from 105 square miles of the 1,485 square mile area of the basin (Delaney & Mendoza, 2016). The project that developed Lake Mendocino originally intended for the construction to occur in two stages with the first developing 122,500 af of storage and the second to enlarge the storage to 199,000 af (USACE, 1986). The second phase has not been implemented and the current storage of the reservoir is 116,500 af (Delaney & Mendoza, 2016). The upper portion of the basin is considered to be the area above the confluence with Dry Creek. Lake Mendocino is operated for flood control with the available conservation storage used to benefit the water supply and environmental flow needs of this region. Lake Sonoma is located on Dry Creek above the confluence with the Russian River. Lake Sonoma has a capacity of 242,000 af and is used to provide water supply and environmental flows for the stream reach along Dry Creek and the Russian River below the confluence (Figure 4.3). The basin is subject to distinct wet and dry seasons. During the wet season of October to May approximately 93% of the precipitation occurs, most of which comes as a result of the "atmospheric river" phenomenon which brings large amounts of precipitation in a short period of time (Delaney & Mendoza, 2016).

4.2 Reservoir Operation

The United States Army Corp of Engineers (USACE) operates Lake Mendocino for flood management according to the guide curve defined in the water control manual (Figure 4.3) (USACE, 1986). The conservation storage is limited to 68,400 af during the wet season to reserve storage space for flood control. The conservation pool is increased during the spring to 111,000 af (Delaney & Mendoza, 2016). When storage levels encroach into the flood storage pool during the winter, flood

release operations are defined by three schedules which define maximum releases. Flood control releases are also restricted when the flow at the downstream gage near Hopland exceed 8,000 cfs.

The Sonoma County Water Agency is the non-federal sponsor of both Lake Sonoma and Lake Mendocino reservoirs and manages releases from the conservation pool in each reservoir. There are multiple uses for the releases in municipal, agricultural and environmental categories. The primary agricultural producers are the vineyards in the basin which divert water for frost protection, irrigation and heat suppression (Delaney & Mendoza, 2016). There are also public water systems that rely on the flows downstream of Lake Mendocino.

Additionally, the SCWA is responsible for ensuring that downstream flows meet minimum environmental limits determined by the Hydrologic Index set forth in the State Water Resources Control Board's (SWRCB) Decision 1610 and the Biological Opinion issued by the National Marine Fisheries Service (NMFS). The minimum flow in the downstream reaches is determined by a set of rules which consider the storage in both Lake Mendocino and Lake Pillsbury as well as the cumulative flow into Lake Pillsbury throughout the year. The Hydrologic Index assigns a category of *Normal*, *Dry*, or *Critical* to the condition in the basin based on the cumulative flow into Lake Pillsbury. Within each category, the minimum flow in the stream reaches below Lake Mendocino are set based on the total storage in both Lake Pillsbury and Lake Mendocino with the flow never to fall below 25 cfs at any time. This rule set is depicted in Figure 4.2 (Fleming, 2012).

The management of releases according to the flood control schedules was interpreted as a method for relating storages with higher risk of exceeding the maximum storage limit with a corresponding higher level of allowable release. Because the urgency of a release in the method presented is informed by the forecasts, the release schedules were ignored with only a maximum release applied. This avoids conflicts that could arise, such as if the forecasts indicated high inflows when storage was in the first flood control schedule and the system found it necessary to make large releases.

4.3 Potter Valley Project

The Potter Valley Project (PVP) is a hydropower generation project owned by Pacific Gas and Electric (PG&E) which stores water in Lake Pillsbury in the Eel River Basin for subsequent release and diversion to the Russian River Basin where the power generating facility is located (Delaney & Mendoza, 2016). The project is allowed to divert up to 300 cfs subject to a license from the

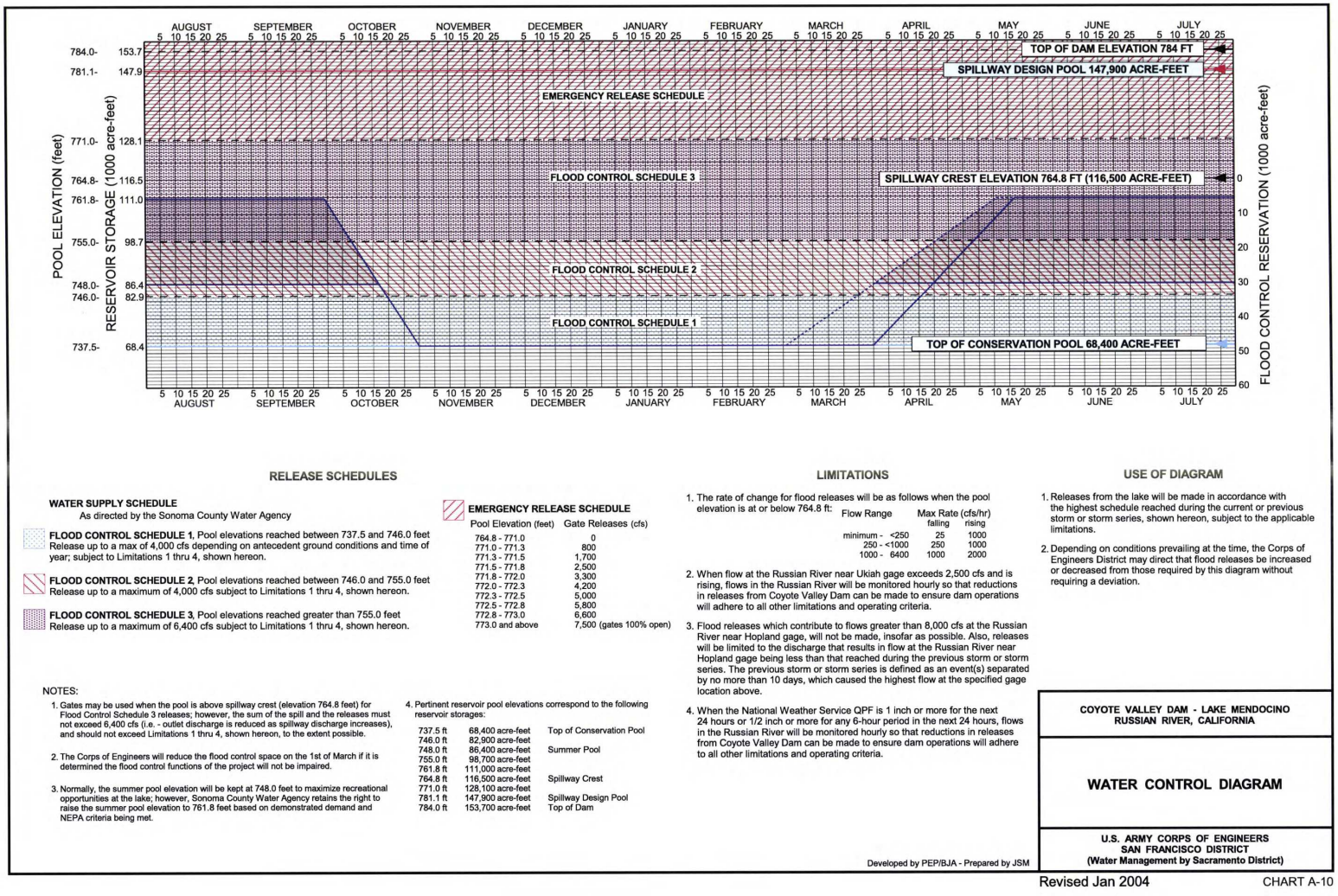

Figure 4.1: Guide curve for the operation of Coyote Valley Dam from the Water Control Manual. The diagram defines conservation and flood control zones with a seasonal variation in the limit of the conservation storage. Above the conservation storage are zones with operations defined by flood control schedules which limit releases when storage encroaches into the flood control zone (USACE, 1986).

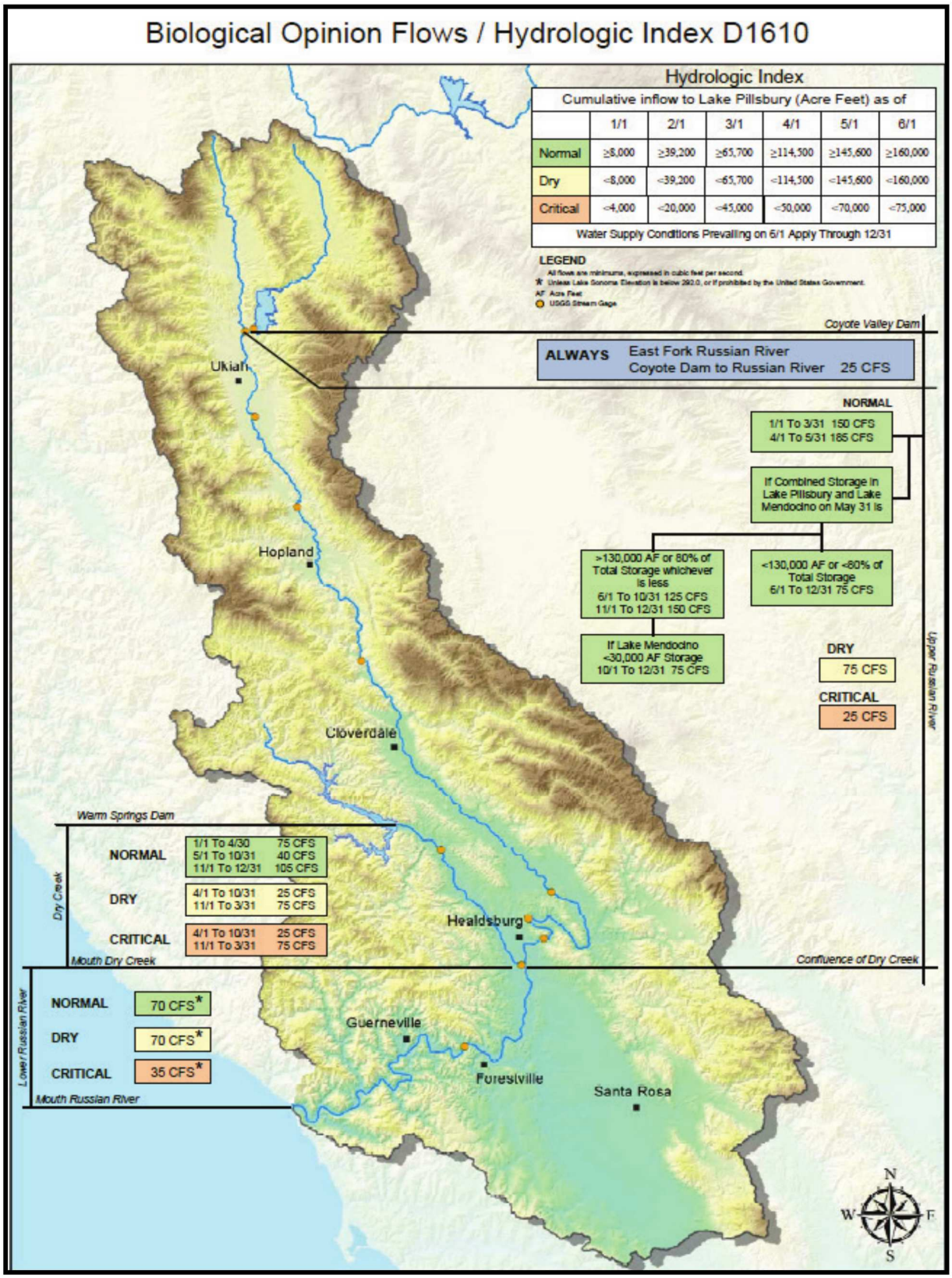

Figure 4.2: Map of the Russian River basin with the definition of the minimum flows as set by the Biological Opinion and the Hydrologic Index from Decision 1610 (Fleming, 2012).

Federal Energy Regulatory Commission (FERC). This license was modified in 2004 to enhance streamflow conditions for the three threatened salmon species in the river which has resulted in a 56% reduction in the annual diversion into the basin and a 46% reduction in the total annual inflow into Lake Mendocino (Delaney & Mendoza, 2016).

As a result of the reduction of flow diverted into the basin due to the change in operations, the SCWA has been required to petition to the SWRCB for temporary reductions to the minimum instream flow requirements in four years since 2006. As a condition of the petition in 2013, the SCWA was required to perform a water availability study. This report was released in 2015 and found that water supply is expected to decline in the future under both historical and projected changing climate conditions and that without the PVP diversion "Lake Mendocino would go dry for some period during a majority of years (over 60 percent)" (Sonoma County Water Agency, 2015).

4.4 Scenarios

The minimum environmental flows are set according to the Biological Opinion and Hydrologic Index. SCWA is currently in the process of requesting a modification to the minimum instream flows as well as the hydrologic index. The current hydrologic index is partly based on the storage in Lake Pillsbury and SCWA contends that this is not a useful indicator for the hydrologic condition in the basin (Delaney & Mendoza, 2016). Additionally, as of January of this year PG&E announced that it will abandon the Potter Valley Project (Kovner, 2019), leaving much uncertainty about the future of the basin. Because of this uncertainty, the environmental flow is considered here as a constant value throughout the year with a value of 75 cfs. Also, because of the uncertainty in the inflow, three scenarios are considered at levels of 1, 0.5, and 0 times the historical inflow contribution from the Potter Valley Project.

The water control manual allows for a maximum release of 6,400 cfs when the storage is in the flood control schedule 3 zone. The FIRO Preliminary Viability Assessment (FIRO Steering Committee, 2017) states that recent analysis has determined that a maximum release of 10,000 af over a two day period is needed to avoid exceeding target downstream flow rates. The present analysis then considers each scenario with two values for the maximum release, 5000 and 12,694 af/day. Also, the water control manual does define operations for storage above the spillway, though

this is an undesirable situation and operators wish to avoid uncontrolled releases. Accordingly, this research considers the maximum storage level to be at the spillway elevation of 765 ft (116,500 af).

4.5 Network Model

The focus of this study reflects the interest of the FIRO project on the operation of Lake Mendocino with consideration of the multiple objectives of flood control, water supply and environmental flows. The system is modeled as a network shown in Figure 4.3. The network has inflow at 10 of the nodes, and losses at 5 nodes, with some of the nodes having both gains and losses. The network includes stream routing for the reaches of the network where routing is present in the HEC-ResSim (USACE, 2007) model. The HEC-ResSim model uses a modified-puls routing method with the storage-discharge relationships contained a table associated with each stream segment. These relationships are shown in Tables 4.1-4.6. The modified-puls method is described in Section 3.3.

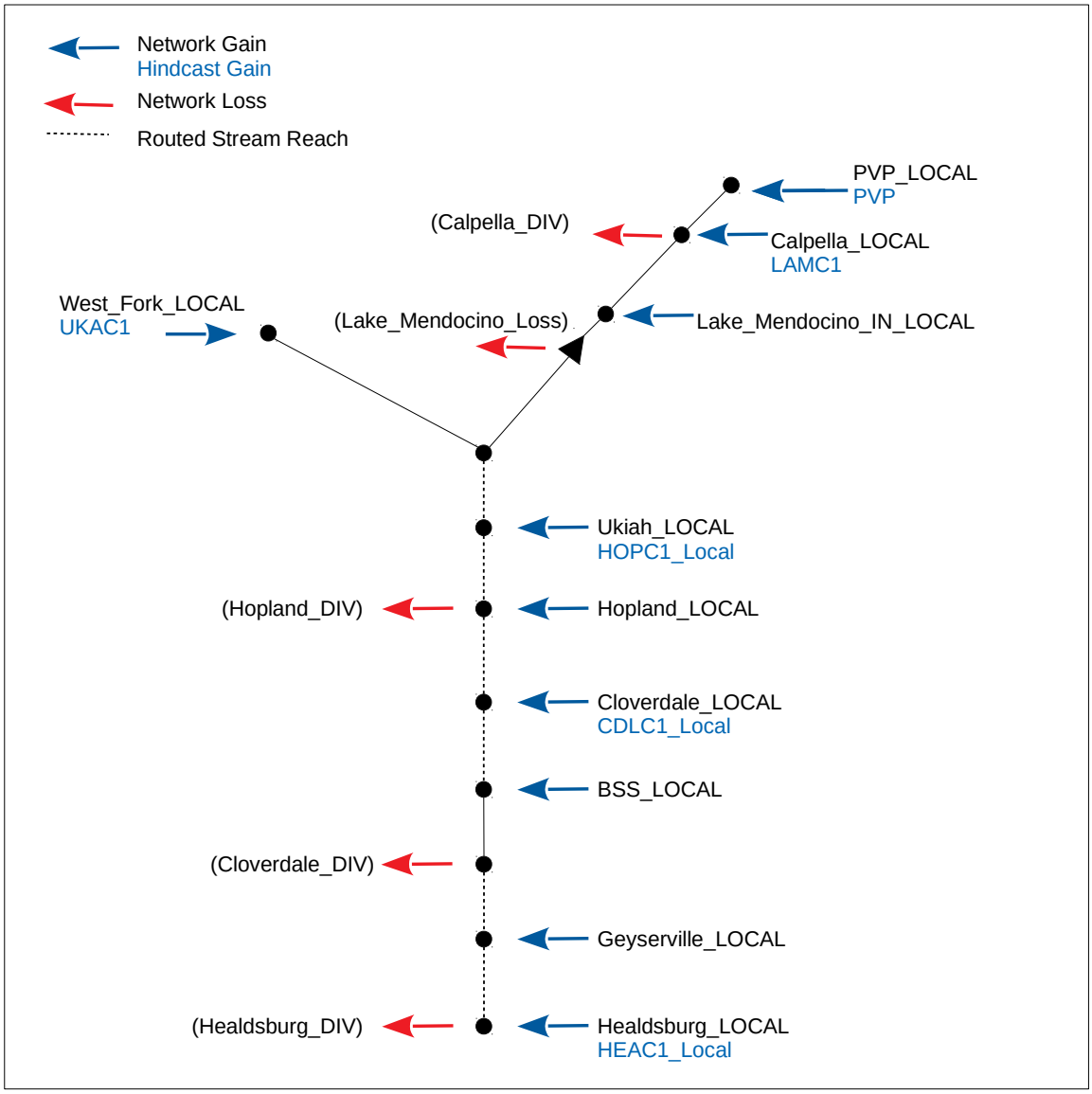

Figure 4.3: Network model of the Upper Russian River Basin. This network is based on the HEC-ResSim model (Fleming, 2012).

Table 4.1: Storage-discharge relationship for the East-West Junction to Ukiah stream segment.

Storage (af)	Discharge (cfs)
0	0
39	25
140	200
248	500
387	1,000
1,268	5,000
1,834	7,500
2,420	10,000
3,996	15,000
6,411	25,000
9,538	40,000
16,233	80,000
32,211	200,000

Table 4.2: Storage-discharge relationship for the Ukiah to Hopland stream segment.

Storage (af)	Discharge (cfs)
0	0
37	25
139	200
280	500
459	1,000
1,587	5,000
2,349	10,000
3,069	15,000
4,399	25,000
6,665	40,000
9,760	60,000
17,669	120,000
38,784	250,000

Table 4.3: Storage-discharge relationship for the Hopland to Cloverdale stream segment.

Storage (af)	Discharge (cfs)
0	0
70	25
243	200
428	500
669	1,000
2,010	5,000
3,736	10,000
5,695	15,000
9,741	25,000
14,342	40,000
19,988	60,000
38,259	120,000
82,893	250,000

Table 4.4: Storage-discharge relationship for the Cloverdale to Big Sulphur Springs (BSS) stream segment.

Storage (af)	Discharge (cfs)
0	0
24	25
98	250
235	1,000
682	5,000
1,129	10,000
1,560	15,000
2,238	25,000
3,245	40,000
4,430	60,000
5,576	80,000
10,700	160,000
22,728	350,000

Table 4.5: Storage-discharge relationship for the Cloverdale Diversion to Geyserville stream segment.

Storage (af)	Discharge (cfs)
0	0
50	25
205	250
496	1,000
1,616	5,000
3,660	10,000
5,273	15,000
8,261	25,000
11,599	40,000
15,822	60,000
19,803	80,000
34,134	160,000
62,007	350,000

Table 4.6: Storage-discharge relationship for the Geyserville to Healdsburg stream segment.

Storage (af)	Discharge (cfs)
0	0
116	25
451	1,000
1,121	5,000
3,465	10,000
6,656	15,000
10,066	25,000
15,728	40,000
23,067	60,000
32,464	80,000
41,418	100,000
82,145	200,000
195,966	350,000

4.6 Reservoir Physical Properties

The data for the physical properties of the reservoir were obtained from the HEC-ResSim model of the basin. To facilitate calculation in the code, curves were fit to the elevation-storage and elevation-area relationships (Figure 4.4). The same was done for the elevation-discharge relationship (Figure 4.5)

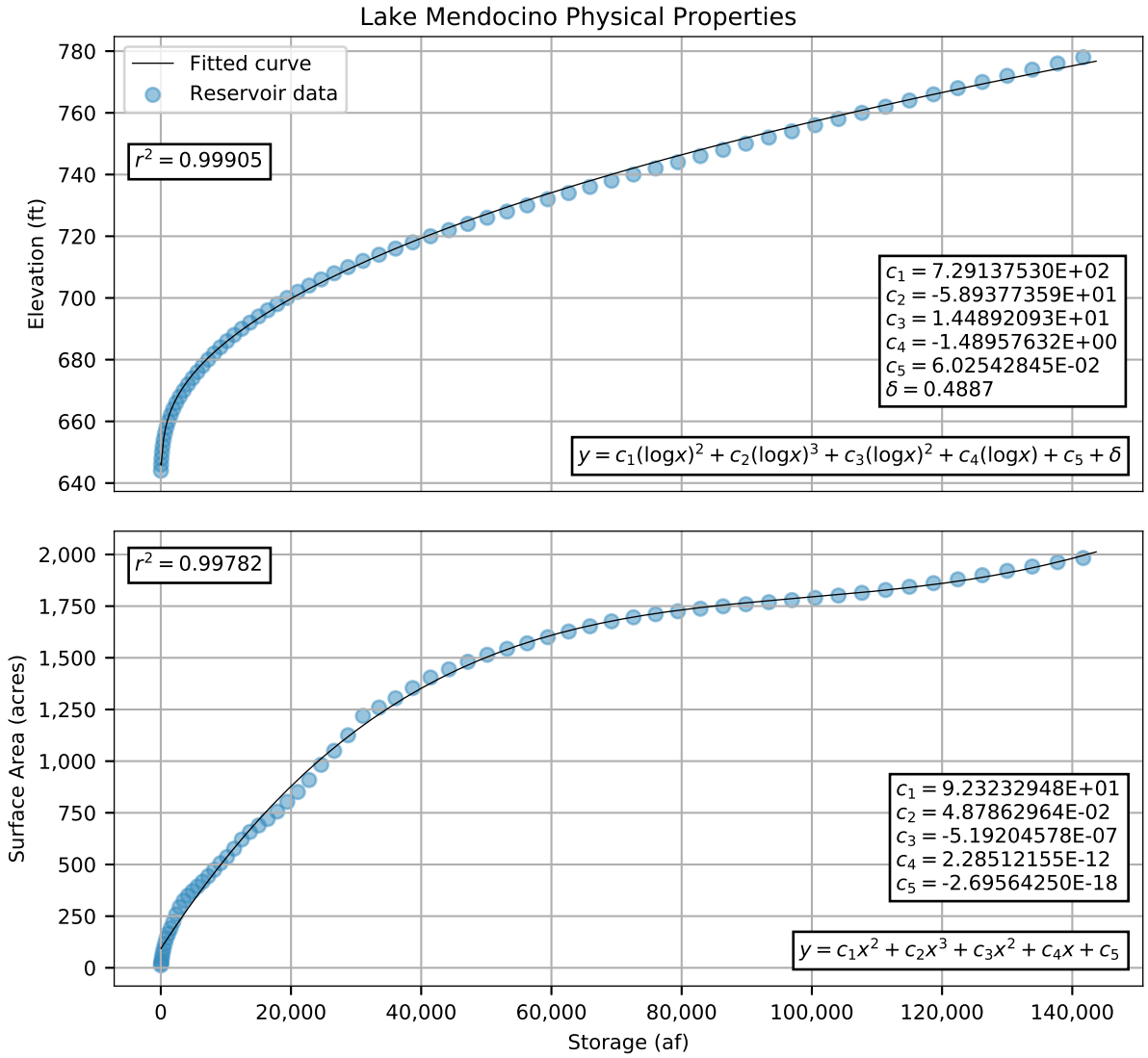

Figure 4.4: Data for the physical properties were obtained from the HEC-ResSim model of the Russian River Basin. Curves were fit to the data and these functions are used in the network model.

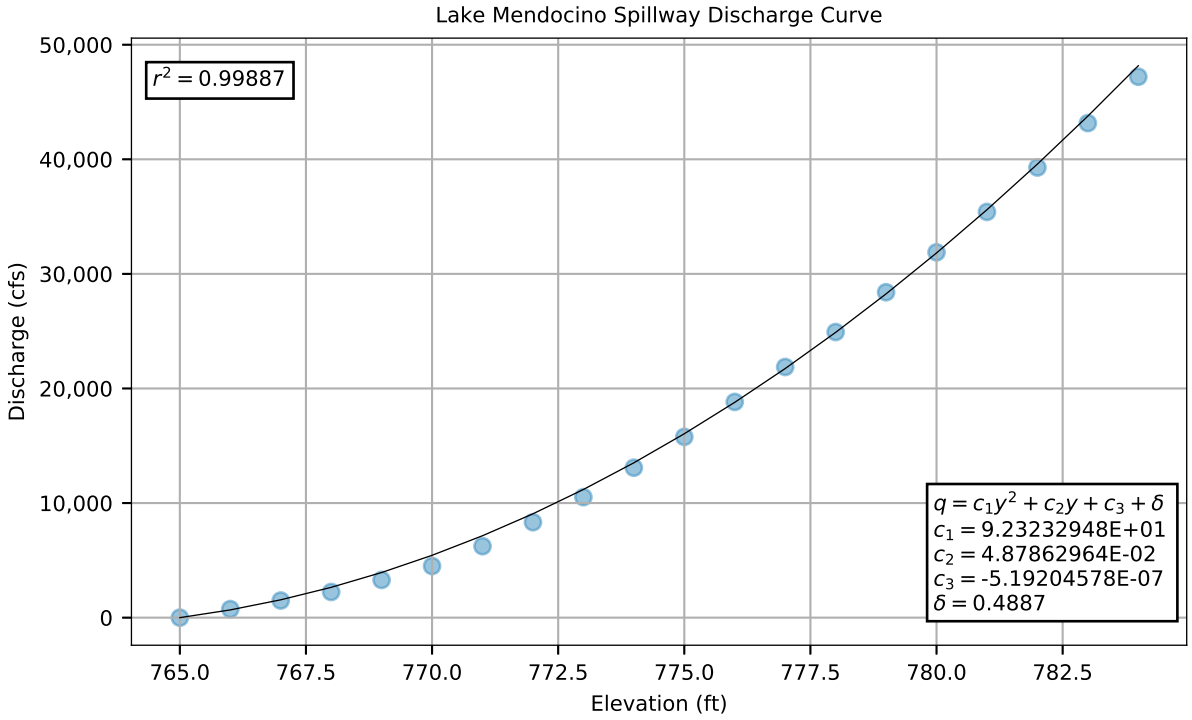

Figure 4.5: Data for the physical properties were obtained from the HEC-ResSim model of the Russian River Basin. Curves were fit to the data and these functions are used in the network model.

4.7 Input Data

The input data for this analysis comes from the current HEC-ResSim model of the basin (Fleming, 2012). This model contains data in hourly increments for the inflows to the river reaches from January 1, 1950 to December 30, 2010. These data represent the unimpaired flows developed by the USGS. The withdrawals from the river are also available at hourly increments from October 1, 1909 to December 31, 2013. The withdrawals are total losses from the stream segment including agricultural and municipal diversions, evapotranspiration and aquifer recharge.

The Potter Valley Project is a transbasin diversion into the Russian River basin from the Eel River Basin operated by Pacific Gas and Electric (PG&E) for hydropower production. The operating license for this project was modified in 2004, and since 2006, diversions have deviated significantly from the long-term average so that the historical data is not representative of expected future flows. The Eel River Model was developed as part of the license renewal to simulate future operations (USACE, 1986). The data for flows from the Potter Valley Project for the historical time period of

October 01, 1909 to October 12, 2014 uses values produced by the Eel River Model as opposed to the actual historical diversions.

Time series data are stored in HEC-DSS database files and can be managed through the HEC-DSSVue interface. Multiple sets of database files are available which seem to represent different scenarios and data generation methods. For this analysis, inflows to the network nodes were obtained from the "Flow_HMS.dss" file. The values originally provided in this data set were one-hour average flows in cubic feet per second (cfs). The time increment was changed to one-day using the "Math Functions" tool. The values for losses at each node were obtained from the "Hourly.dss" file. Average daily flows in cfs were available in this data set when it was obtained and no resampling was required. The inflow data were available from Dec 31, 1949 to Jan 01, 2011. The loss data were available from Dec 31, 1908 to Dec 31, 2014. Because the available hindcasts begin on January 1, 1985, the period January 1, 1951 to December 31, 1984 was selected for the training data.

Tables 4.7 and 4.8 show the average daily values for each of the nodes where gains are input into the system and where losses are taken out of the system. These statistics show the distinctive wet and dry seasons that are present in the basin with unimpaired gains to the system approach zero during the summer months. The diversions show peak values during the summer irrigation season.

Table 4.7: Average daily inflow at each network node by month (acre-feet/day) for the period 1950 - 2010.

	PVP	Cal	LM	WF	Uk	Hop	Clov	BSS	Geys	Heal
Jan	250.5	915.3	117.4	1114.6	738.3	639.9	1523.9	1007.9	680.6	1317.8
Feb	263.7	807.1	113.6	1060.9	780.1	695.8	1573.7	1113.9	740.3	1460.8
Mar	137.9	442.7	75.8	728.6	540.8	454.3	1012.5	769.6	502.4	982.4
Apr	134.4	126.1	32.5	323.7	223.5	189.1	439.1	350.3	218.7	426.9
May	214.9	13.6	10.4	107.0	65.6	51.5	109.6	104.1	56.5	112.9
Jun	242.1	0.9	2.4	24.7	13.2	9.6	21.4	22.7	11.5	25.0
Jul	242.3	0.0	0.3	3.7	1.9	1.5	3.4	3.5	1.8	3.8
Aug	247.8	0.1	0.1	0.7	0.6	0.2	0.5	0.6	0.5	0.7
Sep	233.6	0.4	0.9	0.8	1.4	0.1	0.2	0.2	0.6	0.4
Oct	211.0	12.2	6.7	16.3	9.9	0.7	19.0	12.8	7.0	7.7
Nov	199.1	123.6	26.7	166.9	50.6	29.4	176.1	106.0	53.7	83.1
Dec	229.9	722.4	90.2	821.0	431.9	384.1	1028.3	630.2	397.5	728.4

Table 4.8: Average daily diversion at each network node by month (acre-feet/day) for the period 1950 - 2010

	Cal	LM	Hop	Clov	Heal
Jan	0.0	2.0	9.5	0.6	6.8
Feb	0.0	1.9	9.2	0.5	6.6
Mar	2.5	2.9	12.2	1.3	7.5
Apr	2.1	7.6	16.0	2.3	9.1
May	19.2	4.5	31.9	5.5	12.1
Jun	51.6	7.5	52.3	15.8	19.1
Jul	67.6	10.7	79.8	16.0	38.6
Aug	65.9	11.1	83.5	20.7	49.1
Sep	52.5	8.4	72.2	20.4	67.2
Oct	24.8	5.8	50.7	14.0	34.4
Nov	0.0	2.1	11.9	1.0	8.6
Dec	0.0	1.4	9.7	0.6	6.9

Chapter 5

Results and Analysis

5.1 Hindcasts

5.1.1 Overview

This section will introduce and describe the ensemble forecasts for the inflow to Lake Mendocino and their performance against several common metrics. Analysis of the ensemble forecasts at the other network nodes is contained in Appendix A. Each ensemble forecast originally consisted of 61 members representing meteorological forcings from the years 1949 to 2010. The hindcasts were created for the time period 1985-2010 thus this is the time period which is used to test the forecast based operations strategy. Because the scenarios are developed using the historical observed forcings, the scenarios representing the forcings from a year concurrent with the simulation time step cannot be used as part of a forecast that could have been generated for that time step. That is, the ensemble forecast for January 1, 1985 cannot utilize the scenario generated from forcing observed in 1985. The forcings from years following the simulation year could perhaps be included even though the information was not available at the time of the forecast, as the random variables show little correlation from year to year and could then be assumed independent. The simulation could also be allowed to include scenarios that are generated from years in the designated testing period if the simulation has past that year (i.e. the scenario generated by the historical forcing from 1985 could be used on January 1, 1986), but for this application all of the scenarios for the years 1985-2010 are eliminated from the ensemble for all time steps in the simulation. This results in an ensemble of 35 members and it is the performance of this reduced ensemble that is evaluated in the following sections.

The ensemble forecasts were generated at daily intervals beginning on January 1, 1985. This date is given the counter reference $t = 1$ and the forecasts continue through September 30, 2010 which then has the reference $t = 9398$ (leap days have been removed). An ensemble forecast is represented with the notation q_t, with an individual scenario within the forecast designated with the additional superscript, $q_t^{[j]}$, where $j \in [1, M]$, $M = 35$. Many of the performance measures are only able to represent a scenario as a single value rather than a sequence of values and the cumulative

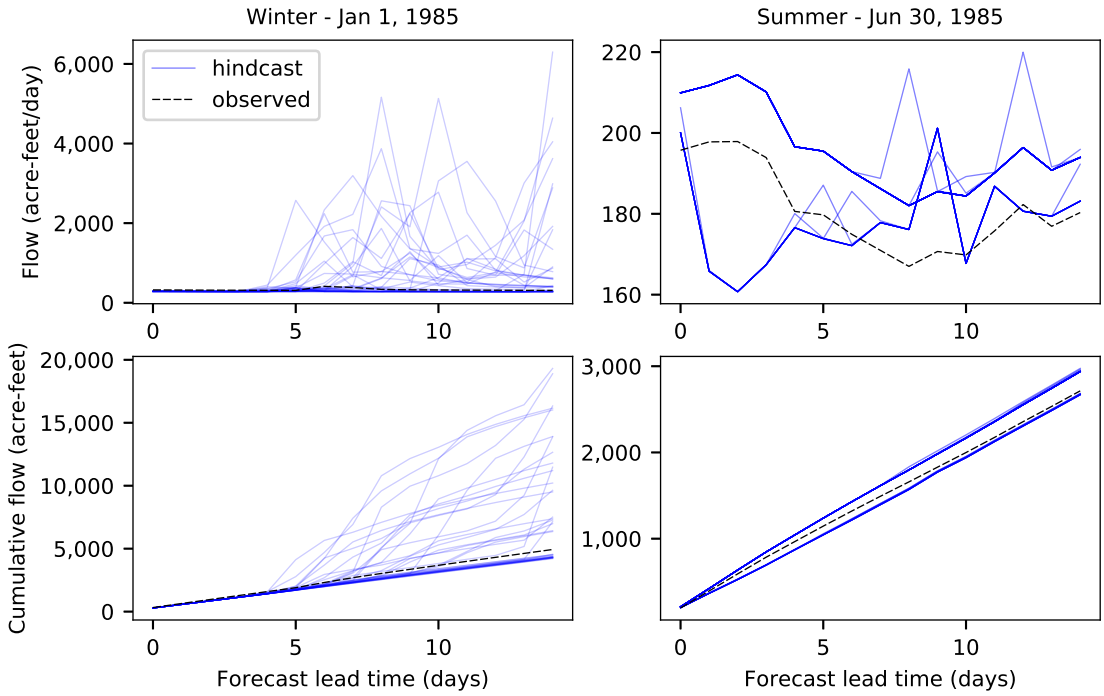

Figure 5.1: Ensemble forecasts for inflow to Lake Mendocino for Jan 1, 1985 and Jun 30, 1985. Figures (a) and (b) show the instantaneous flow, figures (c) and (d) show the cumulative flow.

flow is used his case unless otherwise stated. At times, the mean or median of the scenarios is used to represent the ensemble, and is designated as $q_{fc,t}$. When the scenario is considered as a sequence, an additional counter is needed and this is included as a subscript so that the forecasted flow at time step τ in the forecast horizon is $q_{t,\tau}$. The observation that is the target of the forecast is referred to as $q_{obs,t}$

Two ensemble forecasts are shown in 5.1. One is the forecast for Jan 1, 1985 and the other is the forecast for June 30, 1985. These show typical properties of the ensemble scenarios in each of the wet and dry seasons. The winter forecast shows all scenarios originating from a common point and diverging into unique streamflow sequences. The summer ensemble does not have as much differentiation, with the scenarios grouping near a few common sequences which do not originate from a common point, though this offset may be due to the aggregation of the hourly flows into a daily flow volume.

There are multiple performance measures that have been developed to quantify the performance of an ensemble forecast. Some measures, such as the Brier Score (Brier, 1950), measure the

performance of forecasts of relative to binary events such as exceeding a flow level threshold. Other measures are able to evaluate non-binary outcomes, though in these cases a scenario is only considered as a single-valued outcome, and not as a sequence. The nature of the current application is such that both the quantity and timing of flow throughout the forecast horizon are relevant to the decision selection, but it is currently unknown if such methods exist which are able to evaluate the performance of an ensemble with members which are sequences. Because of this, the following performance analysis will not include any measures which are relative to binary outcomes, and will represent a sequence of flows as the cumulative volume of flow over the sequence as the single value representing a scenario. The effect of the loss of timing information is not known and is perhaps a useful area for future research. In any case, this simplification should be taken into account when considering the values returned by the performance measures.

Throughout this discussion, the data used to represent the observation which is the target of the forecast are the time series of flows obtained from the HEC-ResSim model and used in the simulation. Several metrics compare only a single value representing the entire ensemble forecast. In these cases, the values of the mean and the median of the cumulative flows are used. Some of the metrics compare the forecasts to a reference (or "naive", that is, a forecast that is considered less sophisticated) forecast and often the mean or median of the historical observed values is used. Given the findings of (Pappenberger et al., 2015) the mean or median might be too easy to outperform, so where possible, an additional reference using the value for the previous time step is used and referred to as "persistence." Additionally, because the dry and wet seasons are so greatly different, the forecasts are grouped into *Summer* (the dry season, May to September) and *Winter* (October to April) categories and the measures are calculated separately.

5.1.2 Nash-Sutcliffe Efficiency

The Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) measures the improvement of the forecast compared to the mean according to

$$NSE = 1 - \frac{\sum_{t=1}^{N}(q_{obs,t} - q_{fc,t})^2}{\sum_{t=1}^{N}(q_{obs,t} - \bar{q})^2}, \tag{5.1}$$

where for a set of N observation and forecast pairs, q_{obs} is the observed value, q_{fc} is the forecasted value, and $\bar{q}$ is the mean of the set of observed values. The NSE has a range of $[-\infty, 1]$ with 1 corresponding to perfect forecasts. A positive value of the NSE value indicates that the forecast is better than using the mean, and a negative value indicates that the forecast performs worse than the mean.

The NSE for the winter forecasts is positive for all forecast lead times (Figure 5.2). The NSE increases up to six days lead time, then decreases consistently through the end of the forecast horizon. Decreasing NSE with extended lead time is to be expected as uncertainty increases for time periods further in the future. The reason for the increase through the first six days lead time is uncertain. It is perhaps a result of the nature of the data set which is dominated by values near baseflow with infrequent high flow values increasing the overall seasonal average. Forecasts in the near term then underpredict until the ensemble achieves enough diversity among the scenarios to increase the ensemble mean to near the seasonal mean. The NSE behaves similarly using either the mean or median of the ensemble as the forecast value, with the mean providing higher values. The NSE value for the summer forecasts begins positive at shorter lead times, then decreases as the forecast horizon is extended, becoming negative at three days lead time. The decrease is to be expected as there is more uncertainty in later flows. Again, negative values indicate performance worse than using the seasonal mean and this is perhaps the result of the lower diversity in the ensemble members for the summer forecasts.

For a little more insight, the reference value of the mean, $\bar{q}$, can be replaced by the previous time period's flow for the 1-day forecast horizon, for the 2-day forecast, the sum of the previous two time periods is used and so on, for each day in the forecast horizon:

$$NSE = 1 - \frac{\sum_{t=1}^{N}(q_{obs,t} - q_{fc,t})^2}{\sum_{t=1}^{N}(q_{obs,t} - \sum_{\tau=1}^{FH} q_{t-\tau})^2}. \tag{5.2}$$

This is referred to as the "persistence" reference. With this as a reference, the summer forecasts have higher and consistently positive values indicating both that the forecasts provide more information than the persistence value, and that during the summer the mean is a better (harder to beat) reference. The values in the middle range, in this case, are lower during the middle periods of the

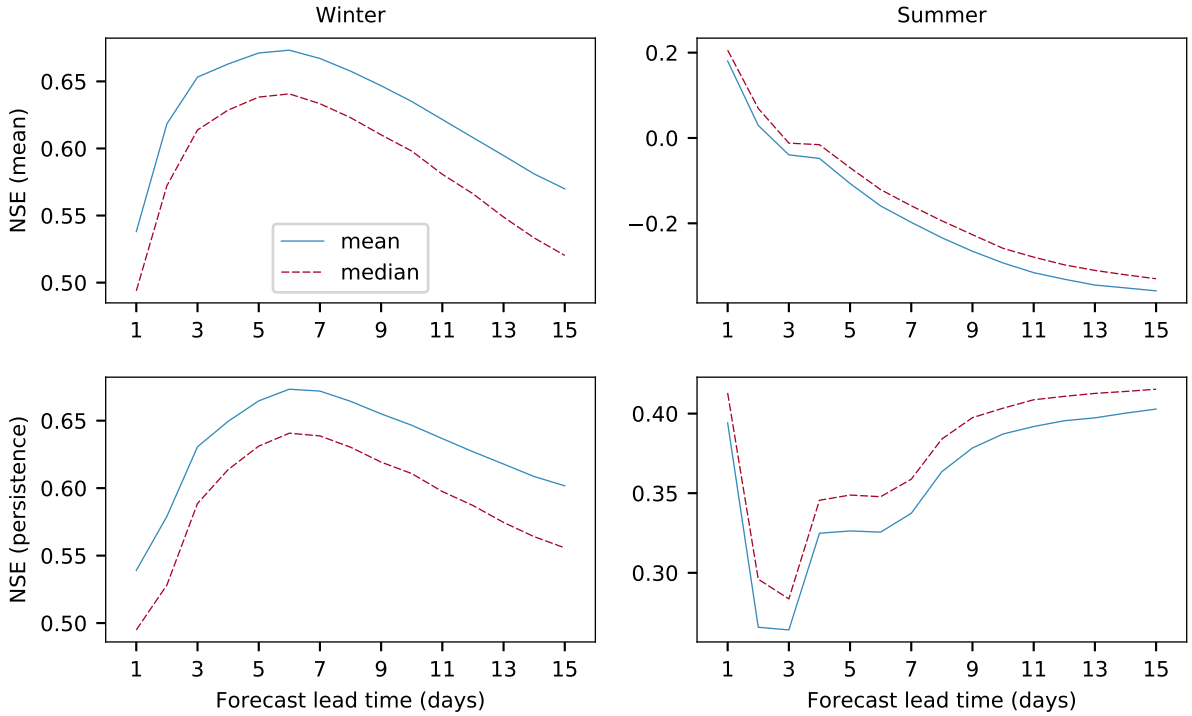

Figure 5.2: NSE values for the ensemble forecasts using the mean or median as representative of the ensemble. The forecasted value is the cumulative flow over the forecast horizon. Figures (a) and (b) use the mean of the seasonal values as the reference (eq. (5.1)), figures (c) and (d) use the persistence value (eq. (5.2))

forecast horizon. During the winter, the NSE is very similar to the values obtained when using the mean flow as a reference.

5.1.3 Percent Bias

The tendency for forecasts to over or under predict can be measured by the percent bias:

$$PB = \frac{\frac{1}{N}\sum_{t=1}^{N}[q_{obs,t} - q_{fc,t}]}{\bar{q}}, \tag{5.3}$$

with negative values indicating a trend toward overprediction and positive indicating underprediction. Figure 5.3 shows the percent bias for the ensemble forecasts for each season using either the mean or median as representative of the ensemble. The winter forecasts show a positive bias throughout the forecast horizon, though the mean of the ensemble does not increase the magnitude of the bias at the same rate as the median. The summer forecasts show negative bias for all lead times during

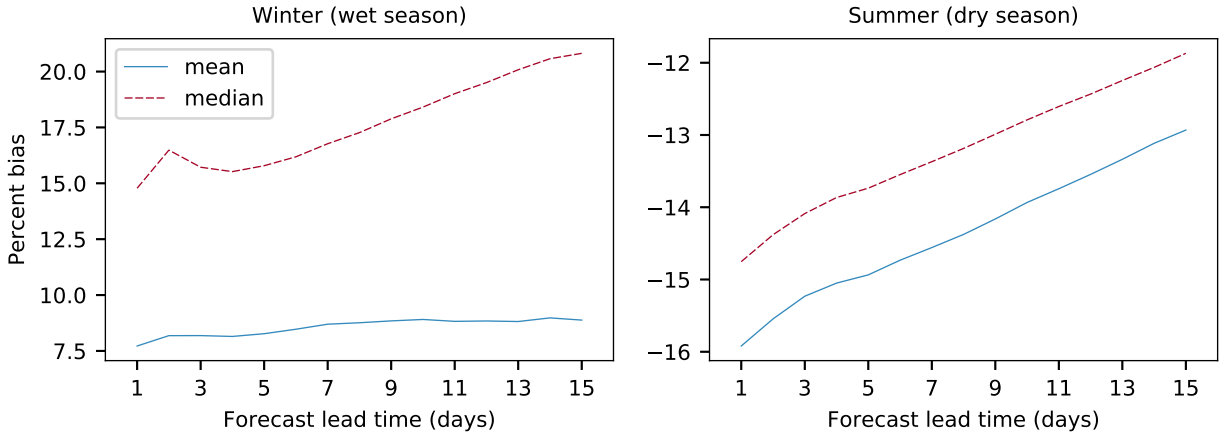

Figure 5.3: Percent bias values for the ensemble forecasts using the mean or median cumulative flow as representative of the ensemble, with the forecast grouped by season. The forecasts during the summer are negative for all lead times which indicates that the ensemble mean or median leads to overprediction. Forecasts during the winter show positive bias (underprediction). The winter forecasts show an increase in the magnitude of the bias with increasing lead time, with the mean showing less increase. The opposite is true for the summer forecasts where the magnitude of the bias decreases over the forecast horizon.

the forecast horizon. The cause of the decrease in the magnitude of the bias with increasing lead time is uncertain but may be a result of regression toward the mean.

5.1.4 Continuous Ranked Probability Skill Score

The continuous ranked probability score (CRPS) compares the entire ensemble of forecast values to the observed outcome, though each scenario, which consists of a sequence of flow values, must still be represented by a single value. The continuous ranked probability score for an ensemble represented by a probability density function, $\rho(x)$ and an observed value x_a is calculated by

$$\text{CRPS} = \text{CRPS}(P, x_a) = \int_{-\infty}^{\infty} [P(x) - P_a(x)]^2 dx. \tag{5.4}$$

where

$$P(x) = \int_{-\infty}^{x} \rho(y) dy, \tag{5.5}$$

and

$$P_a(x) = H(x - x_a), \tag{5.6}$$

and where H is the Heaviside function:

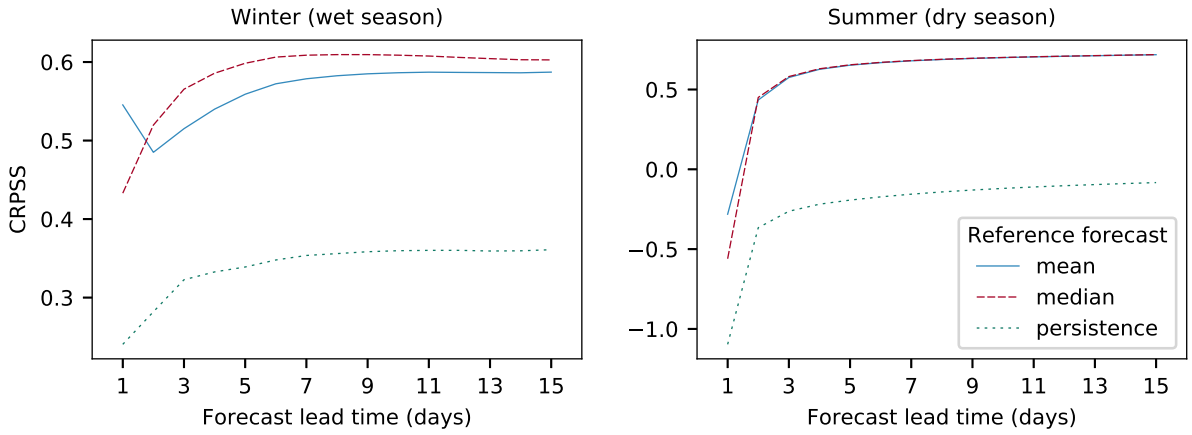

Figure 5.4: Continuous ranked probability skill score for the ensemble forecasts 1985 - 2010 grouped by season with three different values used as the reference forecast. Compared to the mean or median of the historical observed values, the skill score is similar and in both cases results in a positive score for all forecast lead times except for the one-day lead time during the summer. Compared to using the persistence as the reference forecast, the ensembles score lower.

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0. \end{cases} \tag{5.7}$$

The skill score (CRPSS) is a comparison of the score of an ensemble compared to the score using a reference forecast:

$$CRPSS = \frac{\overline{CRPS_{ref}} - \overline{CRPS_{fc}}}{\overline{CRPS_{ref}}}. \tag{5.8}$$

Similar to the NSE, the CRPSS ranges from $[-\infty, 1]$ with positive values indicating that the forecast outperformed the reference. Figure 5.4 shows the CRPSS using the seasonal mean or median as the reference as well as the score compared to using the previous flow as the reference forecast ("persistence"). The forecasts show positive values for all lead times in the winter and all lead times after one-day in the summer when using either the mean or median reference. The forecasts score lower compared to the persistence reference forecast with the summer forecasts receiving a negative score for all lead times.

5.1.5 Coefficient of Variation of the RMSE

The magnitude of the errors can be examined by calculating the root mean squared error:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[q_{obs,t} - q_{fc,t}]^2}, \tag{5.9}$$

which can be scaled by the mean of the observations to obtain the dimensionless coefficient of variation:

$$\text{CV} = \frac{\text{RMSE}}{\overline{q}_{obs}}. \tag{5.10}$$

Additionally, the ensemble as a whole can be used by considering the error for each scenario according to

$$\text{RMSE} = \sqrt{\frac{1}{NM}\sum_{t=1}^{N}\sum_{j=1}^{M}[q_{obs,t} - q_{fc,t}^{[j]}]^2}, \tag{5.11}$$

where M is the total number of scenarios and $q_{obs}^{[j]}$ is the forecast for the scenario with index j. Figure 5.5 shows these measures for the forecast set grouped by season. The RMSE in both seasons increases over the forecast horizon with the magnitude of the error much greater in the winter. The CV scales the RMSE by the mean which provides a measure of the error relative to the mean. A CV value of 0 would indicate perfect forecasts, and 1 indicates that average errors are similar in magnitude to the mean. The CV of the winter forecasts is greater than 1 in the short-term and converges to around 1 as the lead time increases. The CV for the summer forecasts also decreases with increased lead time with all values less than 1.

5.1.6 Rank Histogram

One method of analysis that does not require a reference forecast is the rank histogram. A sorted list consisting of each of the ensemble scenarios and the corresponding observation is created and the rank of the observation is recorded. The collection of the ranks is then presented as a histogram with each bin showing the frequency that the observation occupies a position. Figures 5.6 to 5.9 show the rank histograms for this collection. Because of the large variation in the counts at each position, the histograms are shown in both linear and log scales. In all forecasts, at all lead times, the first position is by far the most frequent position meaning that the observation is most likely less than the smallest value in the ensemble. The second most frequent rank is the highest position (36). Together these two positions make up the large majority of rankings of the observations meaning that the observation rarely falls within the range of the scenario values.

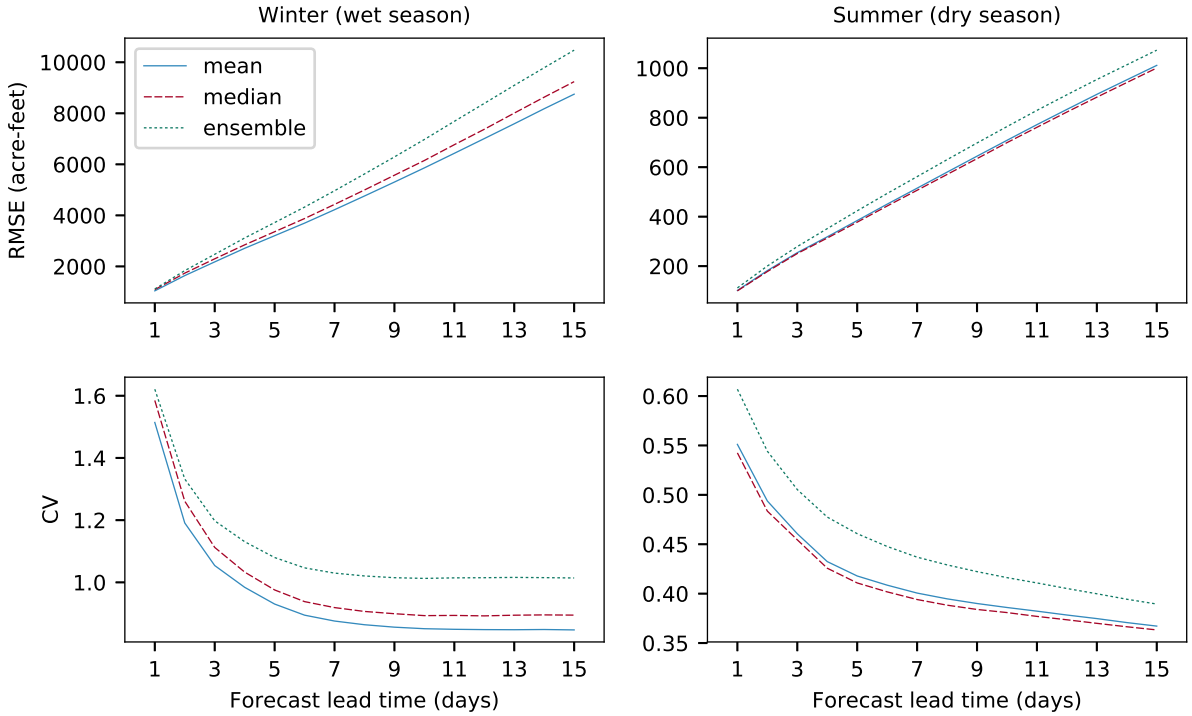

Figure 5.5: The RMSE and CV for each lead time in the forecast horizon grouped by season. The RMSE increases with extended lead time in both seasons with the magnitude much larger in the winter. The CV scales the RMSE by the underlying mean flow over the domain to produce a dimensionless value with values perfect forecasts having a value of zero.

5.1.7 Distance From Ensemble

In the present application, the ensembles are used without assigning a probability distribution to the members. The minimum release is determined for each scenario and the action taken is then the maximum of the set of minimum releases. Because of this, two properties of the forecasts that are of interest are the ability to span the observation and the characteristics of the deviations outside of the scenario range. This is a follow up to the rank histogram which gave the position of the observation in a sorted list. That analysis showed that many of the observations did not fall within the range of the scenarios. This analysis looks at the percentages of forecasts in which the observation lies above or below the range of scenario values and the characteristics of those deviations.

The forecasts were compared to the observed values and recorded if the observation was not within the range

$$\min_j\{q_t^{[j]}\} - \epsilon \leq q_{obs,t} \leq \max_j\{q_t^{[j]}\} + \epsilon, \tag{5.12}$$

where $\epsilon = 10$ af for the winter forecasts and $\epsilon = 1$ af for the summer forecasts. The ensembles with longer lead times more often contain the observation, however, only achieving this approximately half of the time. At shorter lead times, the observation more frequently lies outside of the range of values with over 70% of the observations falling outside of the ensemble for the one-day lead time. The performance is constant over the forecast horizon after two days for forecasts during the summer with over 80% of the forecasts failing to contain the observation at all lead times. In all cases, overprediction is more common, which was observed in the rank histograms.

Figures 5.11 and 5.12 provide more detail on the nature of the deviations of the observations from the ensembles. The red line indicates the mean of the deviations, the box represents the 25-75 percentiles and the whiskers show the maximum and minimum observed deviations for each lead time in the forecast horizon. The magnitude of the mean deviation above and below the scenario range for the winter forecasts is not large relative to the average flows, approximately 100 - 200 af, and a relatively small range contains the 25-75 percentiles. The factor that is of great interest, particularly with relation to the flood control criterion, is the magnitude of the maximum and minimum deviation. This is approximately 2000 af for the one-day lead time when overprediction occurs but is nearly 20,000 af when underprediction occurs. It is these infrequent events that are of concern when operating the reservoir for flood control.

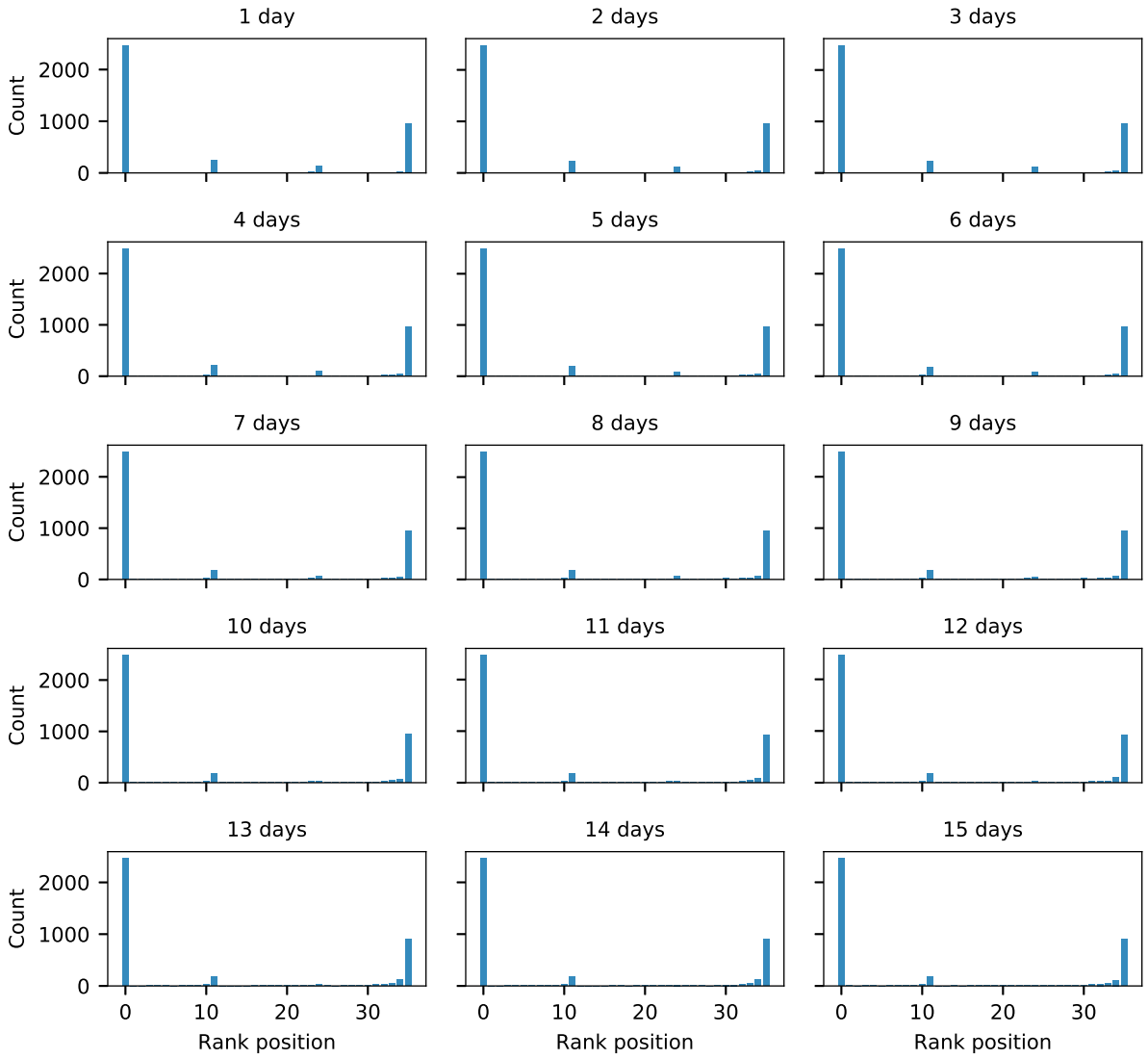

Figure 5.6: Rank histogram for each lead time in the available forecast horizon for the ensemble forecasts during the summer dry season. All lead times show the most frequent position of the observed flow to be at the first position indicating that the observed flow is lower than all of the forecasted values. The second most common position is at the other end of the list (position 36) indicating that the observed value was higher than all of the forecasted values. These two positions occur with an overwhelming majority.

Figure 5.7: Rank histogram for each lead time in the available forecast horizon for the ensemble forecasts during the summer dry season shown with a logarithmic scale. With this representation it can be seen that the observed value does fall between the maximum and minimum forecasted values with non-zero frequency. A regular patter of peaks appears consistently throughout the lead times at positions 12 and 25, though the cause of this is unknown.

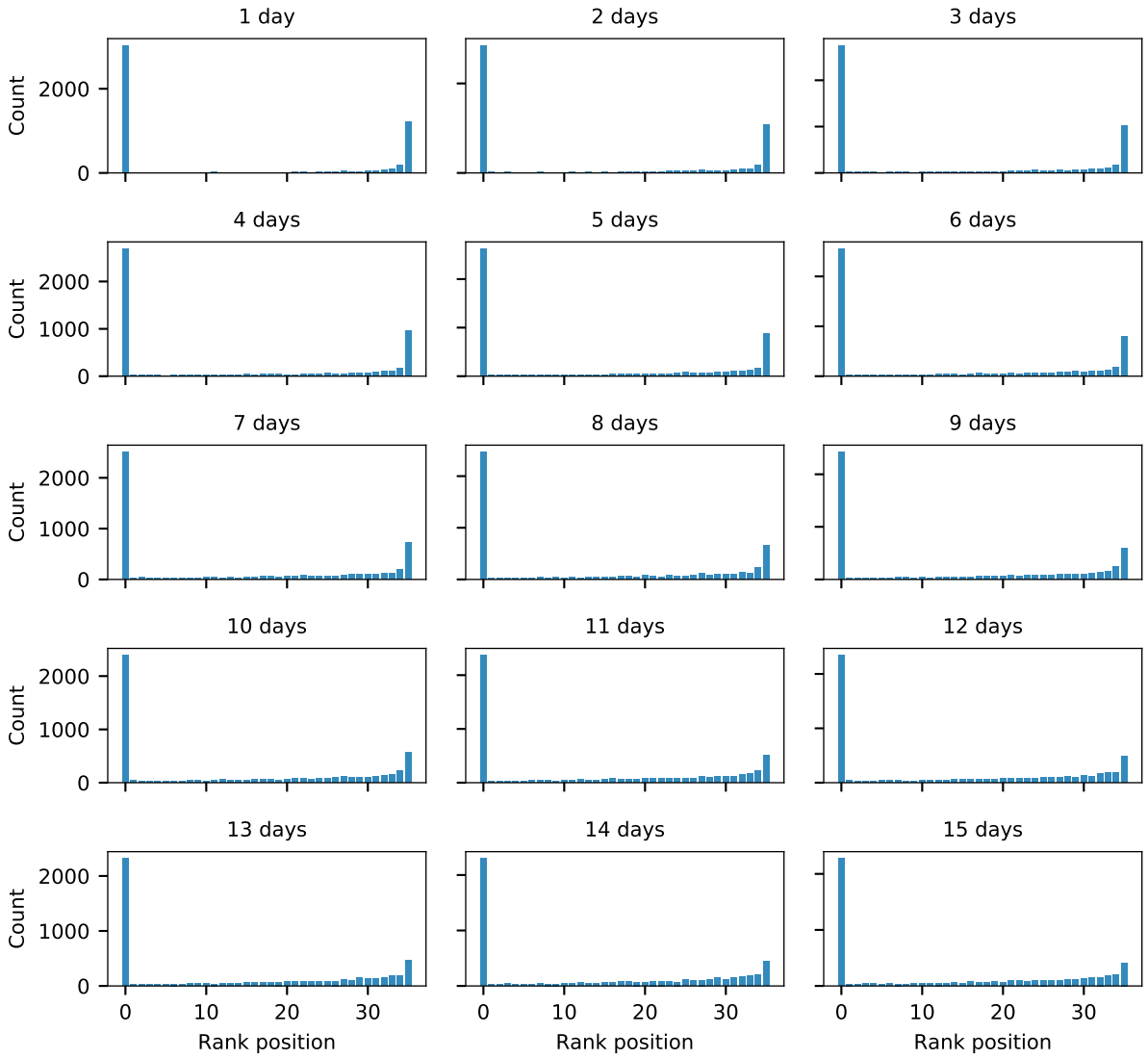

Figure 5.8: Rank histogram for each lead time in the available forecast horizon for the ensemble forecasts during the winter wet season. Similar to the summer forecasts, positions 1 and 36 are the first and second most frequent positions. The observed value falls with the range of the ensemble of forecasted values more frequently than the summer forecasts, with a generally increasing trend toward the higher positions.

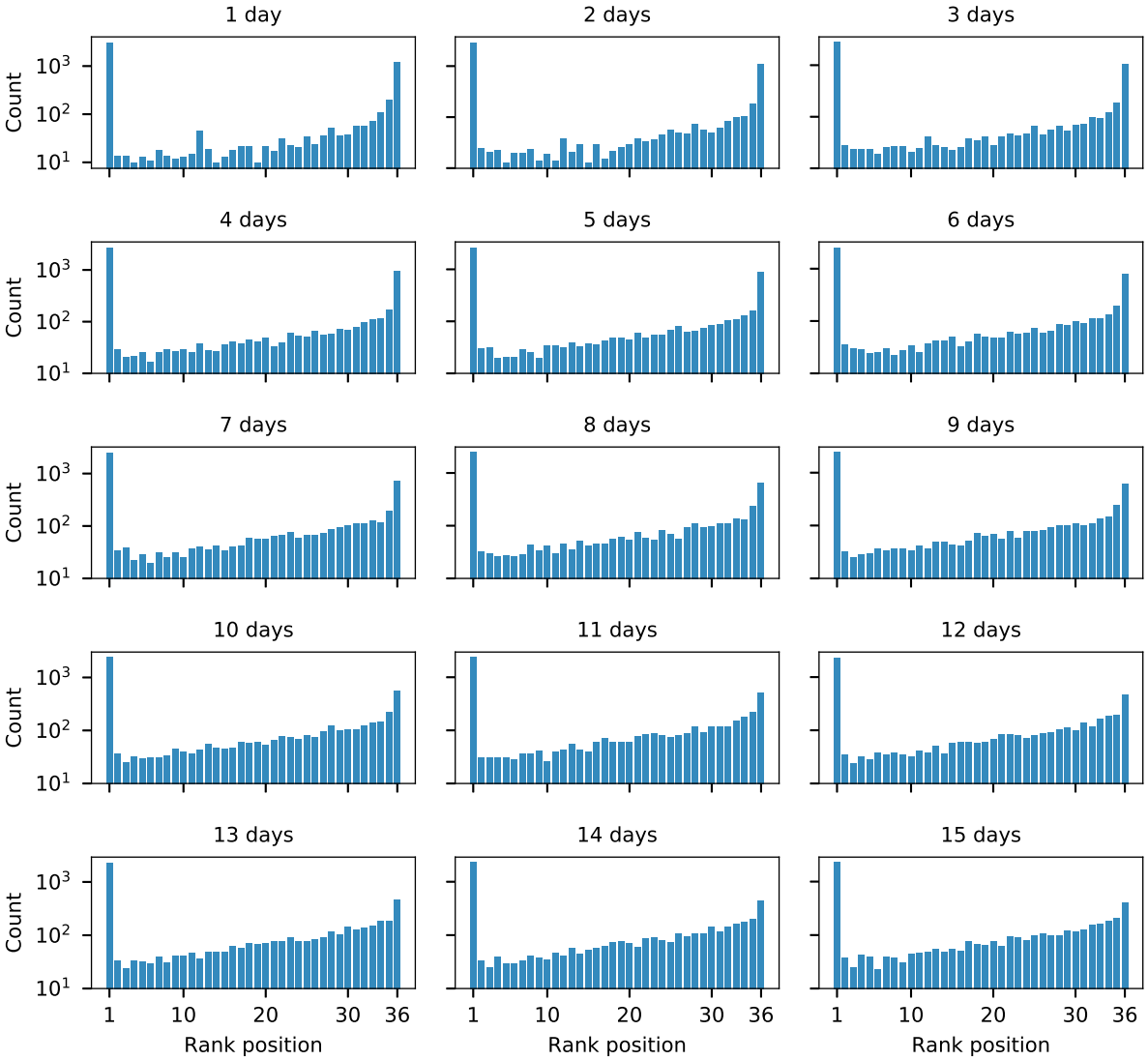

Figure 5.9: Rank histogram for each lead time in the available forecast horizon for the ensemble forecasts during the winter season shown with a logarithmic scale. With this representation, the trend of increase in frequency with increase in position after position 1 can be seen.

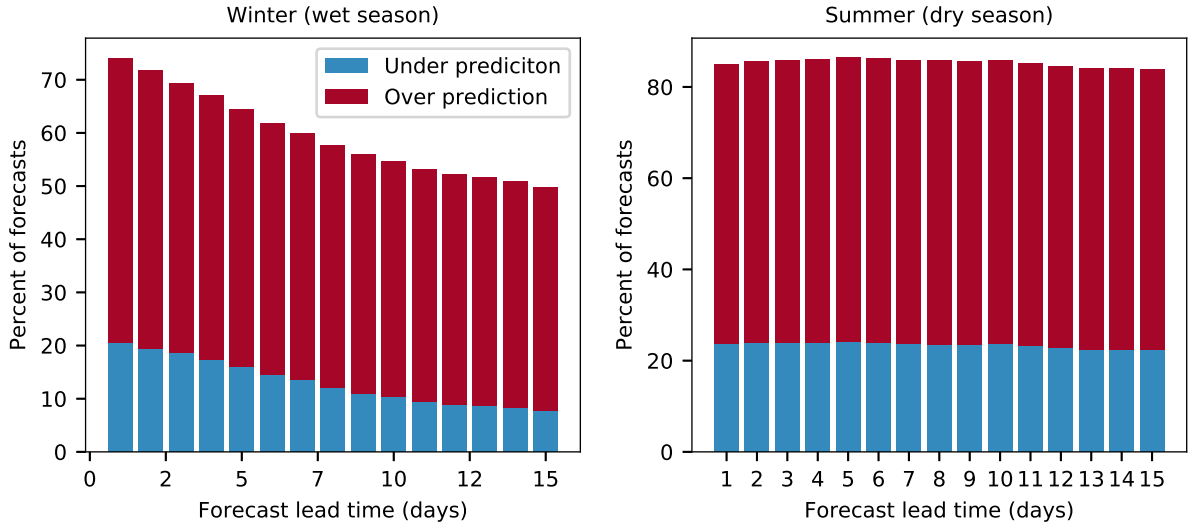

Figure 5.10: Percent of the ensembles in each season in which the maximum is less than the observation (underprediction) or the minimum is greater than the observation (overprediction) considering a tolerance of $\pm\epsilon = 10$af. The

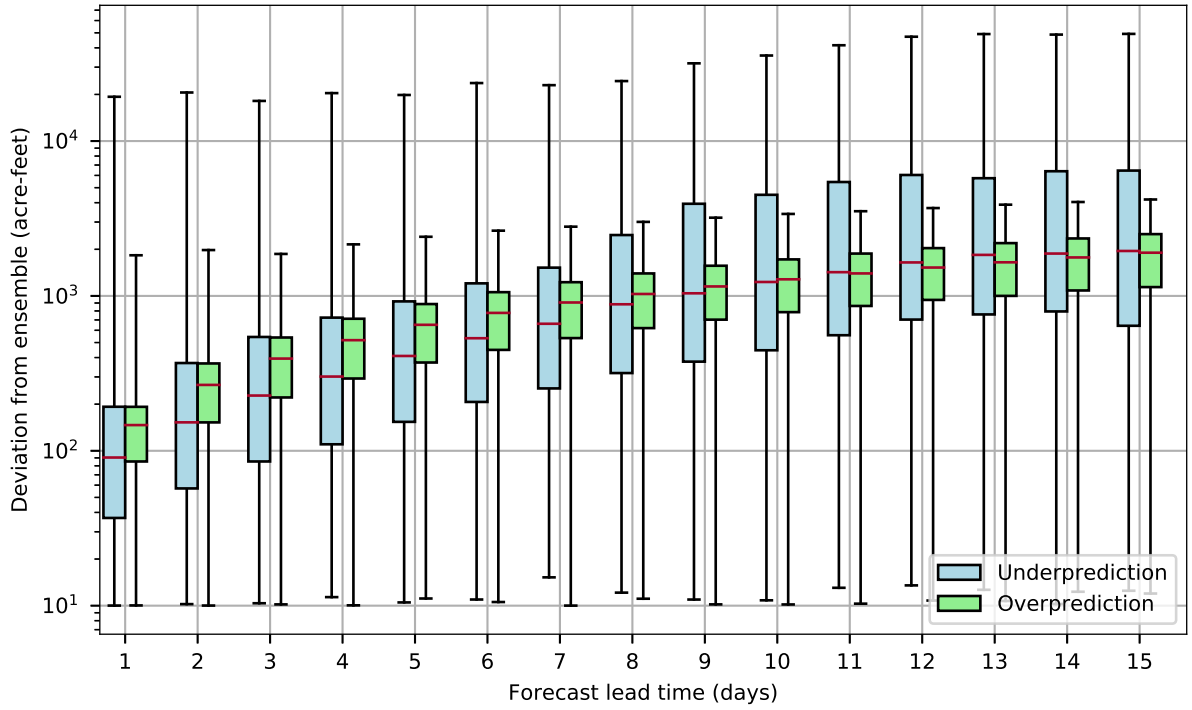

Figure 5.11: Diagram of the deviations from the ensemble range when the observation is either greater than the largest value in the ensemble (underpredicton) or less than the lowest value (overprediction), normalized by the mean of the observations for each lead time, and considering a tolerance of $\pm\epsilon = 10$af, for forecasts during the winter period. The red line indicates the mean of the deviations, the box represents the 25-75 percentiles, and the whiskers are the maximum and minimum values present. The deviations are only included if they are not within $\epsilon = 10$ af of the ensemble range.

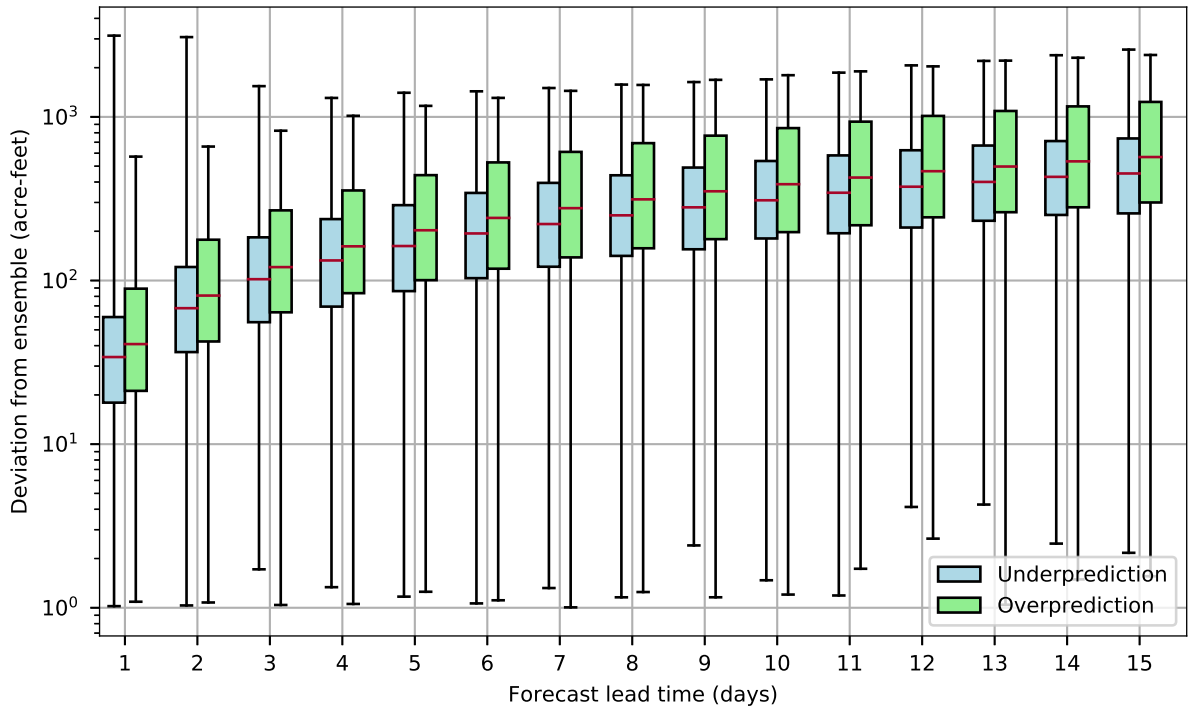

Figure 5.12: Diagram of the deviations from the ensemble range when the observation is either greater than the largest value in the ensemble (underpredicton) or less than the lowest value (overprediction), normalized by the mean of the observations for each lead time, and considering a tolerance of $\pm\epsilon = 10$af, for forecasts during the summer period. The red line indicates the mean of the deviations, the box represents the 25-75 percentiles, and the whiskers are the maximum and minimum values present. The deviations are only included if they are not within $\epsilon = 10$ af of the ensemble range.

5.2 Synthetic Data

The input data used to train the learning models are presented in Figures 5.13-5.14 with the solid line showing the mean value of the total volume of inflow or loss for each calendar day and the shaded area representing one standard deviation. The input from the PVP diversion shows a stepped pattern resulting from the managed flow of the project. The other gains in the system show the natural variation from the stochastic hydrologic conditions in the basin. The winter season shows higher mean flow with accompanying large variations. The summer months have consistently low flow with little variation. The losses from the system show low values present during the winter months with little variation. Diversions increase during the summer irrigation season and show some variability. One distinction between gains and losses being that the standard deviation is larger than the mean for the inflows on many dates. The standard deviation for the diversions does not reach this level of variability.

To study the suitability of the synthetic data, the unregulated gains and losses above Lake Mendocino are summed into one gain and collectively referred to as *LM_Inflow*. The synthetic data generator was run for 10,000 years and the results grouped by the calendar day. Figure 5.15 shows the daily values for the mean and maximum daily flow as well as the 90th, 99th and 99.9th percentiles. The mean daily values of the synthetic data show less variation between adjacent dates and overall lie within the range present in the variability of the input data with the mean of the synthetic data at times greater and at other time less than the mean of the input data. The synthetic data follows the annual wet/dry seasonal pattern again when examining quantiles with a general trend of less frequently being exceeded by the input data until comparing the maximum values in which case the synthetic data exceeds the input data at all dates.

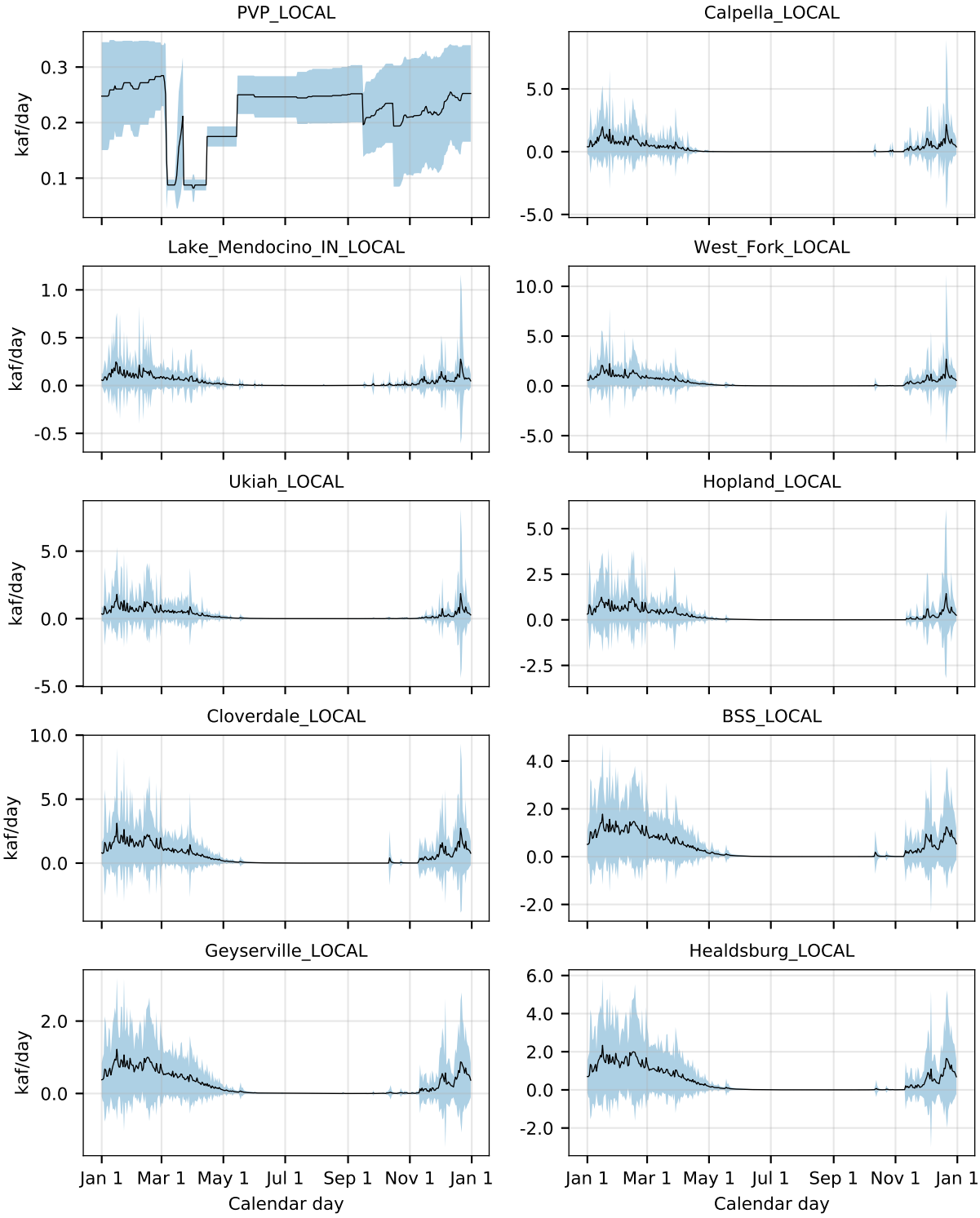

Figure 5.13: Daily statistics for the gains to the network model. The solid line indicates the mean daily flow and the shaded region represents one standard deviation. The value is the total daily volume of flow in thousand acre-feet per day (kaf/day).

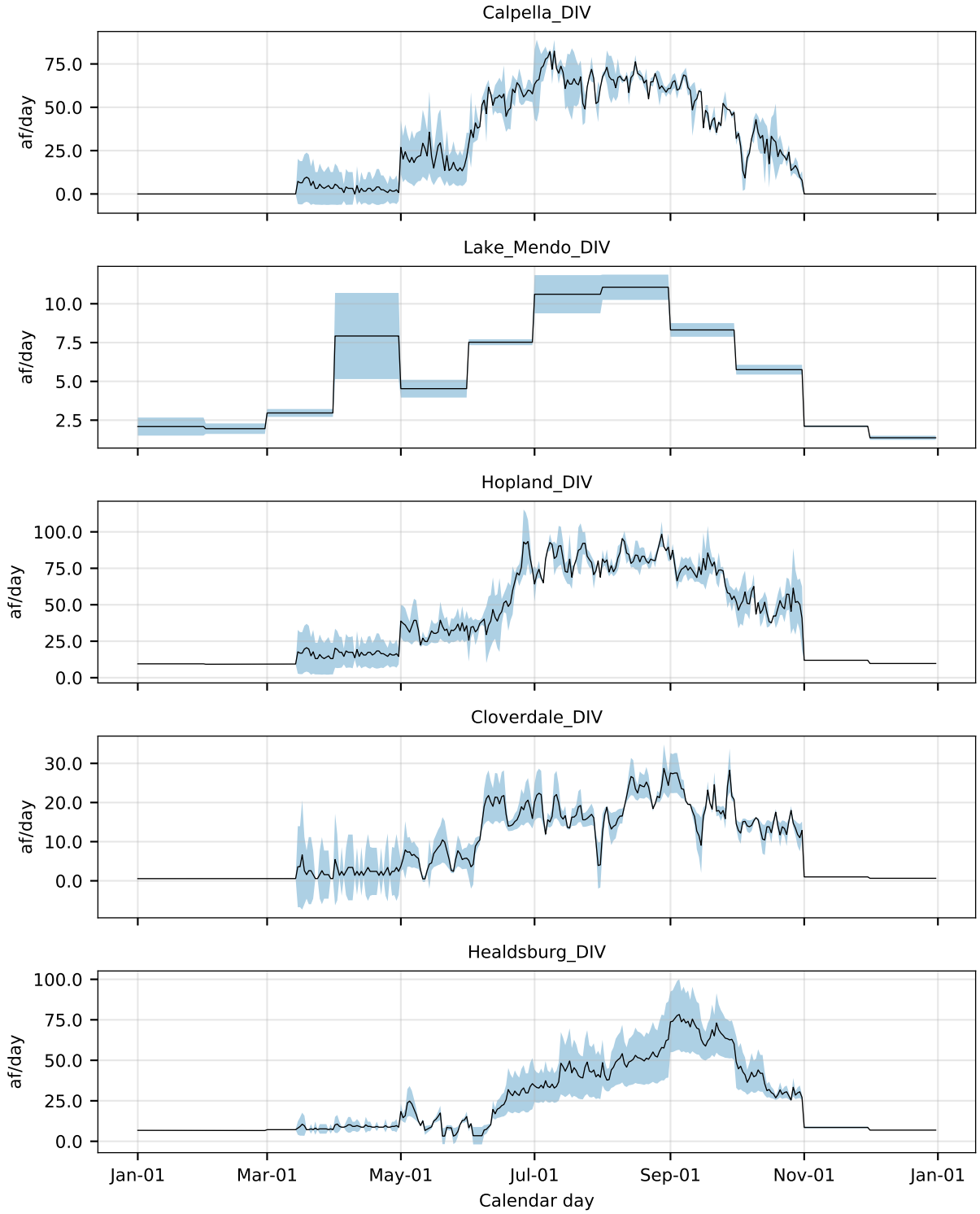

Figure 5.14: Daily statistics for the losses to the network model. The solid line indicates the mean daily flow and the shaded region represents one standard deviation. The value is the total daily volume of withdrawal in acre-feet per day (af/day).

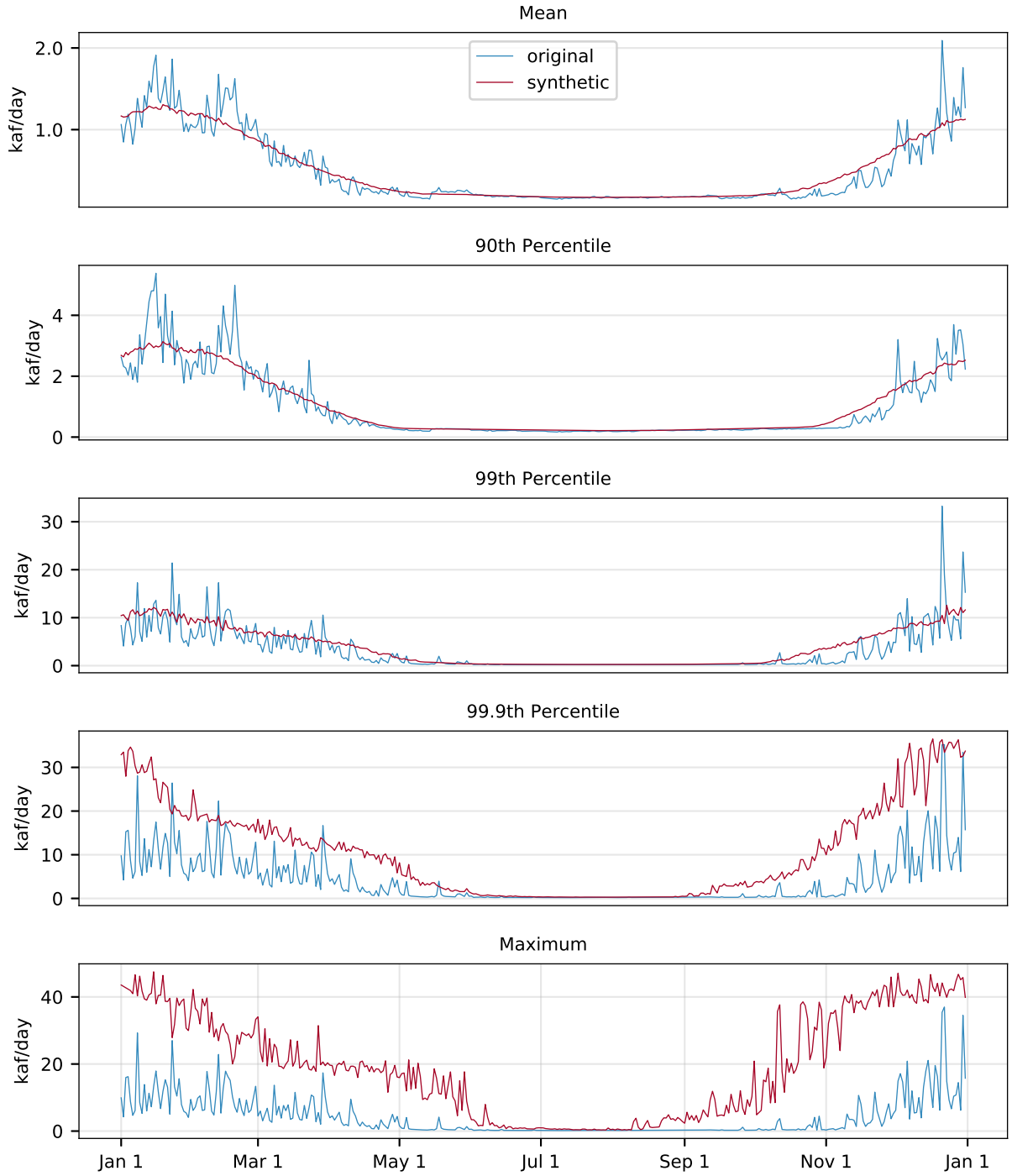

Figure 5.15: Comparison between the original data set and the synthetic data set of daily values of percentiles for inflow to Lake Mendocino. The mean of the synthetic data set falls within the variation of the values in the neighborhood of each date. As the percentile is increased, the synthetic data set more frequently exceeds the original data set. Comparing the daily maximums, the synthetic data set exceeds the original data set at all dates. This behavior aligns with expectation when the original data set is considered a subset of a larger data set.

5.3 Constraints

The entire state space which the system can occupy is given the notation $\mathcal{S}$ and partitioned into three regions: the terminal states above the maximum storage, $\mathcal{S}^+$, the terminal states below the minimum storage, $\mathcal{S}^-$ and the remaining non-terminal states between these limits, $\mathcal{S}^\circ$. The minimum storage in $\mathcal{S}$ is 0 af and the maximum is 140,000 af. The maximum operational storage (non-terminal) in the case study is the level at the spillway, 116,500 af. The minimum operational storage was selected to be 10,000 af. These values were selected to allow for a small region of positive storage values to be terminal states below the minimum operational storage, and for a region of storage that is terminal above the maximum operational storage. It was thought that this might help when attempting to train an ANN to take on the value that is designated as the boundary condition for this region.

In some cases, the boundary between terminal and non-terminal states is designated as a *boundary state*. A boundary state intercepts any state transition that would cross this boundary and assigns the value of the boundary storage level to the ending state. The episode is ended and the storage is randomly assigned at the next iteration. This is used to represent a condition where a storage limit is encountered but not considered a failure according to the objective under consideration. For example, when learning to operate the reservoir for the environmental objective, the storage in the reservoir may exceed the maximum storage in the reservoir. The agent does not benefit from exceeding the maximum storage as this water will exit the spillway. At the same time, this condition is not something that is detrimental to the objective so the agent is not penalized. A similar condition is present when operating the reservoir for the flood control constraint, this time at the lower storage limit of the reservoir. The agent does not benefit from making releases that would go below this limit and storage is maintained at this limit at all times. Again, this is not detrimental to the flood control objective and the agent is not penalized. The important feature of the boundary state is that the value of the $V(s)$ is learned given the conditions of the boundary state. In contrast to a terminal state, where the value of the $V(s)$ function is always the limit of an infinite series of uniform rewards or penalties.

As part of the learning process, the agent is learning a policy which is constrained to an upper and lower release limit. The lower limit is designated as 50 af/day or approximately 25 cfs which is the lowest release that is allowed in all conditions. The FIRO program preliminary viability

assessment (PVA) (FIRO Steering Committee, 2017) states previous analyses determined that a limit on releases from the reservoir of 10,000 af over two days is necessary to avoid exceeding target flow rates. The upper limit is then given a value of 5,000 af/day for this analysis as opposed to the upper limit of approximately 12,700 af/day based on the water control manual. The policy returned by the agent is then a normalized value between 0 and 1 based on these limits. The Hopland Rule constrains the reservoir operations to avoid causing flow downstream at Hopland to exceed 8,000 cfs. The release from the reservoir must still meet the minimum of 50 af/day regardless of the flow downstream.

5.4 Scenarios

The water that is diverted into the Russian River basin by the PVP is a significant portion of the annual inflow. Because the future of this diversion is uncertain, three scenarios were created to study the effects of possible reduction in this flow. The "Full PVP" scenario uses the time series data for the PVP diversion obtained from the HEC-ResSim model, the "Half PVP" uses the same time series reduced by a factor of 0.5, and the "Zero PVP" scenario eliminates the inflow from the PVP project entirely. All other data is consistent between scenarios.

5.5 Value Function Approximation

5.5.1 Objectives and Reward Functions

The operational objectives can be grouped into three main categories: flood control, water supply and environmental flow. The value functions are based on the reward signal that the agent receives as a result of an action taken and the subsequent state transition. The reward signal must then provide information to the agent which correlates with the outcomes which the stakeholders either seek or wish to avoid. Rewards were developed to consider the flood control (FC) and water supply (WS) objectives independently. A reward was developed which is based on the level of flow provided to meet the minimum environmental flow objective (ENV). Water supply is implicitly considered in this scenario as water supply demands are met before any water is provided for environmental flow.

Another reward function was developed which acts as a combination of the flood control, water supply and environmental flow objectives. This function returns a value based on the level of environmental flow provided, conditioned on meeting the upper and lower storage limits (COMBO).

As with the ENV reward function, the water supply demands are met before the environmental flows. This reward is used in a simulation as a verification step of the implementation of the learning algorithm.

The problem is not well defined as objectives are stated in general terms. The FIRO Preliminary Viability Assessment (FIRO Steering Committee, 2017) describes the goal as: "to make modest incremental adjustments to the WCM flood management guidelines to improve reliability in meeting water management objectives and ability to meet environmental flow requirements without diminishing (or actually enhancing) flood protection or dam safety." It was the intent of this study to assume as little as possible beyond what is stated in the available reports. One of the benefits of this reinforcement learning method is that it is not dependent on the form of the reward functions and other functions could easily be substituted if found to align more closely with the stakeholders objectives. There is a benefit though to keeping the reward functions simple as it allows for a better conceptual understanding of the value functions that result from a selected reward function.

In the case of the environmental objective (ENV), the agent is given a reward of

$$r(s,a,s') = \begin{cases} 0 & \text{if } s' \in \mathcal{S}^- \\ \min_k \left(\frac{q_{env} - q_k)}{q_{env}} \right), \forall q_k < q_{env} & \text{if } s' \in \mathcal{S}^\circ \text{ and } \exists q_k < q_{env} \\ 1 & \text{otherwise.} \end{cases} \tag{5.13}$$

The boundary between $\mathcal{S}^\circ$ and $\mathcal{S}^+$ is designated as a *boundary state* as described in Section 3.1.7, and $\mathcal{S}^-$ is a terminal state. In all situations, the water supply demands are met before any water is used to meet the environmental flow objective. Thus, this reward is simply the ratio of the environmental flow provided, up to the full demand.

In the case of the flood control objective (FC), the reward is

$$r(s,a,s') = \begin{cases} -1 & \text{if } s' \in \mathcal{S}^+ \\ 0 & \text{otherwise,} \end{cases} \tag{5.14}$$

and thus the agent only receives a penalty if the maximum storage limit is violated. The Hopland Rule is considered part of the environment so that releases are limited to avoid causing the flow at

Hopland to exceed the maximum regardless of the decision selected by the agent. The policy in this case is then trivial and recommends the maximum release at all states, though learning the state-value function is not simple and requires the addition of the *prioritized replay* technique to be computed (see Section 3.1.10). The boundary between $\mathcal{S}^-$ and $\mathcal{S}^\circ$ is a boundary state in this case and storage is not allowed to end below this level.

For the water supply (WS) objective, the agent is given a reward of

$$r(s,a,s') = \begin{cases} 1 & \text{if } q_k > q_{ws,k} \forall q_k \wedge s' \in [\mathcal{S}^\circ \cup \mathcal{S}^-] \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

where $q_{ws,k}$ indicates the water supply demand at node k. This is then a binary reward with a zero value unless all of the water supply demands are met and the ending storage is below the maximum storage limit.

The reward function that attempts to combine all of the objectives (COMBO) is

$$r(s,a,s') = \begin{cases} 0 & \text{if } s' \in [\mathcal{S}^- \cup \mathcal{S}^+] \vee \nexists q_k > q_{env} \forall q_k \\ \min_k \left(\frac{q_{env} - q_k)}{q_{env}} \right) \forall q_k < q_{env} & \text{if } s' \in \mathcal{S}^\circ \wedge \exists q_k < q_{env} \forall q_k \\ 1 & \text{otherwise.} \end{cases} \tag{5.16}$$

This has the effect of combining the flood control and environmental flow objectives with the highest priority given to the flood control objective followed by the water supply and finally the environmental flow objective.

5.5.2 Discount Factor, γ

The discount factor, γ, is one of the most important parameters in the learning process. Depending on the objective, it can have an affect on the stability of the learning process itself as will be shown in the following experiments. It also must be selected to align with the preferences of the decision-makers and reflect a reasonable discount rate associated with future rewards. There is only a single discount factor used which implies a constant rate at which rewards are discounted for each time step. However, this value may not necessarily be the same between objectives. To

understand the effect of the discount factor, three values of this parameter were selected (0.99, 0.999, and 0.9999) and the learning process was executed for each.

5.5.3 Evaluation and Output

A collection of 50 "experience agents" simultaneously take actions according to the action policy, $\mu(s)$. At the beginning of a learning session, the experience agents are uniformly distributed in the state space, both in the storage dimension and the time dimension. The learning process was conducted for 2500 sessions of 60 time steps (1 day) each. After each time step, the experience, (S, A, R, S'), of each experience agent is recorded in the memory buffer. Then, five stochastic gradient updates are made to the parameters of both the actor and critic ANNs, for a total of 750,000 parameter updates. To track the learning process and to make some evaluation on whether the process was converging, it was decided to observe the mean value of the $V(s)$ function. To evaluate the mean, the $Q(s, a)$ function is evaluated for all values of a grid of size $[141 \times 365 \times 100]$ corresponding to the storage, calendar day, and action dimensions. The maximum over the action dimension then becomes the value of $V(s)$ for each location in the state space. The mean of these values is taken as the mean of the $V(s)$ function.

There is a noticeable time cost to evaluating the $Q(s)$ and $\mu(s, a)$ functions over the entire state and state-action spaces so this is done less frequently at the beginning of the training. The total training time including the function evaluations was approximately 15 hours on a Linux server with 2 Intel® Xeon 14 Core (2.4 GHz) processors and 256 GB RAM and up to five models running on a single server at a time. In general, the results converge, though with some level of noise always present. Because of this noise, after 1,500 sessions the functions are evaluated every 10 sessions. The arrays are stored after each evaluation and the last 100 arrays are averaged at each grid point in the state space to get the output of the learning process.

5.5.4 Combined Objectives

The ultimate goal of the RL application is an estimate of the state-value function for each objective individually. It would be advantageous to be able to directly evaluate the output of the learning model through simulation, however when learning is performed with respect to a single objective, simulating the resulting policy does not provide a means of evaluating the policy. In the

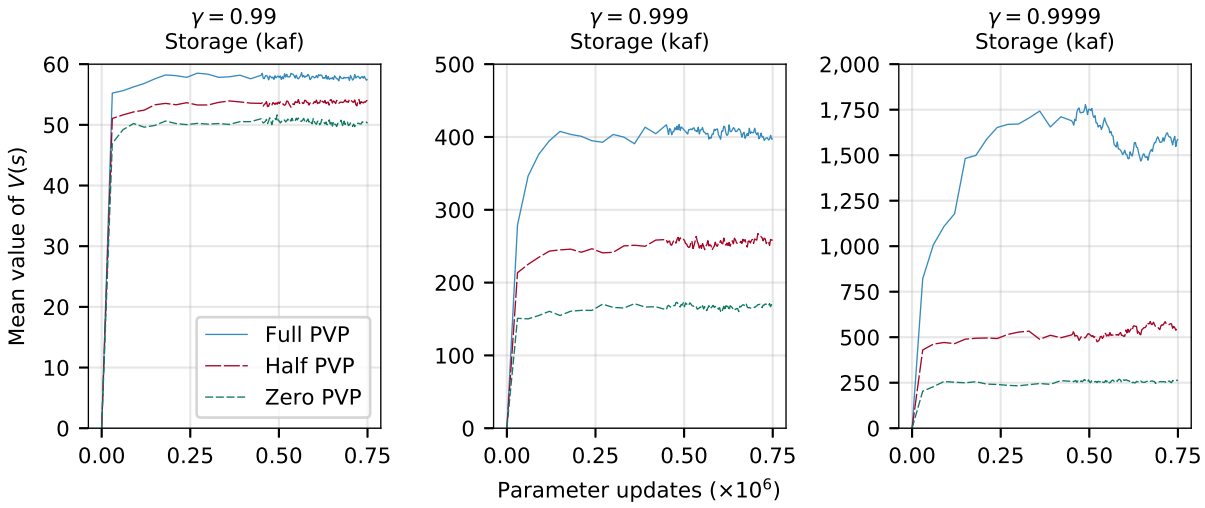

Figure 5.16: Mean value of the $V(s)$ function over the state space for three values of the discount factor,γ, for each level of flow from the PVP project, when using the COMBO reward (Equation 5.16) which is a combination of the flood control, water supply and environmental flow objectives.

case of flood control, for example, the reservoir is simply emptied which does not provide any useful information. To assess the value of the RL model, the three objectives of flood control, water supply and environmental flow were combined into one reward signal (Equation 5.16) and simulated using the embedded network model. The simulation was executed for each level of diversion from the PVP project and each value of the discount factor. This is referred to as the COMBO objective. Figure 5.16 shows the mean value of the $V(s)$ function for each run during the training process. The output of the ANN representing $V(s)$ is not normalized so the value will be affected by the discount factor, γ. For reference, the highest return that could be achieved is $\sum_{t=0}^{\infty} \gamma^t(1) = \frac{1}{1-\gamma}$. With a discount factor of $\gamma = 0.99$, the maximum is 100, and for a value of $\gamma = 0.9999$, the maximum is 10,000. The results show the "Full PVP" scenario to have a higher mean value of $V(s)$ for all values of γ, which is what would be expected as the PVP diversion supplies additional water that can be used to meet environmental flows, though at a low enough daily volume that flood control is not negatively impacted by this additional inflow. The training shows a steep increase in the mean value of $V(s)$ at the beginning of the training period. In the case of $\gamma = 0.99$, the mean value of $V(s)$ appears to converge well as the value remains within a small range of a constant value, though with some noise still present. For the case of $\gamma = 0.999$, the changes in the mean value again increase quickly before slowing, though with perhaps less certainty in the level of convergence. The

"Full PVP" scenario shows more noise in the system as the value varies more than in the other two scenarios. The rate of change is slow but the "Half PVP" scenario shows somewhat of a constant increase that continues through the end of the training period. The "No PVP" scenario appears to converge to a constant value again with all scenarios showing some level of noise. For a discount factor of $\gamma = 0.9999$, the mean value of $V(s)$ converges well for the "Half PVP" and "Zero PVP" scenarios but shows much more variability for the "Full PVP" scenario.

State-Value Function, $V(s)$

Figure 5.17 shows a heatmap plot of the output $V(s)$ function for each scenario. Darker colors indicate higher values, and the different hues are used to group the scenarios by the discount factor and to avoid direct comparison of the magnitudes of the value functions when different values of the discount factor are used. Again, these represent the average of 100 samples of the $V(s)$ function, taken over the last 1000 sessions of the training period (equivalent to the last 300,000 parameter updates). In general, the values decrease as the contribution from the PVP project is decreased, which agrees with the results from Figure 5.16.

The effect of varying the discount factor can be seen by following contours from the middle of the year during irrigation season to the end of the year when the wet season has begun. When $\gamma = 0.99$, all contours up to the maximum storage during the middle of the year will descend to nearly the lowest storage level at the start of the wet season. When $\gamma = 0.9999$ the contours do not descend to the minimum storage at the start of the wet season. This suggests that when the value of $\gamma = 0.9999$ is used, carryover storage into the wet season is more important. The lower level of the discount factor is more short-sighted and does not give as much value to carryover storage. The higher value of γ places more value on future rewards and is able to consider storage between years.

For reference, consider the contribution of a reward that is received one year in the future. When $\gamma = 0.99$, the discounted value of this future reward is $0.99^{365}(1) = 0.025518$, for $\gamma = 0.999$ the discounted reward is 0.69407 and for $\gamma = 0.9999$ it is 0.96416. This means that a reward received one year into the future is given a weight of approximately 96% of an immediate reward when $\gamma = 0.9999$ but only 2.5% when $\gamma = 0.99$. The benefits from rewards this far into the future become insignificant when $\gamma = 0.99$. Thus carryover storage has less value, and the agent will try to use most of the available storage to meet the needs of the current irrigation season.

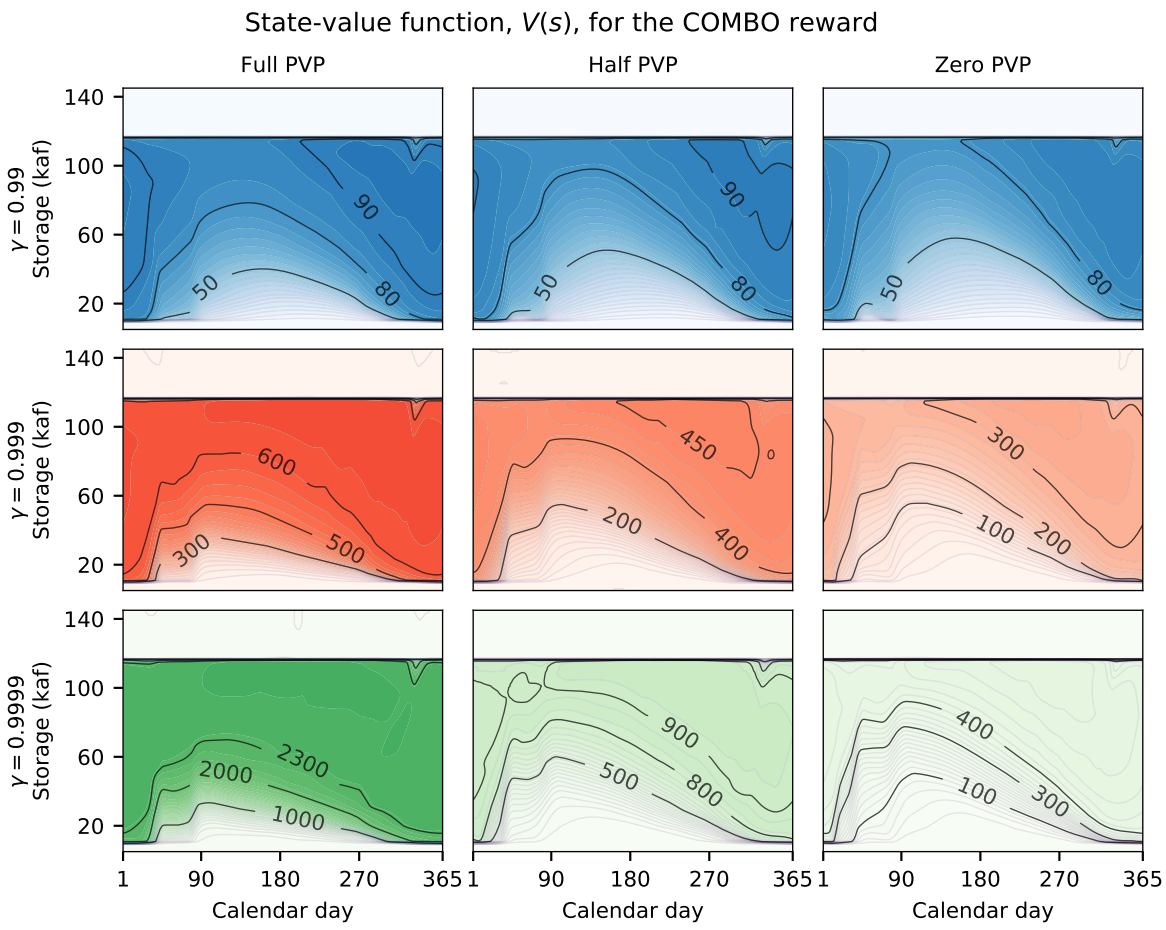

Figure 5.17: Heatmap plot of the $V(s)$ function for each scenario, darker color indicating higher values, the hue indicates the discount factor γ.

Policy Functions, $\mu(s)$

The policy function, $\mu(s)$, is evaluated at the same time as the state-value function. These arrays are again averaged to obtain the policy output of the learning process. These functions are shown in Figure 5.18. The policy returns the ratio of the allowable release, in this case 50 - 5000 af/day. The darker regions indicate higher releases and correspond with the wet season as expected. With a value of $\gamma = 0.99$, the policy continues higher releases through the dry season. As the discount factor is increased, the releases during the dry season are reduced sooner. This agrees with the policy function which is able to consider carryover storage when the larger discount factors are used. As the contribution from the PVP project is decreased, the policy tends to show lower releases at the higher storage levels during the wet season as more supply must be retained to provide environmental flows during the summer.

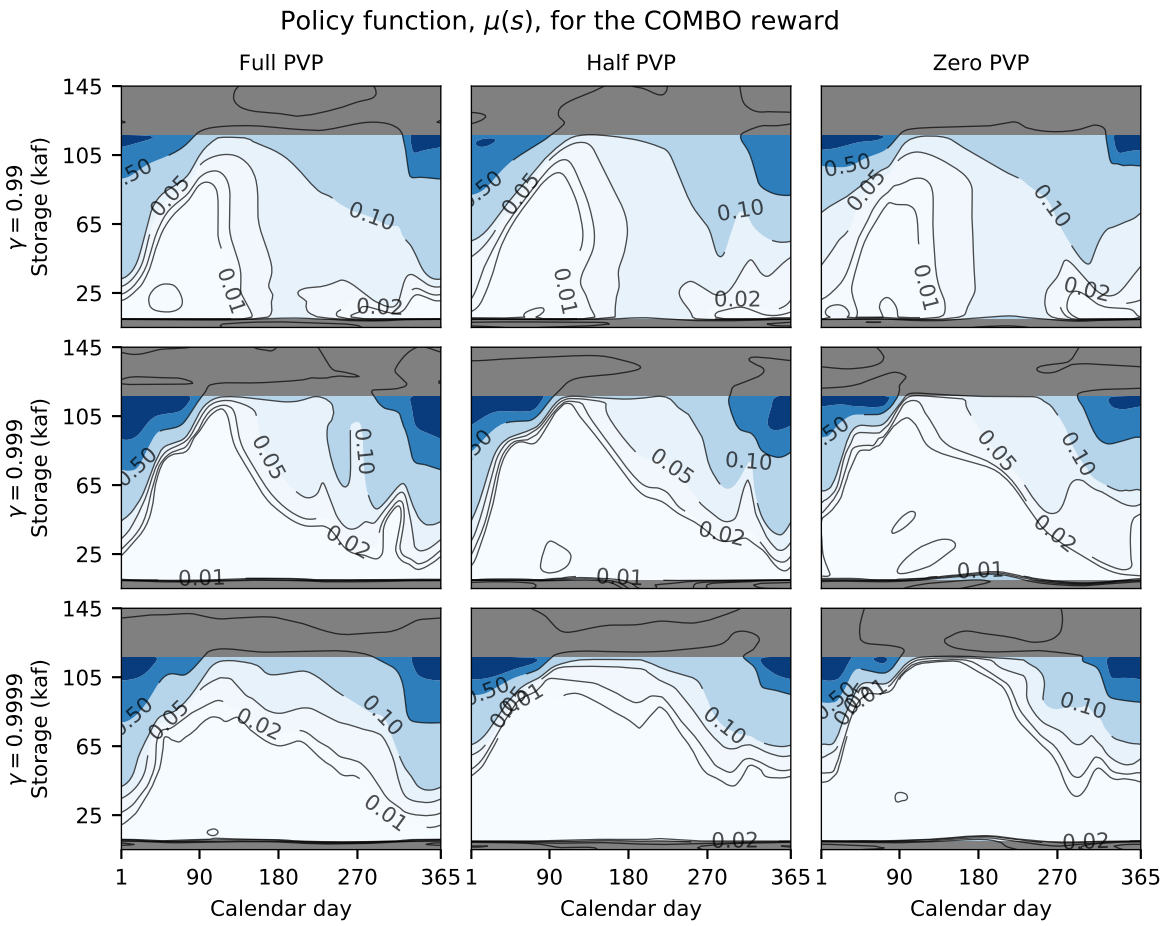

Figure 5.18: Heatmap plot of the $\mu(s)$ function for each scenario, darker color indicating higher values. The space above and below the storage limits is shaded grey as the policy is only applicable for the regions between the boundaries because the learning process terminates episodes when the agent encounters the boundary states.

Simulation

These policies were used in a simulation over the training and testing periods using the embedded network model and compared to a simulation of the guide curve operations from the HEC-ResSim model. The guide curve operations were modified to include a maximum release constraint of 2520 cfs or 5000 af/day to match the constraint on the learning model. When storage is above the guide curve, the HEC-ResSim model will make releases to return to the guide curve given the maximum release constraint as well as the downstream maximum flow constraint at Hopland. When the storage is below the guide curve, the model will release the minimum amount needed to meet the water supply demands and the minimum environmental flow objective. This then represents somewhat of an idealized operating plan as no excess water will leave the network.

The simulation which uses the policies obtained from the learning process does not have information about the demands at each time step. The policy reflects the learning model's attempt to make releases which have the highest expected value of the return for each state. Because of this, the operation is less efficient as at times the volume released will exceed the downstream demands and flow out of the network. At other times, the release will be less than the demand and the objectives will not be met.

Storage Plots: Full PVP

The storage plots for the "Full PVP" scenario for the training and testing periods are shown in Figures 5.19 and 5.20. From the perspective of the water supply and environmental flow objectives, the dominating event during the training period is the extreme drought during 1977 when the usually wet winter provided very little supply for the following irrigation season. The shaded region represents the time period beginning on October 1, 1976 and extending through December 1, 1977. In the simulations which used an RL policy with a value of $\gamma = 0.99$ and with the guide curve operations, storage reaches the minimum storage for extended periods. When the discount factor is increased to 0.999, the minimum storage during this period is 10,616 af, just above the minimum. When the discount factor is increased further to 0.9999, the minimum storage is 13,756 af. This suggests that the higher discount factor would represent a more conservative approach to water supply operations as releases during the dry season are reduced at low storages to prevent emptying the reservoir.

The test period 1985-2010 does not have such an extreme event as the drought in 1977. In Figure 5.20 we can see the storage levels resulting from the simulation over the testing period. In general, a lower discount factor results in more drawdown during the irrigation season.

The minimum storage constraint drives the trade-off between avoiding an empty reservoir and meeting downstream water supply and environmental flow demands. The performance with respect to these objectives is discussed later. We will first consider the effect of reducing the PVP flows by examining the storage plots from the Half PVP and Zero PVP scenarios.

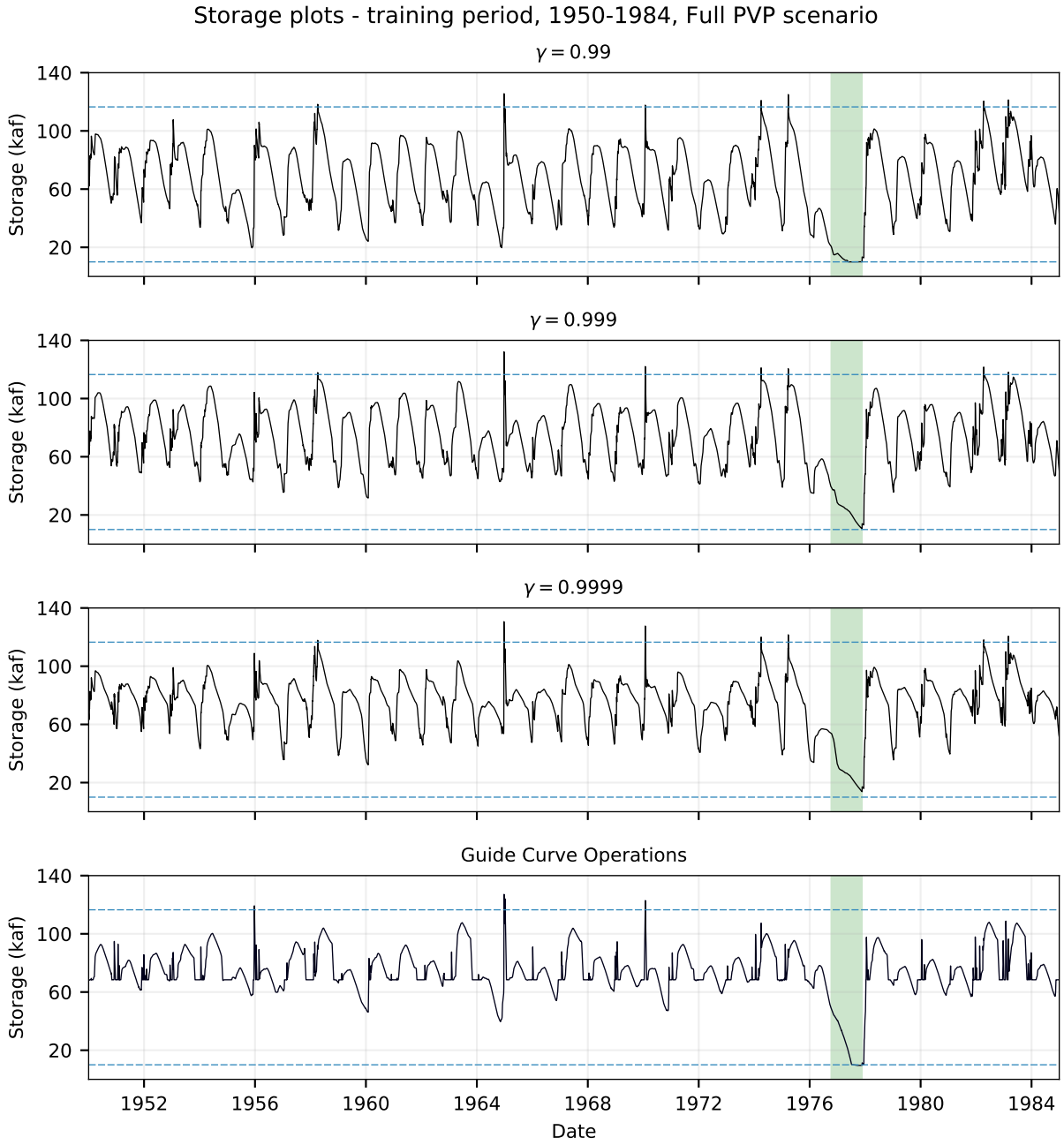

Figure 5.19: Plot of the storage level during the training period, 1950 - 1984, in the Full PVP scenario. Three policies obtained from the learning model are compared to the current guide curve operation. The policies from the learning model were obtained by changing the discount factor. These are compared to the storage resulting from the simulation which uses the current guide curve. The shaded region represents the severe drought that occurred in 1977. The results suggest that the higher discount factor may enable the model to become more far-sighted and learn from rare events.

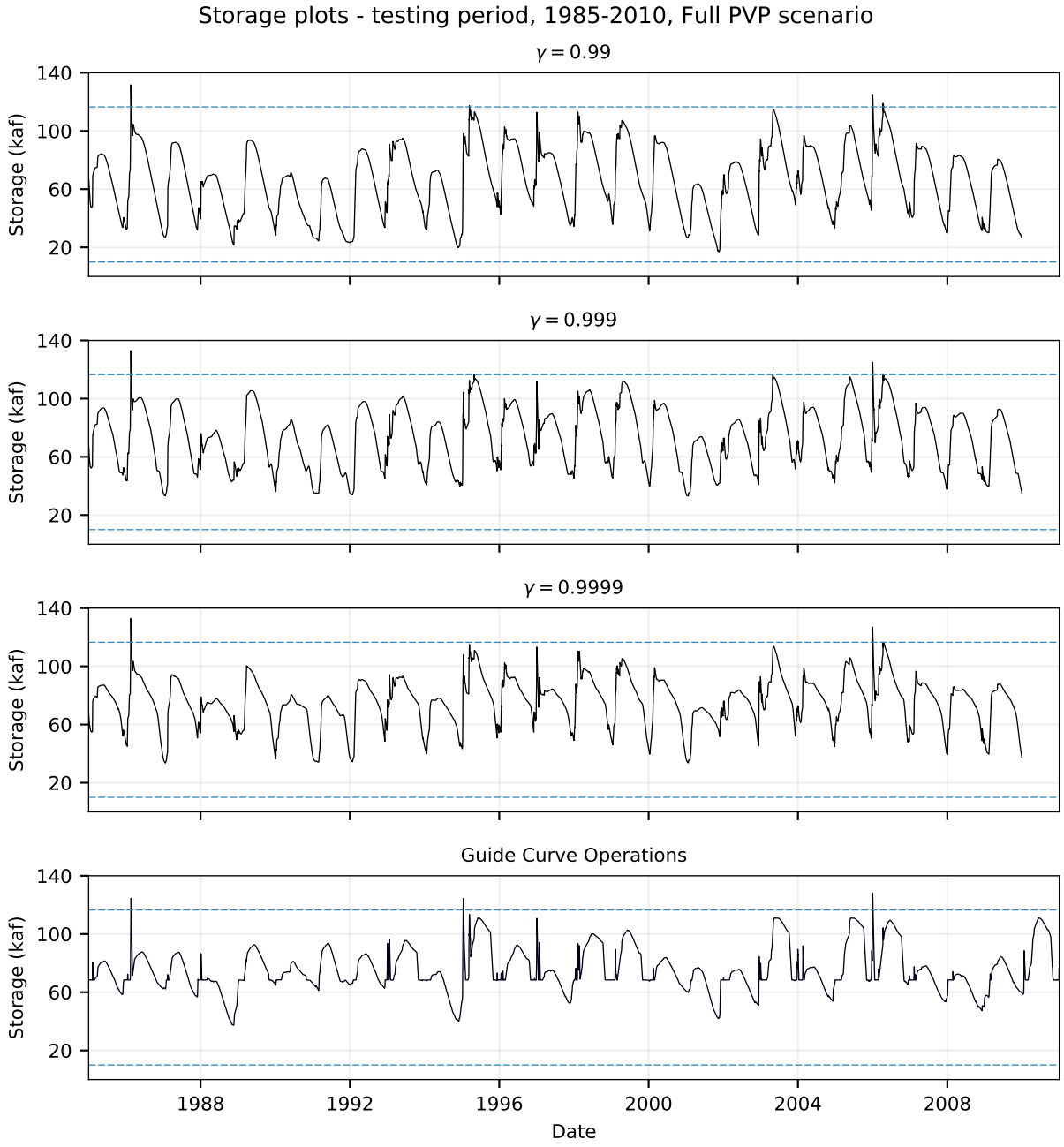

Figure 5.20: Plot of the storage level during the testing period, 1985-2010, for each policy obtained from the learning model and compared to the current guide curve operations. The increasing the discount factor generally results in higher minimum storage levels after the irrigation season.

Storage Plots: Half PVP

The storage plots for the "Half PVP" scenario are shown in Figures 5.21 and 5.22. We can observe a similar trend with higher discount factors showing increasingly conservative releases at low storages to avoid reaching the minimum storage during the drought period in 1977.

Storage Plots: Zero PVP

The storage plots for the "Zero PVP" scenario are shown in Figures 5.23 and 5.24. Again, the guide curve operations and the learning policy with discount factor 0.99 both reach the minimum storage during the drought in 1977. In this case, the guide curve operations also reach the minimum at multiple other times during both the training and testing periods.

Flood Scenarios

Another trade-off that is present results from the maximum storage limit of the reservoir. The flood control objective seeks to maintain capacity in the reservoir during the wet season to avoid exceeding the maximum storage level in the event of a large storm. The environmental flow and water supply objectives desire that the storage be maintained at higher levels during the wet season to meet demands during the following dry season. The current model is not appropriate for a detailed examination of performance during flood control operations. The major hindrance is the modeling of discharge through the spillway. The daily time step is too long to accurately model the change in storage that occurs when large inflows raise the elevation of the reservoir above the spillway. The elevation of the reservoir and the corresponding discharge through the spillway defined by the storage-discharge table are only calculated at the beginning of each time step. This discharge is then considered to continue for the full 24-hr period. The result is that when the elevation is above the spillway, very large discharges for the next time period are calculated reducing the storage in the reservoir more quickly than would be observed. The opposite is true when storage begins below the spillway and ends above the spillway. The storage will rise quickly as no spill is calculated for that period.

We can, however, gain insight into the learning process by observing the operations before the storage exceeds the spillway. The maximum storage levels reached during flood events in the long-term simulation are affected by the storage prior to the event. To evaluate the performance

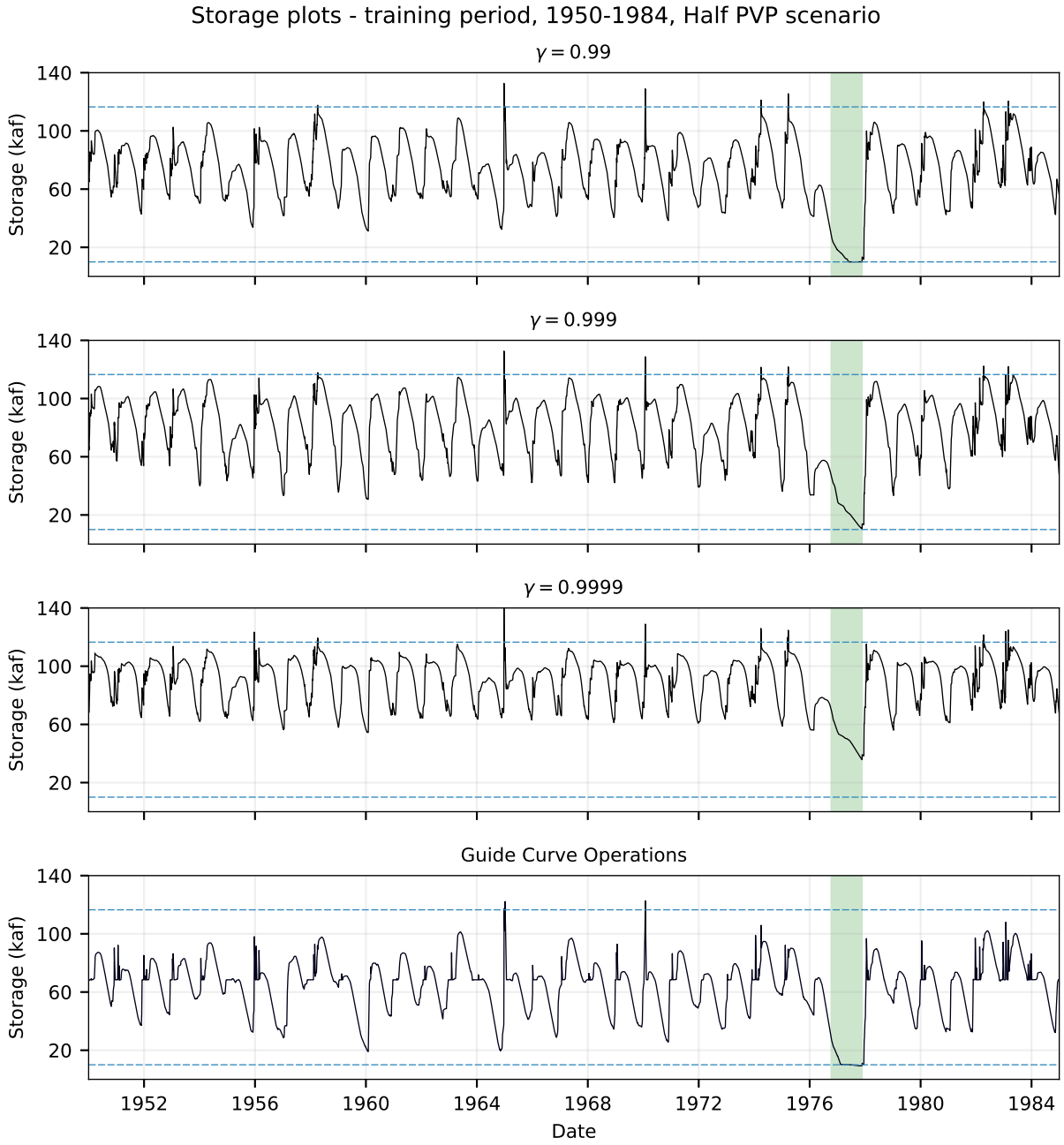

Figure 5.21: Storage plot for each policy in the Half PVP scenario. The guide curve operations and the $\gamma = 0.99$ policy both reach the minimum storage during the drought in 1977. The policies with higher discount factors maintain the storage above the minimum.

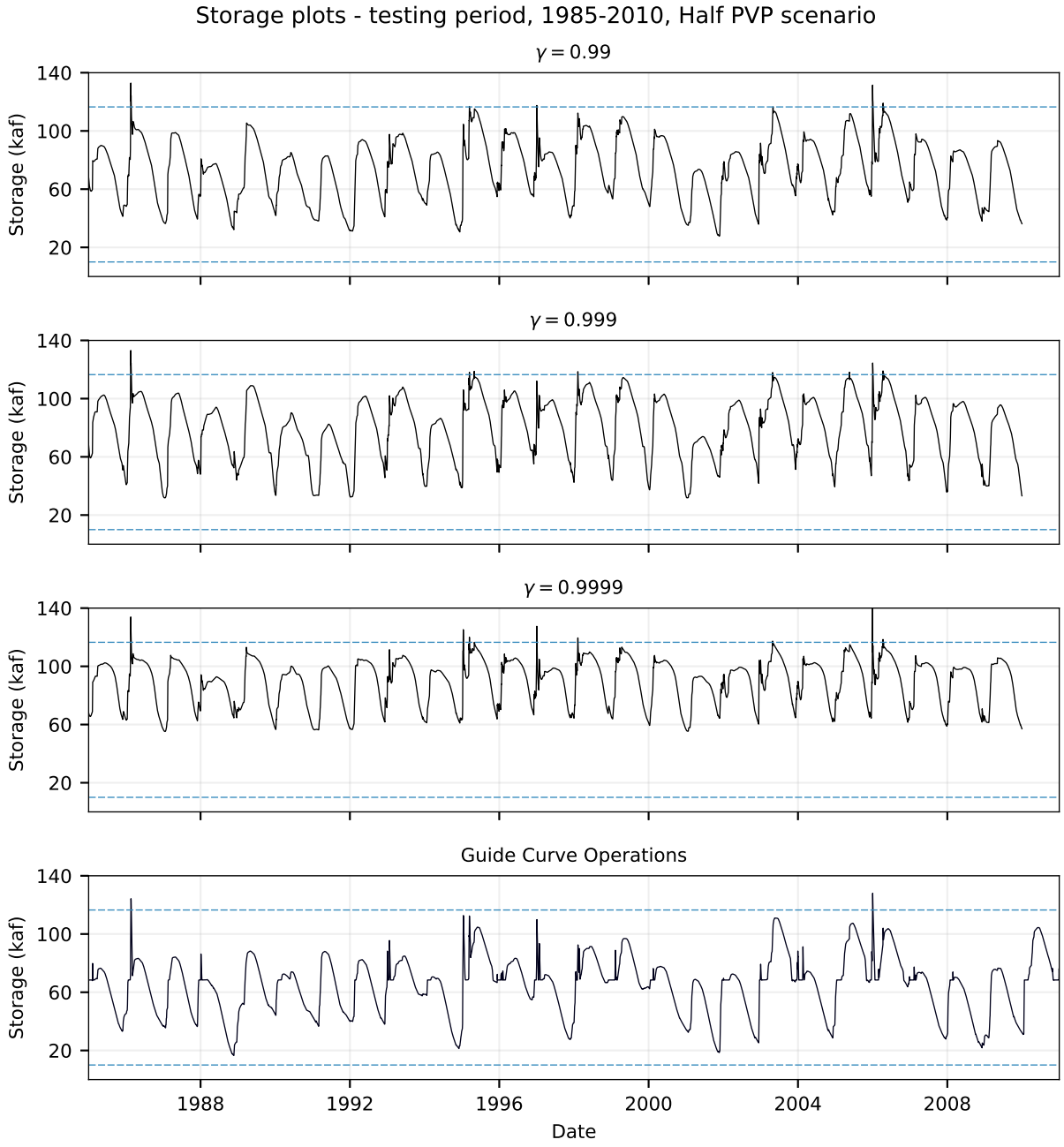

Figure 5.22: Plot of storage levels from the simulation during the testing period in the Half PVP scenario. The policies from the learning model maintain storage above the minimum throughout the testing period, with a higher discount factor generally maintaining higher storage levels.

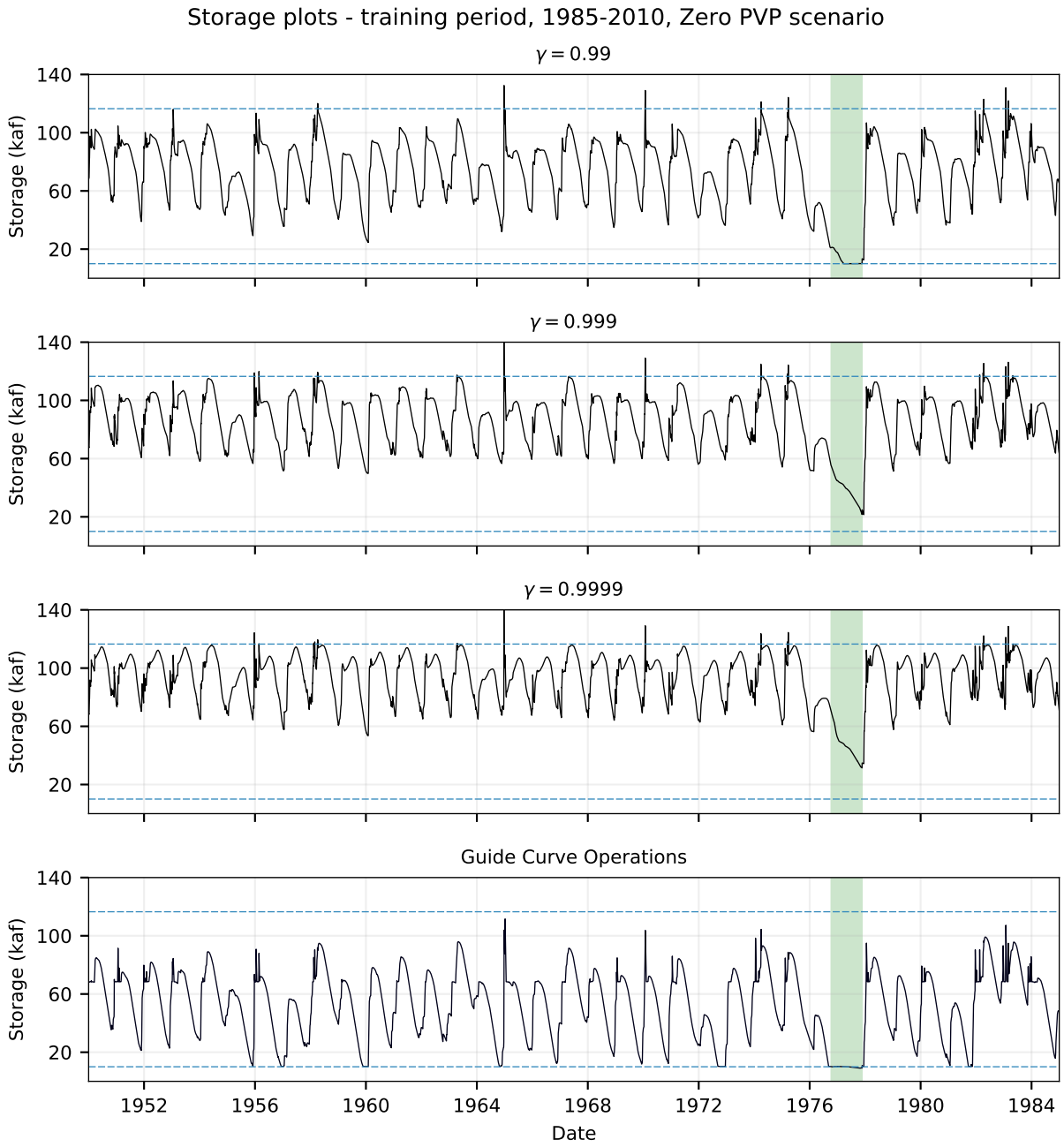

Figure 5.23: Plot of the storage level during the training period in the Zero PVP scenario. Storage under the guide curve operations reaches the minimum multiple times during the training period. The policies from the learning model which use a value of 0.999 or 0.9999 for the discount factor maintain the storage above the minimum throughout the training period, though with a trade-off in the performance with respect to flood control.

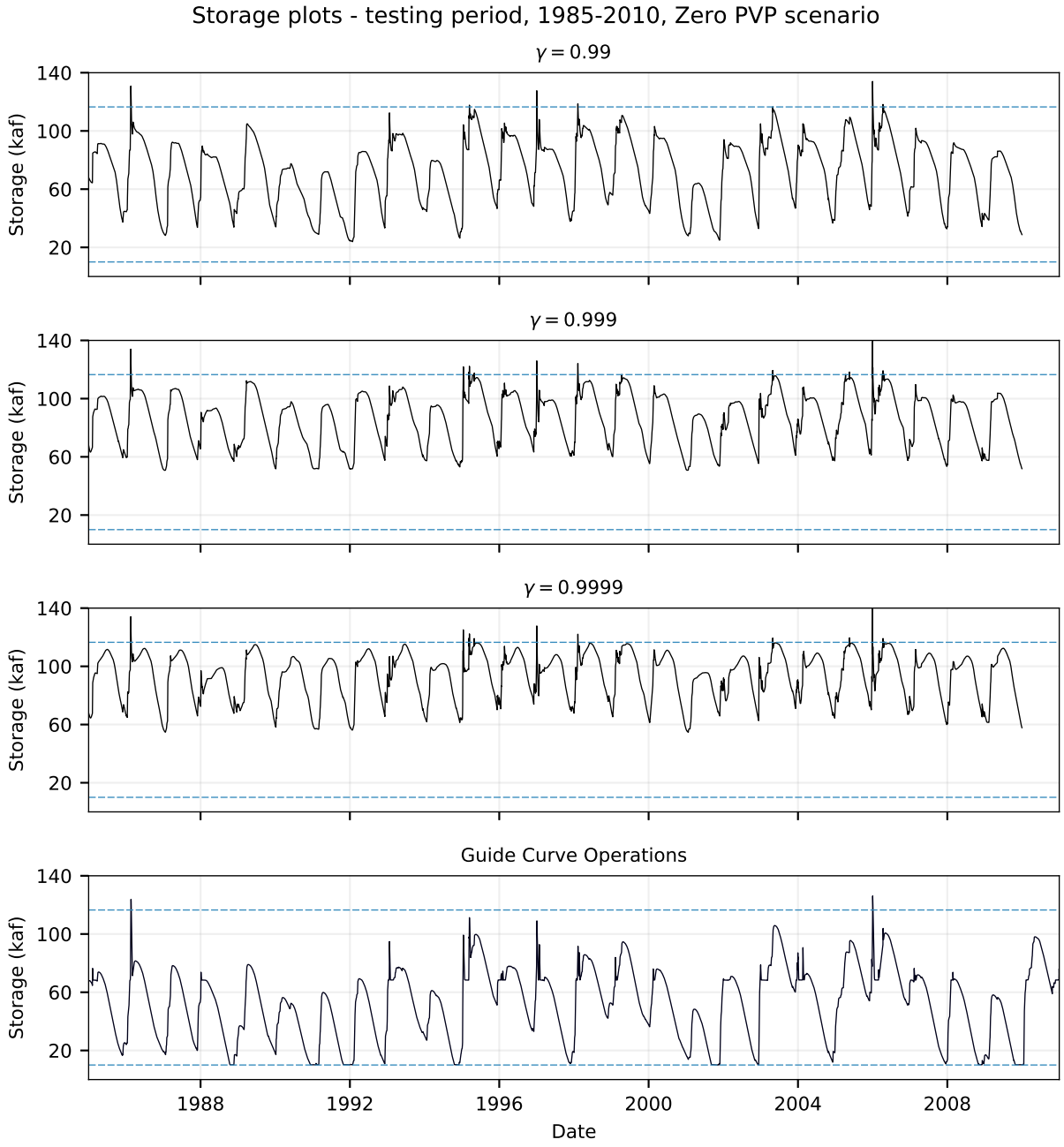

Figure 5.24: Storage levels during the testing period in the Zero PVP scenario. Reduced inflow to the basin causes the learning model to be more aggressive with respect to the flood control constraint, attempting to maintain higher storage levels to provide supply during the irrigation season.

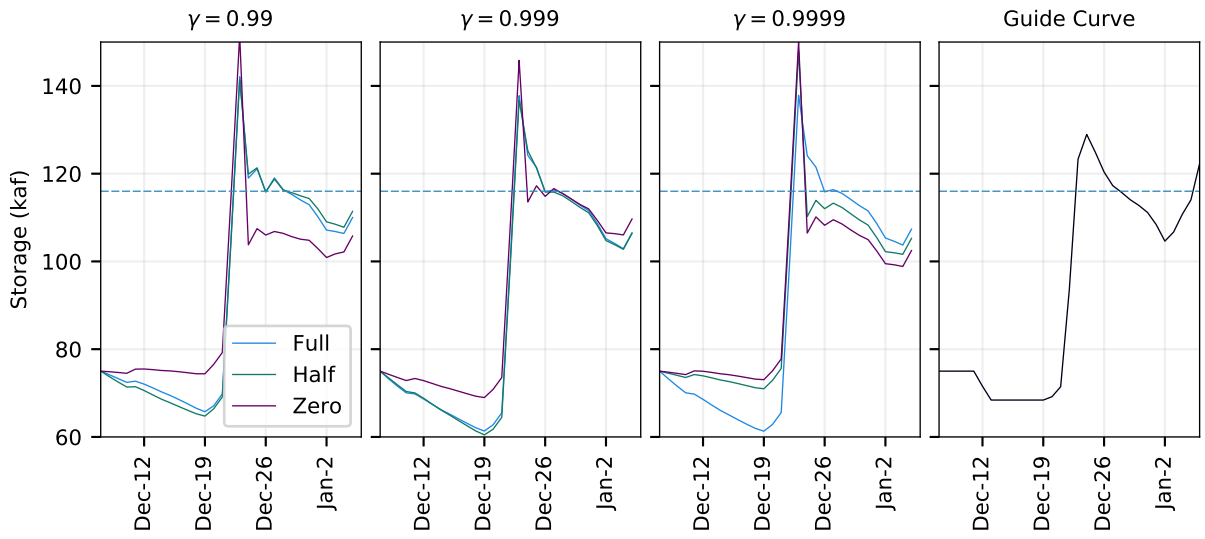

Figure 5.25: Performance of multiple policies during the flood event that occurred in December 1964. The policies for each level of diversion from the PVP are shown for three values of the discount factor. When the diversion is lower (Zero PVP), the policies tend to store water at higher levels, anticipating the lower supply that will be available for the irrigation season. This event is in the middle part of the wet season and when all of the policies from the learning model maintain the lowest storage levels.

of each policy with respect to flood control operations, two events were simulated; a flood event in December of 1964 and one in February of 1986. Each simulation began with a storage level of 75,000 af. This value was selected because it is slightly higher than the conservation pool level during the flood season defined by the guide curve operations. This allowed for a comparison of any drawdown prior to the flood event to be observed. Nine policies from the learning model and the current guide curve operations were simulated.

The results of the simulation of the event in December of 1964 are shown in Figure 5.25. The policies from the "Full PVP" scenario draw down the storage prior to the flood event for all values of the discount factor. The policies from the "Half PVP" scenarios draw down the storage prior except when the discount factor has a value of 0.9999. The "Zero PVP" policies do not draw down the storage or do so very little. Note that the releases made prior to the storm are based only on the current state of the system and no forecasting information is used in this simulation. The guide curve operations draw down the storage to the top of the conservation pool.

The results of the simulation for the flood event of February 1986 are shown in Figure 5.26. The policies from the learning model show similar behavior compared to the flood event of 1964, with the policies from the "Full PVP" scenario maintaining lower storage levels prior to the event,

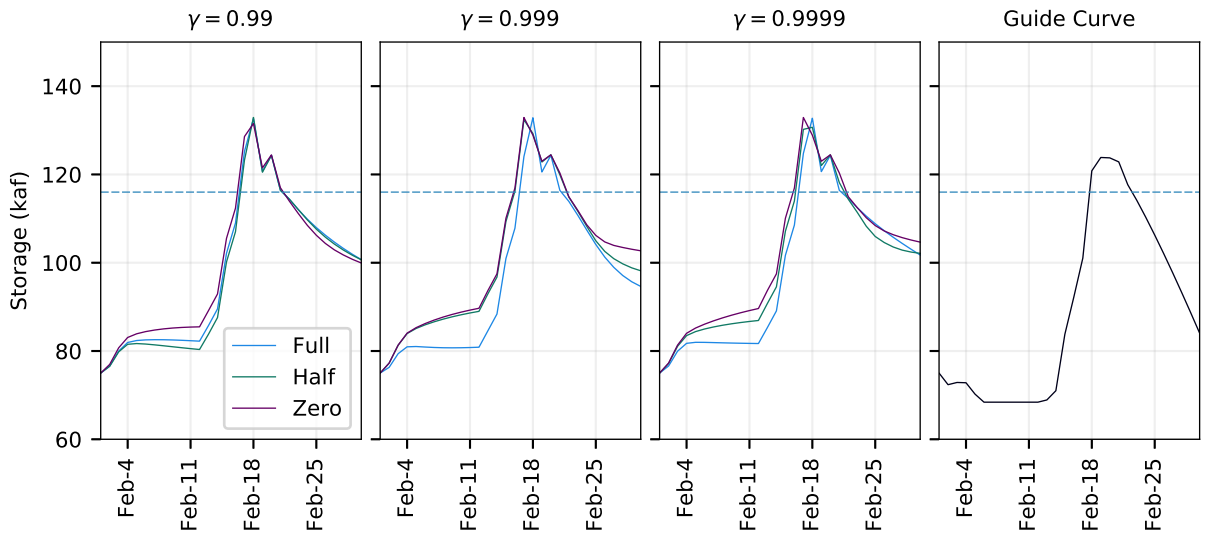

Figure 5.26: Performance of the policies during the flood event that occurred in February of 1986. The policies that were trained with the lower level of diversion from the PVP again favor higher storage levels. All policies favor higher storage compared to the event that occurred in December 1964 as this event is closer to the end of the wet season.

though in this case not decreasing the storage. This event is later in the wet season and all of the policies allow for higher storage at this time of year compared to the event in December 1964. In all cases, the policy is not able to maintain the storage below the spillway, including with the guide curve operations.

Water Supply and Environmental Flows

The reward that the agent received was based on the portion of the minimum environmental flow that was supplied, conditioned on water supply demands having priority and the requirement to meet the maximum storage limit at all steps. The environmental flow level was set to 149 af/day (75 cfs) and for the "Full PVP" scenario, the deficit in the environmental flow at the Healdsburg node is shown in Figures 5.27 and 5.28. The guide curve operation meets the environmental flow objective at all times during the training period except when the reservoir is emptied during the 1977 drought, in which case the deficit is the full value of the minimum flow, 149 af/day. The policies from the learning model show an increase in the frequency of the deficits, with most having a value less than 149 af/day except for during the drought in 1977 when deficits do reach the maximum value and do so for a longer period of time compared to the guide curve operations.

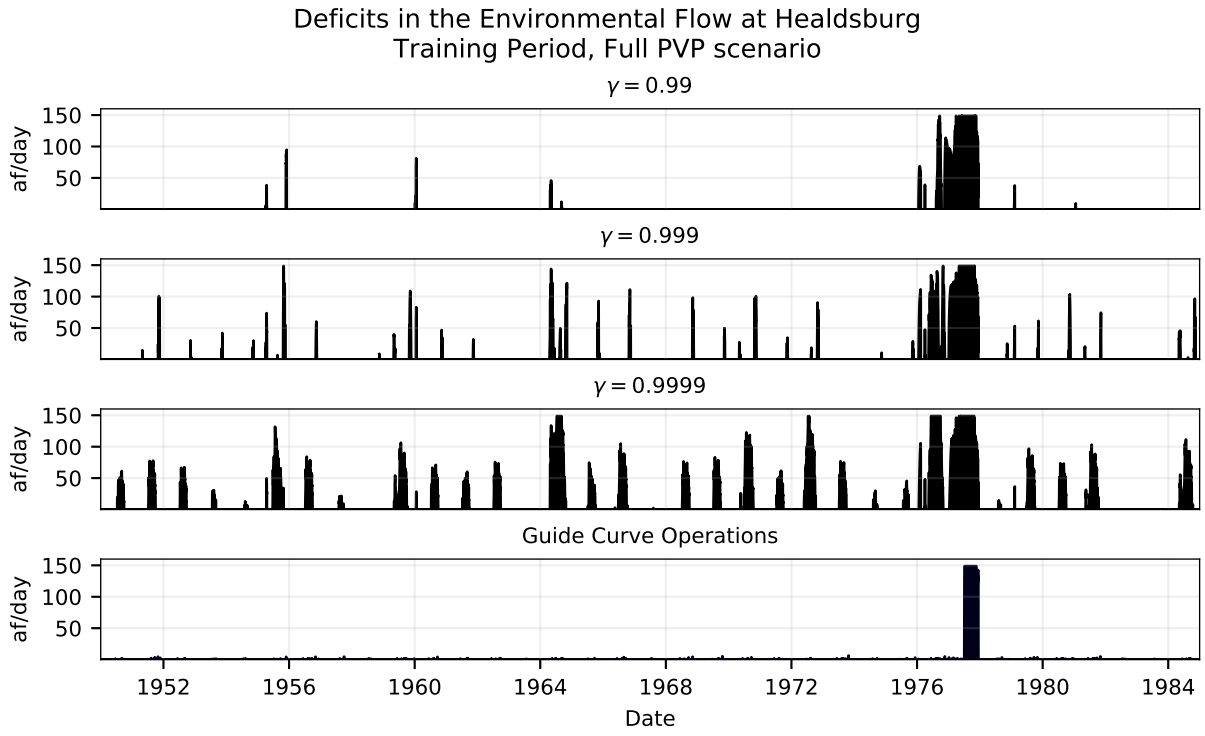

Figure 5.27: Deficits in the environmental flow requirement of 149 af/day (75 cfs) at the Healdsburg node for the Full PVP scenario during the training period. The guide curve operations meet the environmental flow target at all time steps except during the drought event of 1977. The learning models all show more frequent deficits with the higher discount factor showing the most frequent occurrence of deficits.

The guide curve operations represent an idealized scenario where only the minimum amount of water needed to satisfy the environmental flow is released, though without the ability to reduce flow to avoid emptying the reservoir in the case of a long term drought such as that during 1977. The policies from the learning models do not have the benefit of knowledge of the downstream diversions and thus are less efficient, leading to more deficits. We can observe the advantage that the learning model has over the guide curve operations by examining the upstream node at Hopland and consider the water supply deficits at both nodes. Healdsburg is the furthest node downstream in the network and diversions are not assigned a priority, so any water that is available is taken to meet upstream demands before it reaches Healdsburg. Figures 5.29 and 5.30 show the water supply deficit as a fraction of the demand for the Healdsburg and Hopland nodes. At Healdsburg, the water supply deficit is the full amount of the diversion for all of the learning models during the drought. The policy with a discount factor of 0.999 is comparable to the guide curve operations, but the other two policies show more frequent deficits.

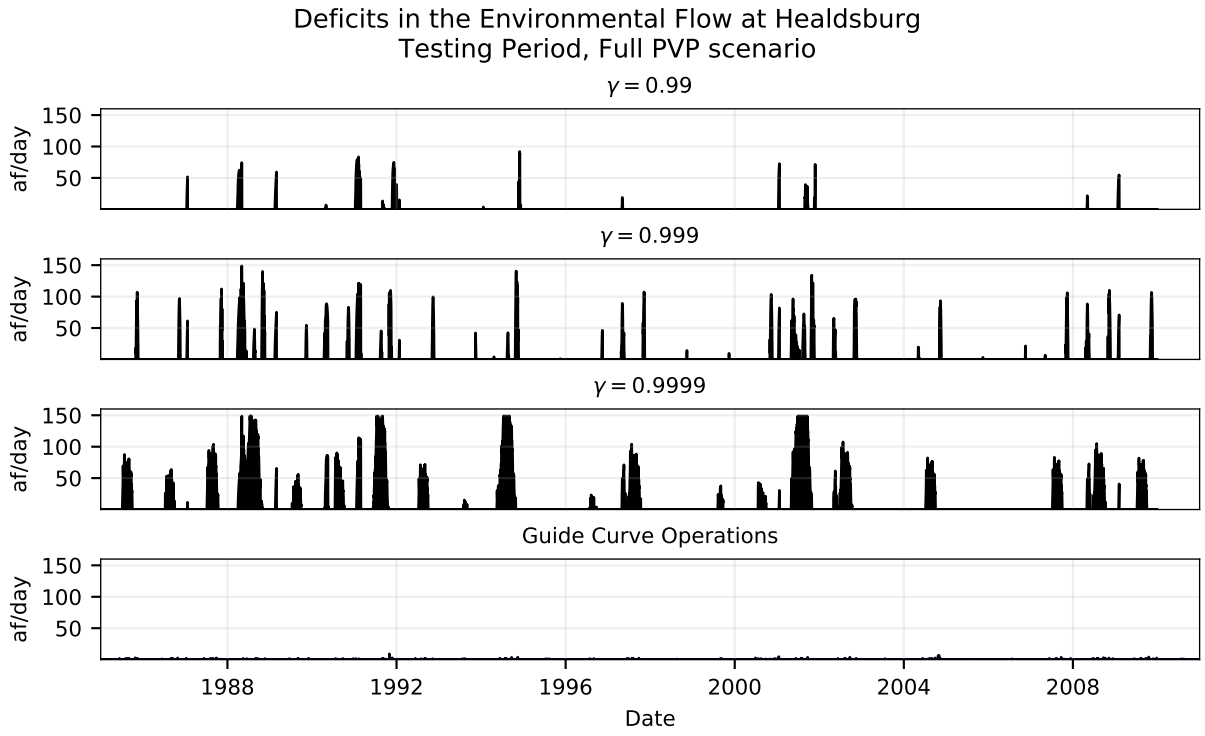

Figure 5.28: Deficits in the environmental flow requirement at the Healdsburg node for the training period. The guide curve operation is able to avoid any deficit during the testing period. Policies from the learning model show increasing frequency of deficits as the discount factor is increased.

When we look at the upstream node at Hopland we see that the occurrence of a full deficit is much less frequent for the policies from the learning model, but the guide curve operations still has a full deficit at the same time steps as when there is a deficit at Healdsburg. In the guide curve operations, the system is left with no more supply during the drought and is unable to meet demands at any point downstream. The policy with a discount factor of 0.99 appears to be too shortsighted and encounters the same problem as the guide curve operation no storage is remaining to meet any demand. When the discount factor is increased, the model is able to learn to avoid this situation and provides some level of supply throughout most of the drought period.

The difference is even more noticeable as the water supply stress to the system is increased. The water supply deficits in the "Zero PVP" scenario are shown in Figures 5.31 and 5.32. When the PVP diversion is eliminated, the guide curve operation more frequently results in full deficits for water supply at all nodes in the network. The policies with a discount factor of 0.99 meets the water supply demand at nearly all times until the reservoir is emptied during the drought leading to full water supply deficits for an extended period. The policies with higher discount factors are able

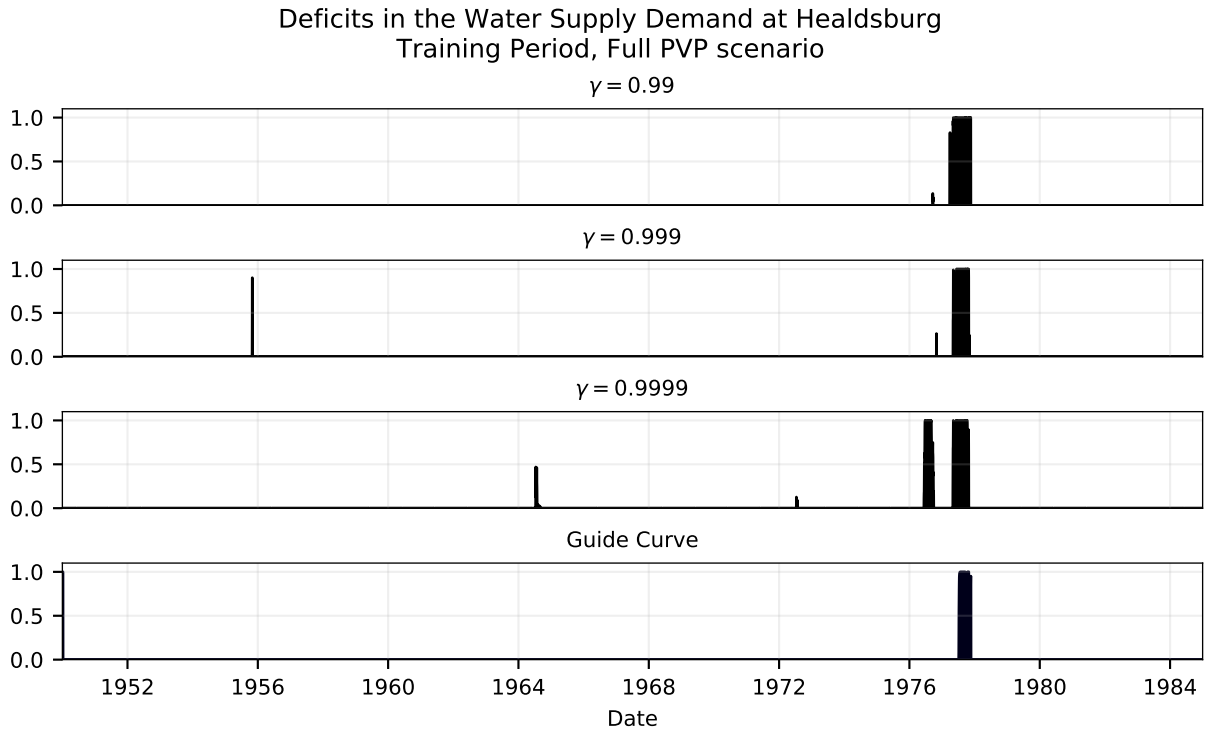

Figure 5.29: Water supply deficits at the Healdsburg node during the training period for the Full PVP scenario. All policies show full deficits during the drought period, with the policies from the learning model showing longer duration of the deficit period.

to avoid this situation by increasing the frequency of lesser deficits in order to avoid long periods with full deficits. The trade off is the increased flood risk associated with the higher storage that these policies maintain.

Performance Measures

To quantify the performance, a basic set of performance measures are introduced here. These generally align with standard calculations of reliability, resiliency and vulnerability with some simplification. Reliability to meet the environmental flow objective here is defined as the fraction of time periods in which the environmental flow target is met at all nodes to within a specified level of tolerance:

$$P_{rel} = \frac{1}{NK} \sum_{j=1}^{N} \sum_{k=1}^{K} \mathbb{1}_{\{q'_{j,k} > q_{env} - \epsilon\}}, \tag{5.17}$$

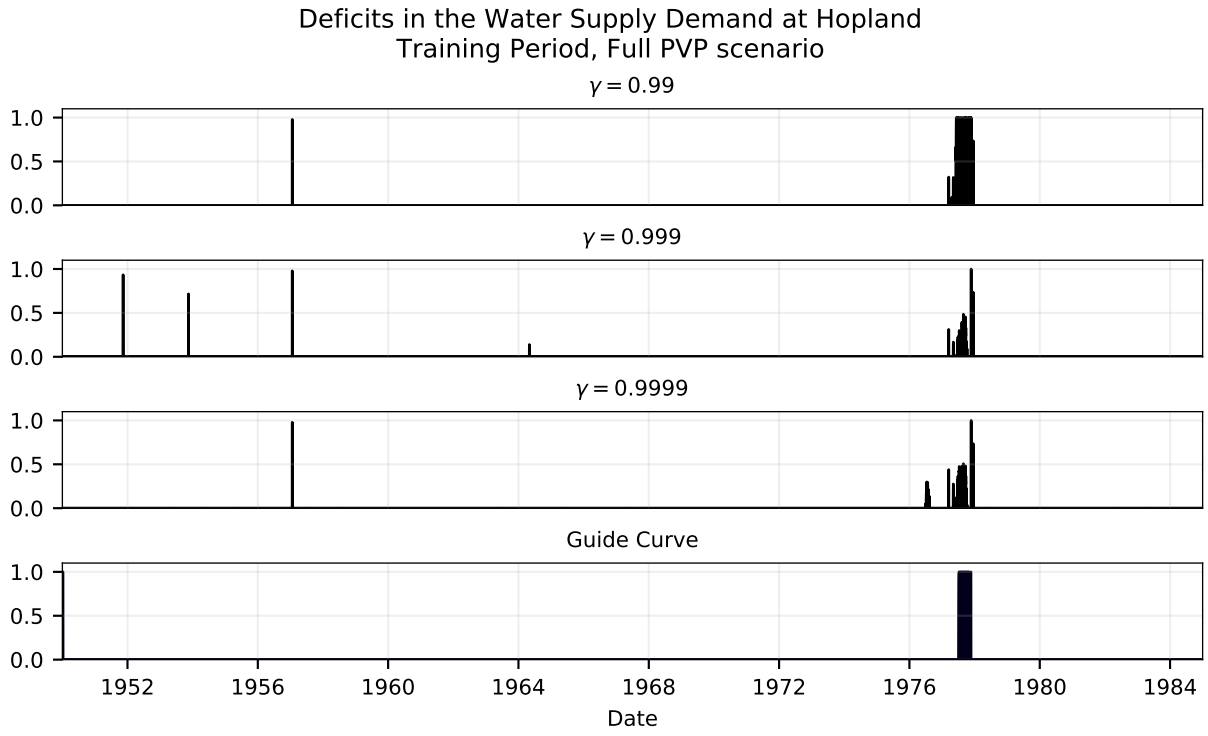

Figure 5.30: Water supply deficit at the Hopland node during the training period in the Full PVP scenario. The guide curve operation again shows full deficits during the drought. The policies from the learning model with higher discount factors show a shorter duration of the full deficit during the drought.

where N is the number of time steps in the simulation, q_{env} is the target environmental flow rate, K is the number of stream segments below the reservoir and ϵ is the tolerance. The operations which use the guide curve arrive at a decision by iteratively selecting a trial release, observing the downstream effects and adjusting the original release to meet the downstream demands. The HEC-ResSim model appears to do this to within approximately 10 af of the environmental flow demand, therefore a value of $\epsilon = 10$ af is used in this calculation.

Resiliency is defined here as the expected value of the number of consecutive time steps in which the environmental flow target is not fully met:

$$P_{res} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{M_k} \sum_{j=1}^{M_k} s_{j,k}, \tag{5.18}$$

where M is the size of the set of periods of consecutive failure and s is the length of each period of consecutive failures. Periods of consecutive failure are considered separately at each node.

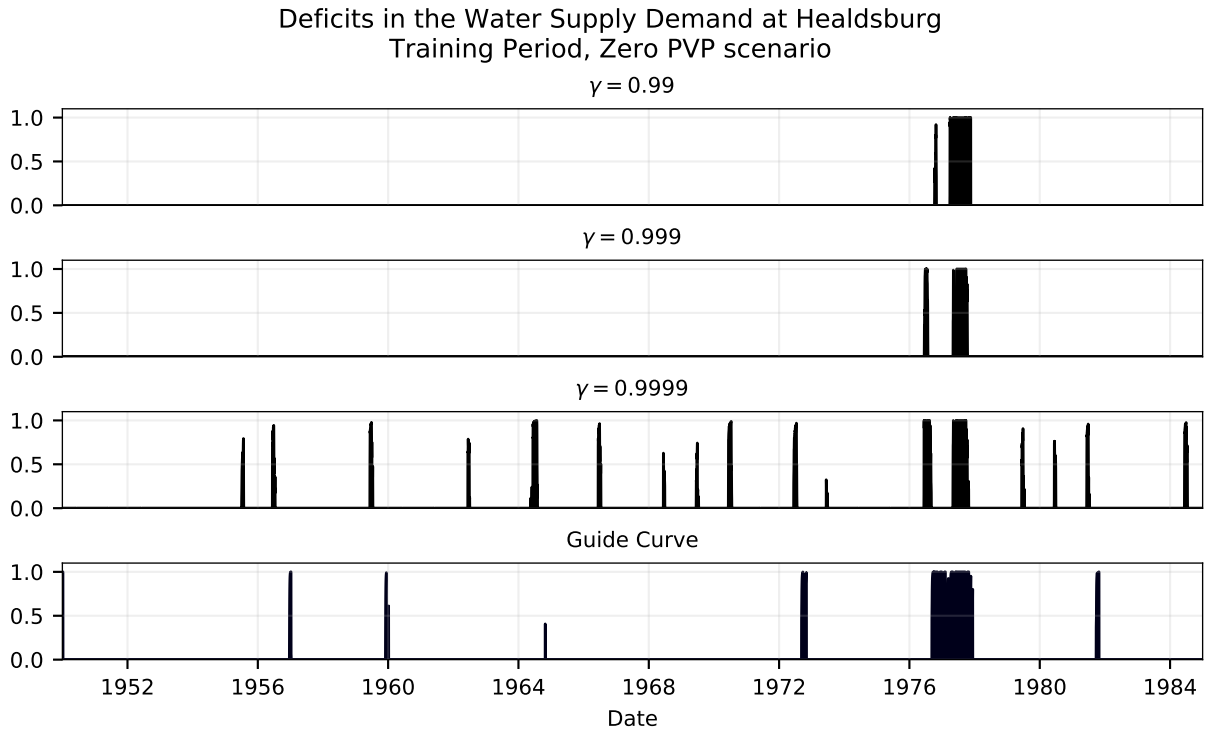

Figure 5.31: Water supply deficit at the Healdsbrug node during the testing period in the Zero PVP scenario. The guide curve operation shows multiple occurrences of full deficit with an extended period during the drought. The RL policies show full deficits of shorter duration during the drought. The policy with a discount factor of $\gamma = 0.9999$ shows the highest number of occurrences of deficit.

Vulnerability is defined here as the expected value of the deficit when a failure to meet the environmental flow target does occur:

$$P_v = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{L_k} \sum_{j=1}^{L_k} d_{j,k}, \tag{5.19}$$

where L is the number of time steps in which a deficit occurs, and d_j is the magnitude of the deficit at each time step. Again, deficits are considered for each node individually.

Performance - Flood Control

Tables 5.1 and 5.2 show the performance measures for each policy for the training and testing periods. The learned policies show similar values for reliability for each diversion level with a trend toward higher vulnerability values corresponding to higher discount factors as higher storage is favored leaving the system more vulnerable to exceeding the maximum. There is less distinction between the vulnerability values for the "Zero PVP" diversion which is due to the lower overall

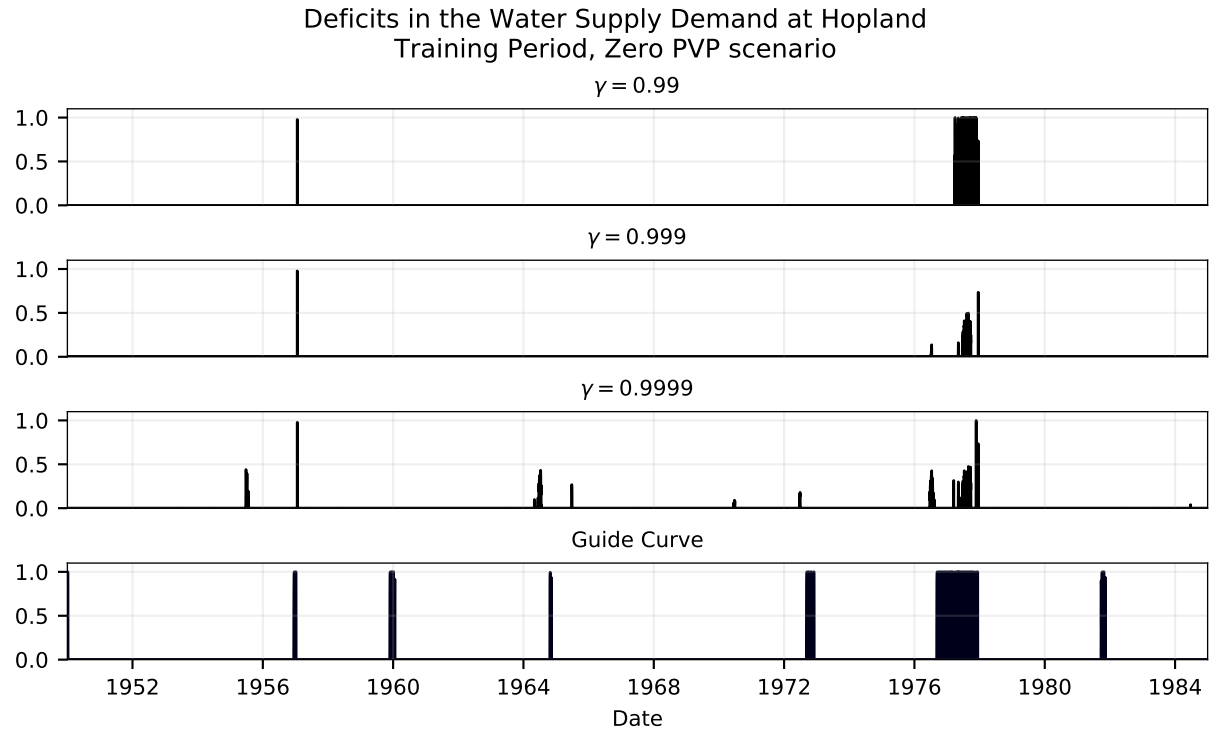

Figure 5.32: Water supply deficit at the Hopland node during the training period in the Zero PVP scenario. The guide curve shows full deficit at the same time steps when deficit occurs at the Healdsburg node. The policy with a discount factor of 0.99 shows full deficit during the drought, the other policies maintain some level of supply throughout the drought.

storage in the reservoir, leading to fewer and smaller excursions above the maximum storage limit. In the "Full PVP" scenario, the guide curve operations show reliability and vulnerability values similar to the $\gamma = 0.9999$ policy, though with a higher resiliency value, meaning that fewer events occurred where the storage exceeded the maximum, though each event lasted longer. As the diversion level is decreased, the performance of the guide curve operation with respect to the flood control objective improves over each measure. The guide curve operations do not compensate for lower values of supply by favoring higher storage levels, which results in lower overall storage levels providing more capacity to manage large storm events.

The performance during the testing period shows similar values for the reliability and resiliency measures, though with generally higher values of vulnerability, due mainly to the two large storms of February of 1986 and Jan 2006. The performance of the guide curve operation does not improve as dramatically as the diversion is increased during the training period as natural inflow is generally more available leading to higher overall storage levels.

Table 5.1: Performance measures for the training period.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Full PVP										
099	0.9988	1.9	3093.8	0.98844	20.1	45.4	0.9604	18.1	100.6	83
0999	0.9987	2.0	3730.3	0.99136	11.4	34.2	0.9273	10.3	76.8	85
09999	0.9988	1.9	4485.0	0.98607	15.3	35.0	0.8880	22.1	71.1	85
GC	0.9988	3.8	4483.2	0.99139	20.6	48.3	0.9878	32.7	142.7	92
Half PVP										
099	0.9982	2.3	3616.3	0.98868	20.7	45.9	0.9656	17.4	96.7	86
0999	0.9984	2.1	4287.1	0.98797	15.4	34.5	0.9292	19.0	80.1	87
09999	0.9980	1.7	5380.8	0.98740	16.1	35.0	0.8684	20.9	68.0	87
GC	0.9993	3.0	3367.1	0.98729	16.8	43.8	0.9766	42.6	142.1	92
Zero PVP										
099	0.9980	2.2	4669.7	0.98729	14.8	43.3	0.9663	16.4	105.1	89
0999	0.9974	1.6	4613.9	0.99029	17.7	34.6	0.9568	16.9	86.5	88
09999	0.9977	1.6	4749.5	0.97370	12.4	29.4	0.8251	34.1	88.8	86
GC	1.0000	0.0	0.0	0.97604	14.8	40.5	0.9460	28.1	138.6	89

Table 5.2: Performance measures for the testing period.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Full PVP										
099	0.9988	2.2	5014.0	0.99996	1.0	11.4	0.9763	7.6	42.2	39
0999	0.9986	1.9	4522.5	0.99985	1.3	14.4	0.9224	10.3	60.2	39
09999	0.9990	3.0	6799.3	0.99759	6.6	25.5	0.8898	18.2	59.1	39
GC	0.9986	4.3	5662.2	1.00000	0.0	0.0	0.9999	1.0	86.9	58
Half PVP										
099	0.9986	1.9	5199.7	0.99996	1.0	11.4	0.9779	9.1	47.0	38
0999	0.9978	2.0	4242.6	0.99949	3.5	16.7	0.9305	15.5	61.5	39
09999	0.9977	2.1	5826.7	0.99686	7.2	23.5	0.8707	19.9	57.4	40
GC	0.9991	4.5	6043.1	1.00000	0.0	0.0	0.9998	1.0	70.7	54
Zero PVP										
099	0.9980	2.2	5335.8	0.99996	1.0	11.4	0.9879	6.4	39.9	40
0999	0.9965	1.6	4508.1	0.99996	1.0	11.4	0.9666	8.9	44.4	39
09999	0.9969	2.0	5402.2	0.97932	13.2	27.8	0.8319	30.4	85.8	40
GC	0.9991	4.5	5452.3	0.98985	12.0	32.3	0.9560	19.6	132.7	53

Performance - Water Supply

With respect to the water supply objective, for the "Full PVP" and "Half PVP", the policy with a discount factor $\gamma = 0.999$ showed the best performance of the learned policies with respect to each performance measure. For the "Zero PVP" level of diversion, this policy had the highest reliability value, although the policy with a value of $\gamma = 0.999$ had the lowest resiliency and vulnerability values. The policies which used $\gamma = 0.99$ generally show high reliability values with corresponding high resiliency and vulnerability values as this is a more near-sighted strategy which seeks to more fully meet demand in the present at the risk of larger shortfalls in the future. The guide curve operation shows higher reliability in the "Full PVP" scenario compared to the learned policies, though with corresponding high values for the resiliency and vulnerability. Guide curve operation is the most near-sighted of the policies as it will fully meet demands until no supply is available resulting in performance comparable to the $\gamma = 0.99$ policy. The guide curve is outperformed by the $\gamma = 0.999$ policy in the "Half PVP" and "Zero PVP" scenarios. The $\gamma = 0.9999$ policy outperforms the guide curve operation in the "Half PVP" scenario and though it does not quite match the reliability of the guide curve operations in the "Zero PVP" scenario, it shows much better resiliency and vulnerability values.

During the testing period, the guide curve operation outperforms the learned policies in the "Full PVP" and "Half PVP" scenarios. As the supply further decreases in the "Zero PVP" scenario, the performance deteriorates and is comparable to the $\gamma = 0.9999$ policy. The $\gamma = 0.99$ and $\gamma = 0.999$ policies show the same performance measures which outperform the guide curve for each measure.

Performance - Environmental Flow

As for the water supply objective, the guide curve operation and the $\gamma = 0.99$ policy represent near-sighted policies which provide higher reliability values with the corresponding high resiliency and vulnerability values. The $\gamma = 0.999$ and $\gamma = 0.9999$ policies do not match the reliability of the other two policies but show better vulnerability measures. The $\gamma = 0.999$ policy also shows resiliency values near or better than those of the $\gamma = 0.99$ policy for each level of diversion.

During the testing period, again, the guide curve operation performs well until water supply becomes limited in the "Zero PVP" scenario and the reliability decreases with high values of resiliency and vulnerability. The $\gamma = 0.99$ policy which represents a policy which is slightly less

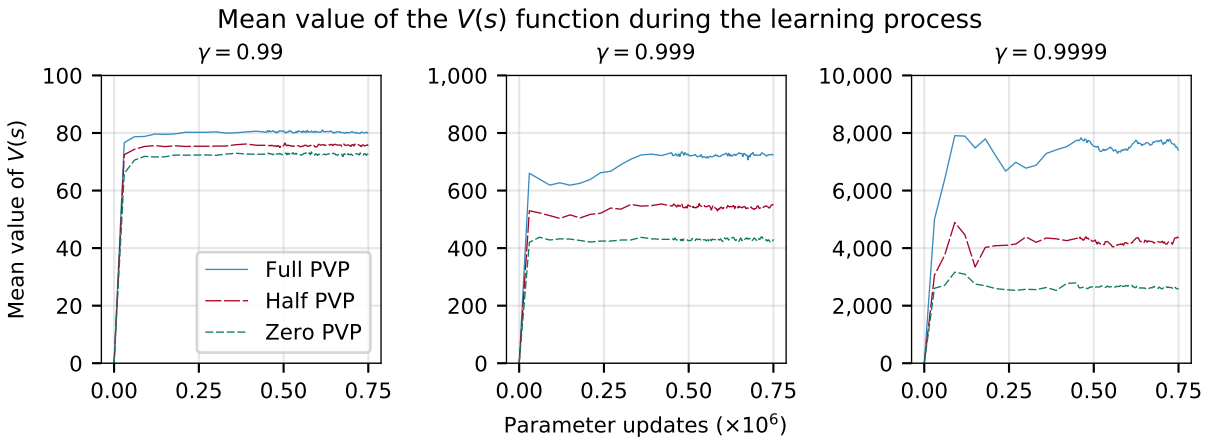

Figure 5.33: Mean value of the $V(s)$ function during the learning process for three values of the discount factor and each level of PVP diversion when the learning process is given the ENV reward. With a discount factor of 0.99, the mean value increases quickly then converges to a constant value with little noise. All results appear to show convergence within some level of tolerance with more noise present with higher discount factors.

near-sighted than the guide curve operation, shows the best performance of the learned models in the "Full PVP" and "Half PVP" scenarios, and the best overall for the "Zero PVP" scenario.

5.5.5 Environmental Flow Objective

The results obtained using the reward signal which combined the objectives suggest that the RL model is able to learn to operate the reservoir with some level of sophistication. The other objectives were then examined to estimate state-value functions to be used later with the ensemble forecasts.

The same combinations of discount factor and PVP diversion level were run with the learning process given the ENV reward defined by Equation 5.13 to estimate the state-value function for the environmental flow objective. The mean value of the $V(s)$ function during the learning process is shown in Figure 5.33. Figure 5.34 shows the state-value functions, $V(s)$.

Repeatability

The progress of the learning process was monitored by following the value of the mean of the $V(s)$ array. The results showed that the mean value generally converges to a constant value with some noise still present. However, the method relies on stochastic processes both when gathering experiences and when performing mini-batch gradient descent to adjust parameters of the ANNs. The question then arises as to how the learning method performs when repeated for an objective

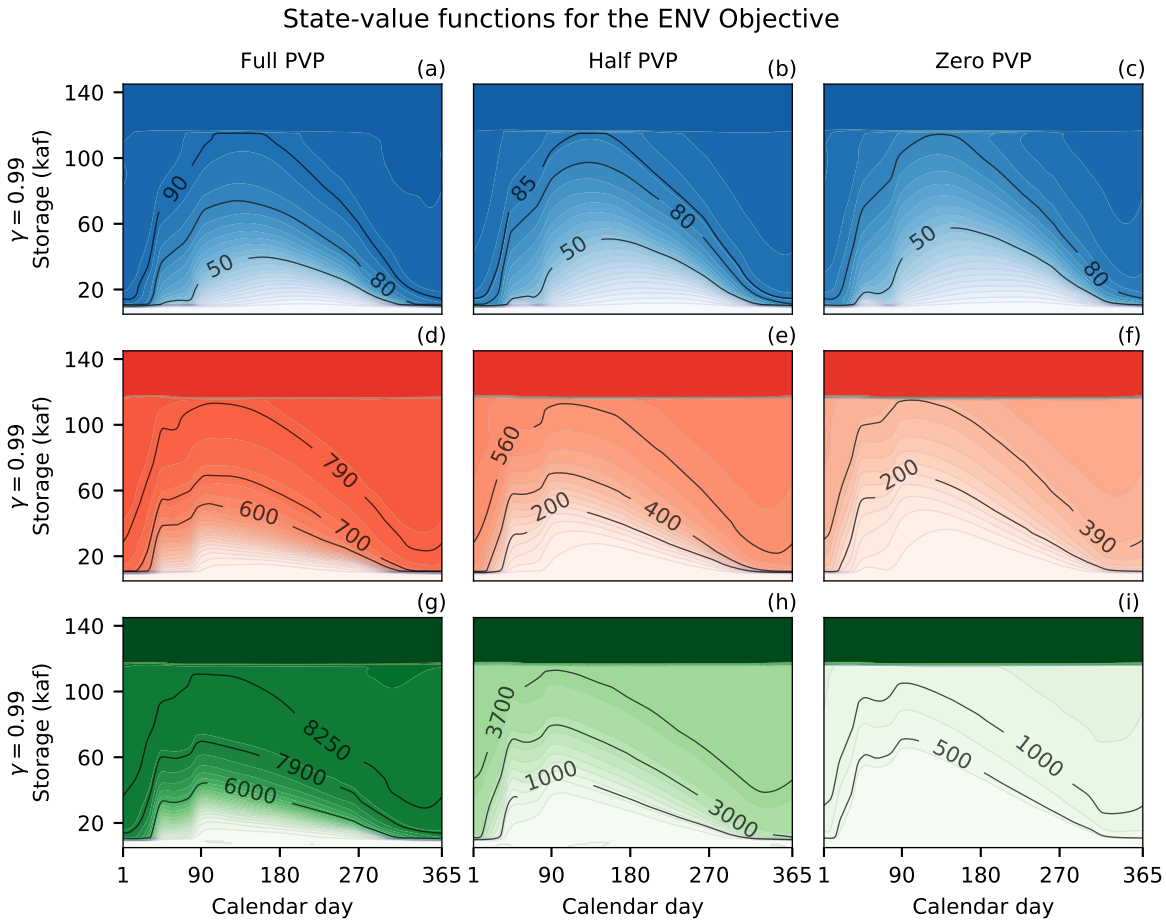

Figure 5.34: Heatmap plot of the state-value functions obtained from the learning process with the ENV reward. Darker colors indicate higher values with the hue corresponding to the discount factor. Each array is an average over the last 300,000 parameter updates. The value for storage above and below the storage limits is defined by the boundary conditions with storage above the spillway having the maximum value based on the discount factor, $1/(1-\gamma)$, and storage below the minimum having a value of zero.

with the learning parameters fixed. Also, the mean of the array over the state space only partially describes each array as there are multiple arrays which could have the same mean value.

To assess the repeatability of the method, the learning process was run five times for each value of the discount factor in the Zero PVP scenario. Figure 5.35 shows the progress of the mean of the $V(s)$ function and Figure 5.36 shows the resulting $V(s)$ function overlaid for comparison. The mean values of the $V(s)$ function show convergence to similar values for each session with the spread appearing to increase as the discount factor is increased. We can see this also by observing the contour plots of the $V(s)$ function. The results show that each training run produces contours that match well, with the contours for the lower value of the discount factor more closely matching each

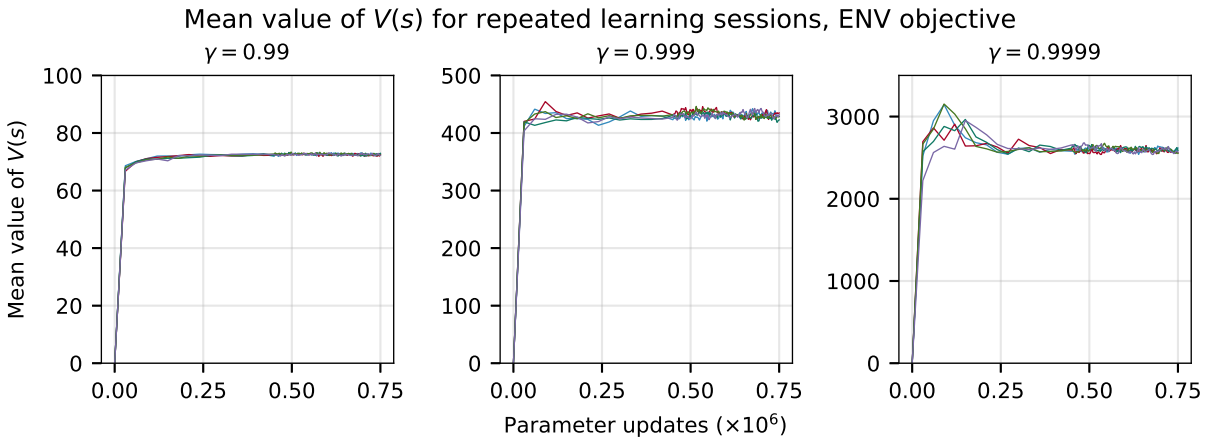

Figure 5.35: Mean value of the $V(s)$ function for 5 learning sessions at each value of the discount factor for the ENV reward in the "Zero PVP" scenario.

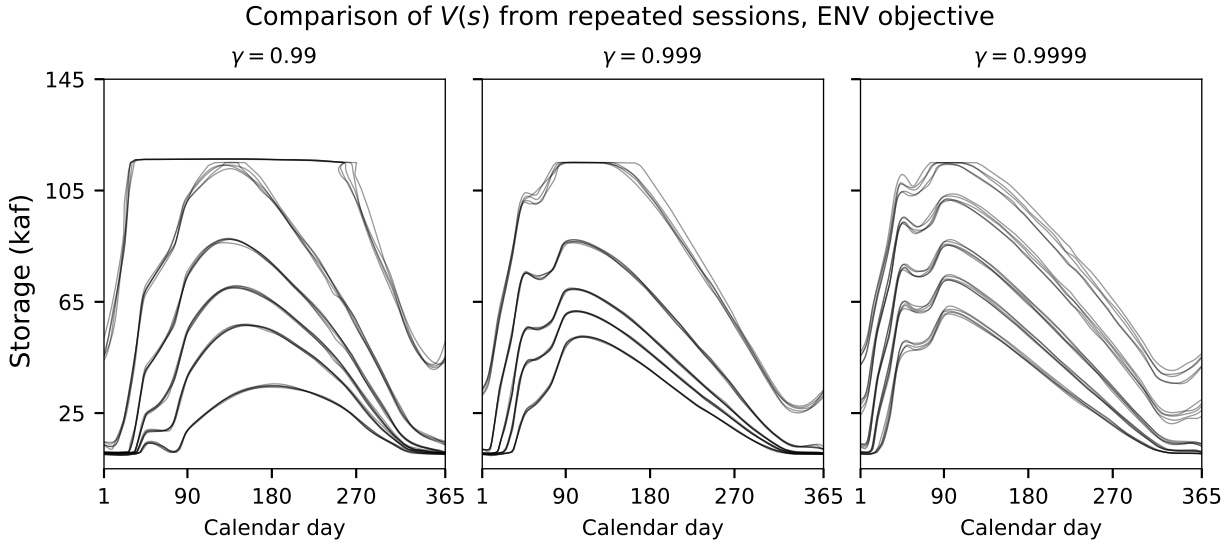

Figure 5.36: Contour plots of the results of running the learning process five times for each value of the discount factor for the ENV reward in the "Zero PVP" scenario.

other and there is more spread present in the contours with a higher discount factor. From these results we can only provide a qualitative assessment of the repeatability of the learning process. To assess the repeatability in some quantitative manner should take into account the use of the value function and is left for future research. For now, the method is determined to be suitable for the purposes of this research. Though this is restricted to the current objective under consideration.

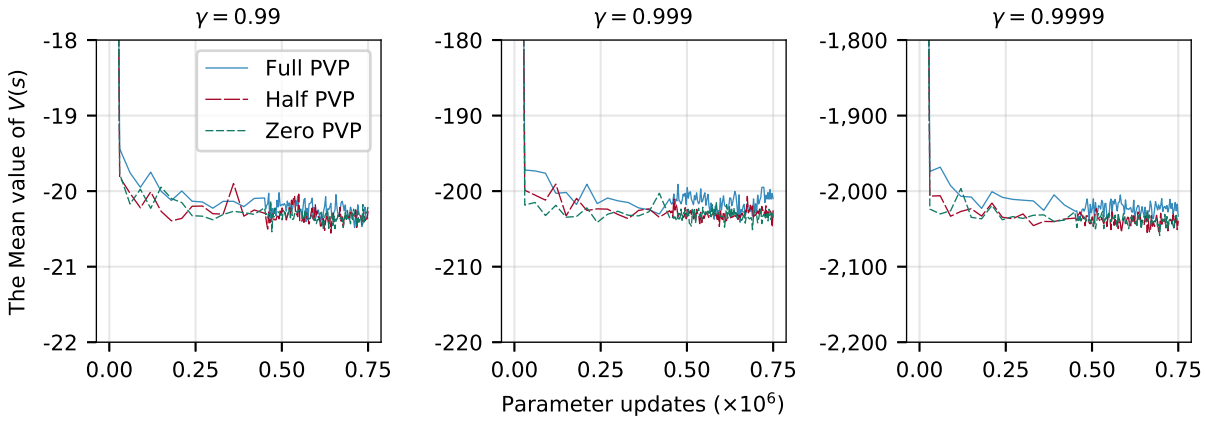

Figure 5.37: Mean value of the $V(s)$ function during the learning process for each scenario and discount factor for the FC reward.

5.5.6 Flood Control Objective

Throughout the development of the implementation of the DDPG algorithm, the FC reward showed more instability and ultimately required the addition of the prioritized replay method to make the learning process stable. The learning process was given the reward for the flood control objective (Eq. 5.14) and again executed for each level of the PVP diversion and three values of the discount factor.

The progress mean value of the $V(s)$ function during the learning process is shown in Figure 5.37. The mean value shows a similar tendency to converge during the learning process as was observed for the environmental flow and combined objective rewards. Figure 5.38 shows the resulting state-value functions and here we can see much different behavior than that observed in the COMBO and ENV objectives. For a value of $\gamma = 0.99$, the pattern of the contours follows generally what we would expect with the lower values associated with high storage in the winter months. The rate of diversion into the reservoir is only a small fraction of the maximum release and so we would not expect much variation in the contours as we change the level of the PVP diversion, yet the patterns appear quite different. Additionally, the state-value function in the "Half PVP" scenario is bimodal along the storage axis during the middle of the year. This is counter intuitive as more storage would seem to provide more likelihood that the maximum storage limit is maintained.

When a discount factor of $\gamma = 0.999$ or $\gamma = 0.9999$ is used, the results show very rough agreement with high storage during the winter having a low value, however, the paths of the contours are

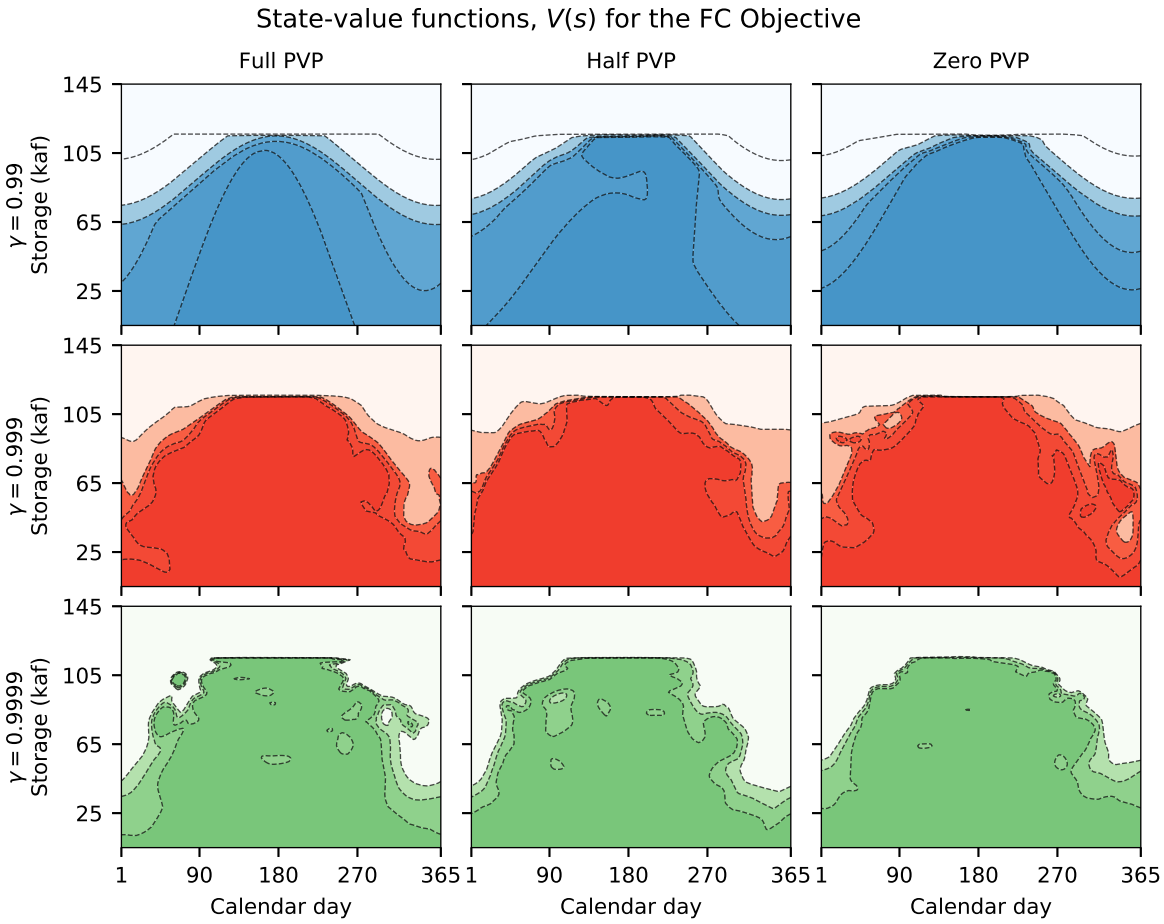

Figure 5.38: State-value functions obtained from the learning process given the FC reward and executed for each level of diversion and discount factor.

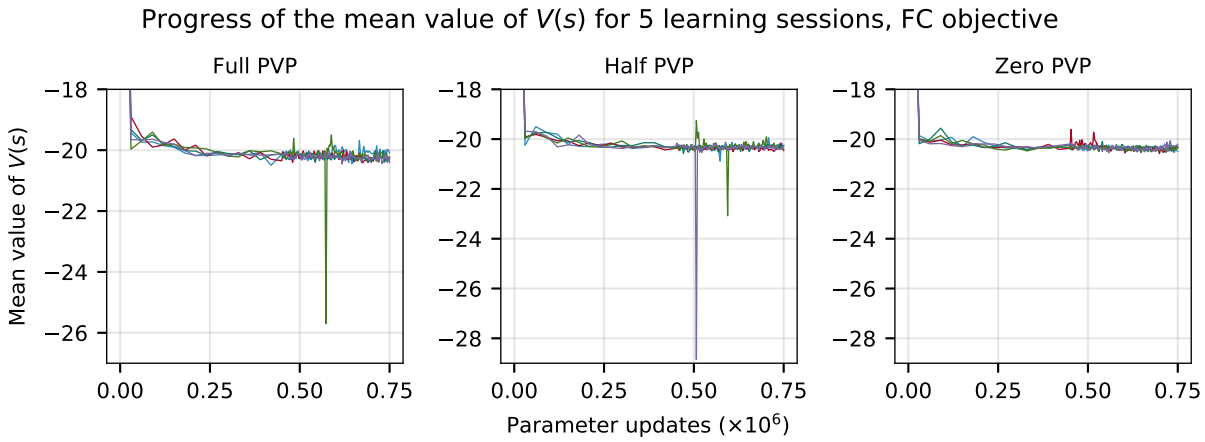

Figure 5.39: Mean value of the $V(s)$ function during the learning process given the FC reward, for each scenario and with a discount factor of $\gamma = 0.99$.

highly irregular. The cause of this behavior is unknown at this time. It is noted that the lower value of the discount factor would be more suitable to this objective as the outcome of an event with respect to flood control that comes a year into the future would seem to have little significance to decision-making in the present.

The learning process was run five more times for each level of diversion using a discount factor of $\gamma = 0.99$. Figure 5.39 shows the progress of the mean value of the $V(s)$ function for the repeated experiments. A difference observed in this case is the occurrence of rapid deviation from and return to the convergence value. Another interesting behavior is the appearance of a distortion such as in Figure 5.40 (a), (b), (h), (i) and (o). This is visually distinct from the strange pattern observed in the contours of the "Half PVP" and "Full PVP" scenarios in Figure 5.38. It was suspected that the distorted figures may correspond to the sessions that show large deviations in the mean value of $V(s)$ during the learning process, but this was not the case. Figure 5.41 shows the plots of the mean value of the $V(s)$ function with the traces which correspond to the distorted functions shown in red. We can see that although in the "Full PVP" scenario, the distorted function does correspond with the mean trace which showed the large deviation, this relationship is not shown in the "Half PVP" and "Zero PVP" scenarios.

Perhaps the learning process could be continued further to remove the distortion and provide better repeatability between learning sessions. Figure 5.42 shows the state-value functions from the flood control objective, in this case the average of the last 60,000 parameter updates of the

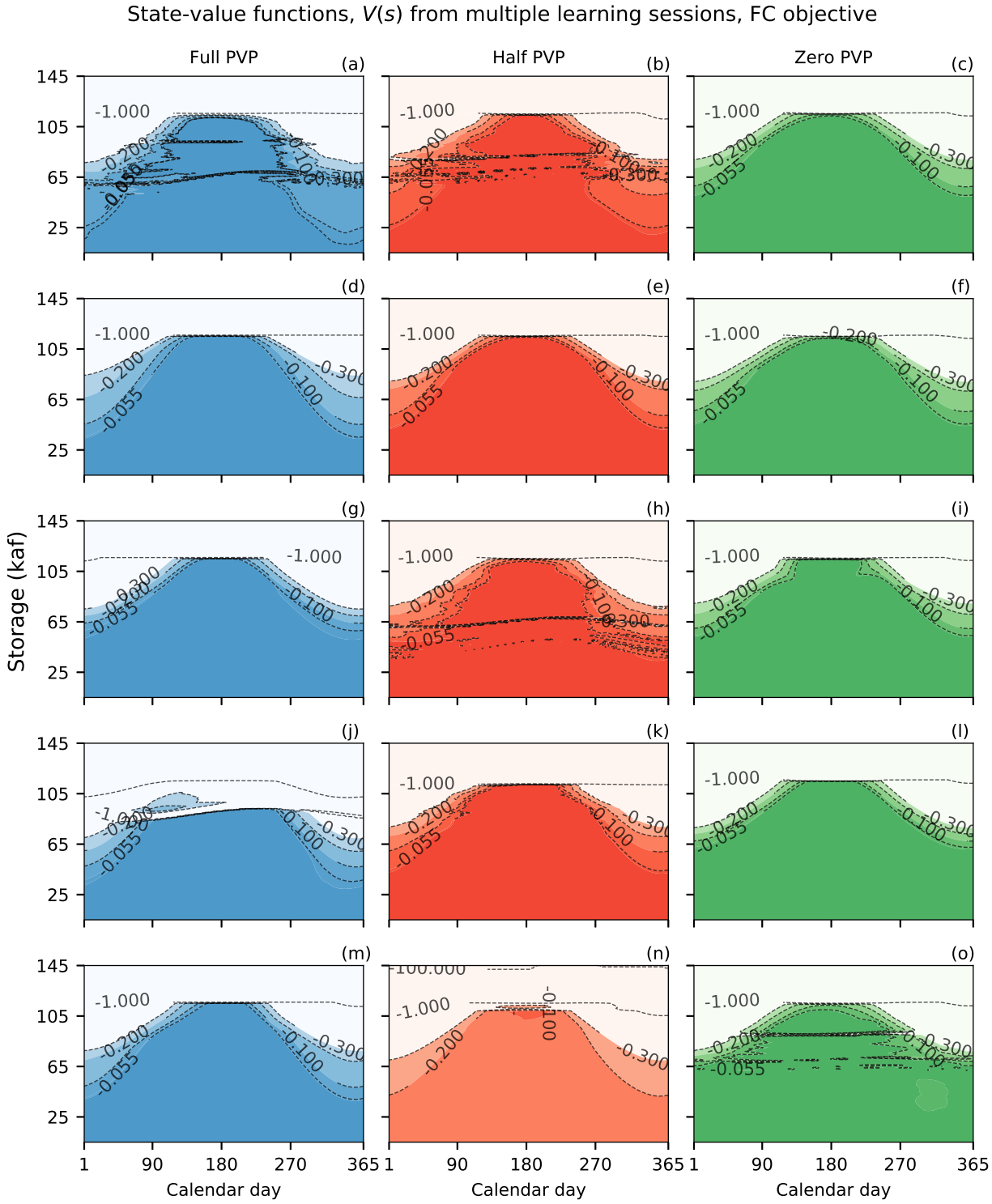

Figure 5.40: State-value functions obtained from the learning process given the FC reward and executed for each level of diversion and using a discount factor of $\gamma = 0.99$.

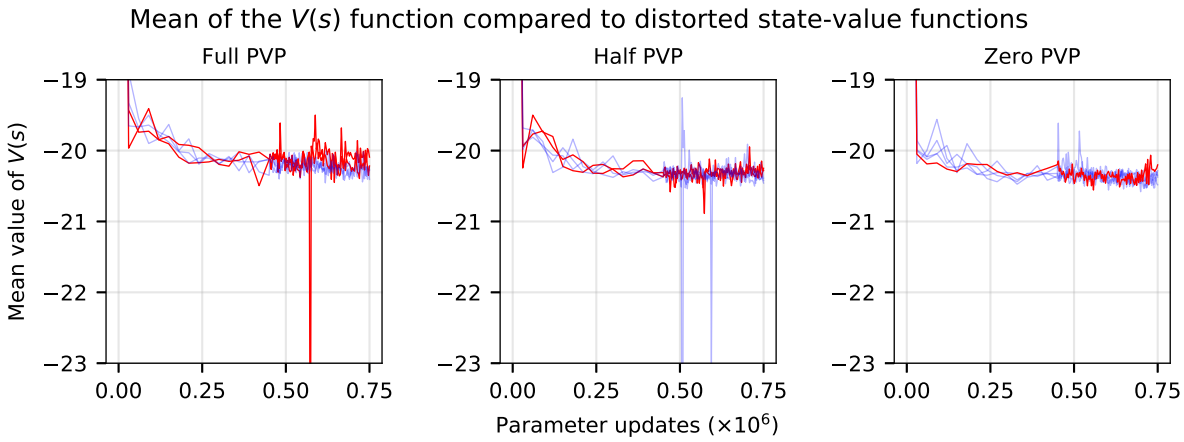

Figure 5.41: Mean value of the $V(s)$ function during the learning process. The red lines correspond to the functions which show distortion in Figure 5.40. In the "Full PVP" scenario, the trace with the large deviation does correspond to the distorted function, however, in the "Half PVP" scenario, the distorted figure does not correspond with a trace which shows the large deviation.

training process is shown. The average had been taken over the last 300,000 parameter updates. There appears to be less distortion which suggests that the distortion is occurring earlier in the training process and then reducing as learning progresses. Further research could examine the effect of extending the learning process.

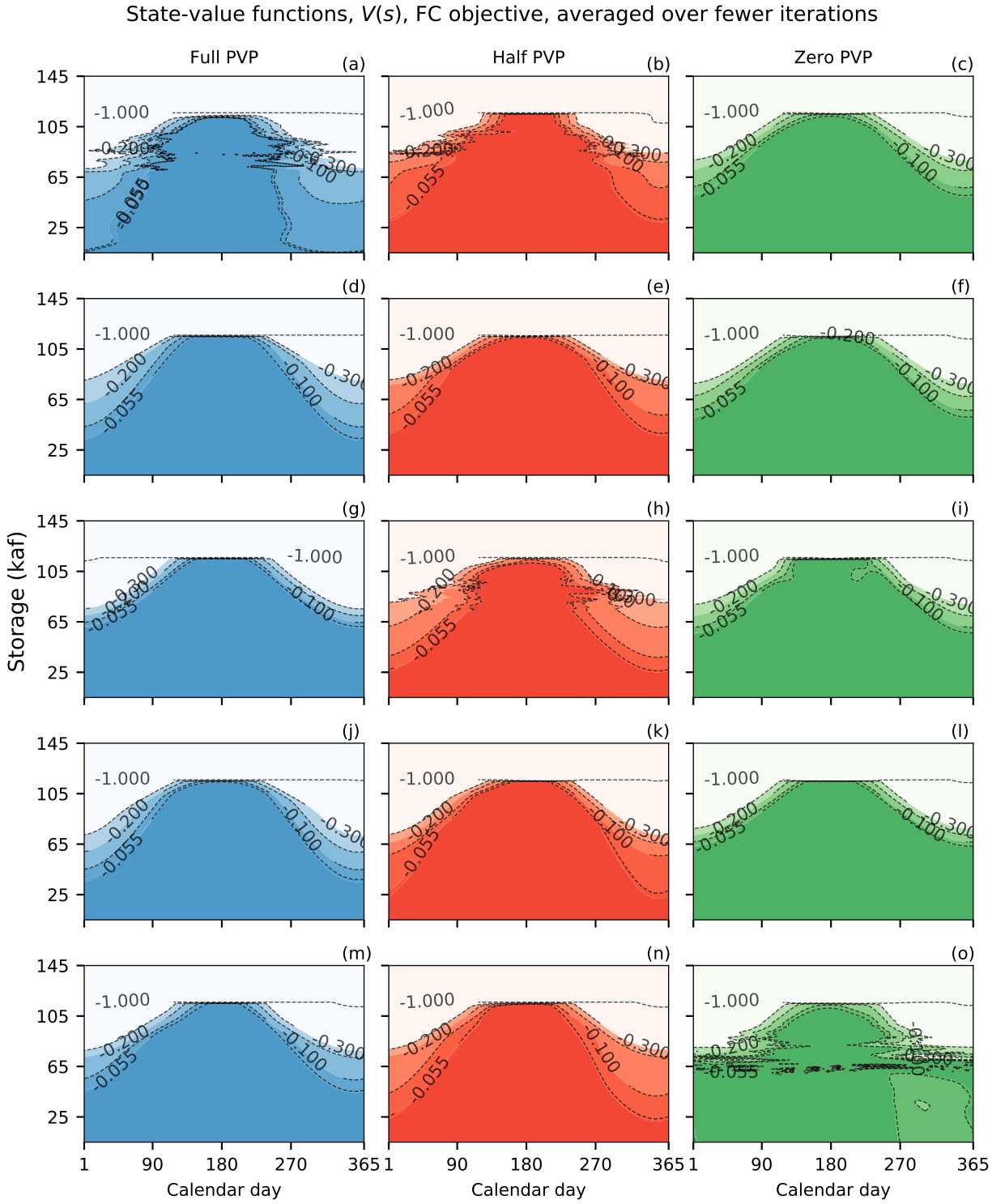

Figure 5.42: State-value functions from the repeated learning sessions averaged over the last 60,000 parameter updates, in contrast to the functions in Figure 5.40 which are averaged over the last 300,000. The functions in this figure show less or no distortion compared to those in Figure 5.40.

5.6 Operations Using Ensemble Forecasts

The state-value functions computed in Section 5.5 represent the value over all states of the system when no forecast information is present. When forecasts are used, this state-value function represents the value of the state at the end of the forecast horizon which is the estimated expected value of the discounted sum of all rewards that will be received after the forecast horizon. Section 5.5 showed the value functions obtained when the learning process is given a reward that represents a single objective. The discount factor is a critical value that is assumed in this process which may not be the same for different objectives. Because the objectives are considered independently, the simulation using forecasts can be carried out using a state-value function based on the discount factor that is appropriate for the objective.

For this research, the first experiments using forecasts considered the "Full PVP" scenario where the full amount of diversion from the PVP is used. The state-value function obtained using a discount factor of 0.99 was used for the flood control objective. Because the environmental flow objective is a longer-term objective which is affected by carryover storage, the state-value function with a discount factor of 0.9999 was used. The output of the reinforcement learning process is an ANN trained to approximate the action-value function, $Q(s, a)$. The proposed method requires the state-value function, $V(s) = \max_a Q(s, a)$. To avoid the excessive computing time that would be involved with performing the maximization step every time the state-value function is evaluated, the state-value function was approximated by sampling at points in a grid of size $[100 \times 365 \times 100]$ for the (storage, day, release) axes, with the maximum value along the action axis becoming the value of the state-value function at that location in the (storage, day) axes. In the simulation, the value of $V(s)$ then is determined through linear interpolation of the values of the nearest storage locations for the given day.

The information provided by the state-value functions allows the decision-maker to evaluate the trade-offs between objectives when selecting a decision. A decision-making framework is proposed here to show the benefit of the value function-based approach to using ensemble forecasts. In order to run a simulation, the set of rules for decision selection must be defined prior to execution. In practice, the decision selection process need not follow this particular form nor would it need to consist of inflexible rules. One of the benefits of this approach is that it does not replace the decision-maker. It simply illuminates the trade-offs between objectives.

5.6.1 Fitness Function

In the proposed decision-making framework, contours of the state-value functions are interpreted as levels of constant performance. A fitness function is defined which can be used by a particle swarm optimization algorithm to find optimal decisions during the forecast horizon for each scenario in the ensemble. This fitness function reflects generally a ranking of preferences for meeting the storage limits of the reservoir, meeting the flood control performance contour at the end of the forecast horizon, satisfying the water supply demands and finally meeting the environmental flow target. A hedging mechanism is included for reducing the level at which the environmental flow is met depending on the amount of storage relative to the target level of performance defined by a contour of the environmental flow objective state-value function. Also included in the reward is the value of the $V(s)$ function for the environmental flow objective at time steps within the forecast horizon. This value is discounted to show preference for scheduling releases later in the forecast horizon.

This function then takes the form:

$$f_t = \sum_{\tau=0}^{T-1} \gamma_f^{\tau+1}(r_{1,\tau} + r_{2,\tau} + r_{3,\tau} + r_4) + r_T \tag{5.20}$$

where τ is the step counter during the forecast horizon, γ_f is the discount factor used in the fitness function and T is the forecast horizon. The maximum storage constraint is given the highest priority through the term $r_{1,\tau}$:

$$r_{1,\tau} = c_1(s_{t+\tau+1} - s_{max})\mathbb{1}_{\{s_{t+\tau+1} > s_{max}\}} \tag{5.21}$$

where a value of $c_1 = -10^3$ is used to turn this into a penalty for exceeding the maximum storage.

The decisions are guided toward meeting the environmental flow requirements through the term $r_{2,\tau}$ which is equivalent to the reward for the environmental criterion from the reinforcement learning process:

$$r_{2,\tau} = \begin{cases} \min_k \left(\frac{q_{env} - q_{t,t+\tau+1}}{q_{env}} \right), \forall q_k < q_{env} & \text{if } \exists q_{k,t+\tau+1} < q_{env} \\ 1 & \text{otherwise} \end{cases} \tag{5.22}$$

Preference for storing water is included using the term $r_{3,\tau}$ which returns a reward based on the value of the state of the system at each step during the forecast horizon obtained from the

state-value function from the environmental flow criterion:

$$r_{3,\tau} = c_3 V_{ENV}(s_{t+\tau+1}) \tag{5.23}$$

where a value of $c_3 = 10(1-\gamma)$ is used. The term $(1-\gamma)$ scales the values of V_{ENV} to the range [0,1], and the value of 10 increases the weight given to this term.

Due to the boundary conditions set for the environmental flow objective when performing the reinforcement learning process, the state-value function can show dramatic increases near the maximum storage limit. Because of this, when a large storm event led to storage near the boundary, the value returned by the $r_{3,\tau}$ term would come to dominate all other terms, resulting in the system preferring to maintain the high storage levels rather than meet the environmental flow requirement. To avoid this, an additional term was included to set penalize storages that were within 2,000 af of the maximum storage limit:

$$r_{4,\tau} = c_4 \mathbb{1}_{\{s_{t+\tau+1} > s_{max} - 2000\}} \tag{5.24}$$

where $c_4 = -10$.

These four terms are discounted based on their distance into the forecast horizon using a discount factor $\gamma_f = 0.85$. This allows the fitness function to place more emphasis on accruing rewards earlier in the forecast horizon and with the inclusion of the $r_{2,\tau}$ term then creates the preference for storing water in the near term and making necessary releases later in the forecast horizon.

At the end of the forecast horizon, we are back to the condition of having no information beyond that of the historical record which means that the storage at the end of the forecast horizon should be at or below the contour that was determined to provide acceptable performance for the flood control objective. This is ensured by using the term r_T:

$$r_T = c_3(V_{FC,min} - V_{FC}(s_{t+T}))\mathbb{1}_{\{V_{FC}(s_{t+T}) < V_{FC,min}\}} \tag{5.25}$$

where a value of $c_3 = -10^5$ turns this term into a penalty with magnitude proportional to the distance from the flood control contour.

5.6.2 Decision-Making During the Forecast Horizon

The ensemble forecasts are treated as possible sequences of streamflows that could occur. No assumption is made about the form of a distribution that could be suggested by the collection of scenarios. This method even refrains from assuming each scenario is equiprobable, only that if the scenario is in the ensemble, then it is possible and should be accounted for. Each scenario is used intact and the entire ensemble is used. No scenario reduction or scenario tree methods are used.

Selecting the release decision begins with determining the optimal sequence of decisions considering each scenario as a deterministic forecast. Figure 5.43 shows the ensemble forecast for the inflow to Lake Mendocino for six days during the flood event that occurred in February 1986. Figure 5.44 shows the same forecasts represented as cumulative flow.

Using the fitness function defined above, the PSO determines the optimal sequence of decisions for each scenario. This process is shown in Figures 5.45 and 5.46. Figure 5.45 shows the storage for each scenario as the PSO progresses. Initially, decisions are selected randomly and as the optimization progresses, the PSO is able to find sequences of decisions which are able to ensure that the final storage is at or below the flood control contour (dashed blue line). There are multiple paths that the system could take to result in this final storage and because the fitness function includes the term $r_{II,\tau}$, preference is given to storing water in the near term for release toward the end of the forecast horizon. These decisions are shown in Figure 5.46. Again, the decisions are randomly initialized between the maximum and minimum release limits with the final decisions showing the preference for making high releases at the end of the forecast horizon.

The intent of this research has been to avoid assigning a distribution to the scenarios in an ensemble and thus each scenario is considered a possible outcome. It is because each scenario is considered a possible outcome, with no consideration of probability, and because each sequence of decisions is optimized to meet the flood control objective and the level at which the environmental flow requirement is satisfied and to do so by minimizing the amount of release in the earliest time steps, the release made from the reservoir during flood control operations is then the maximum of the first day decisions over the set of scenarios. A hedging mechanism is built into the release decision so that when the flood control criterion has been satisfied and when the storage is below the environmental flow contour, the release is reduced by the ratio of the current state-value to the environmental contour:

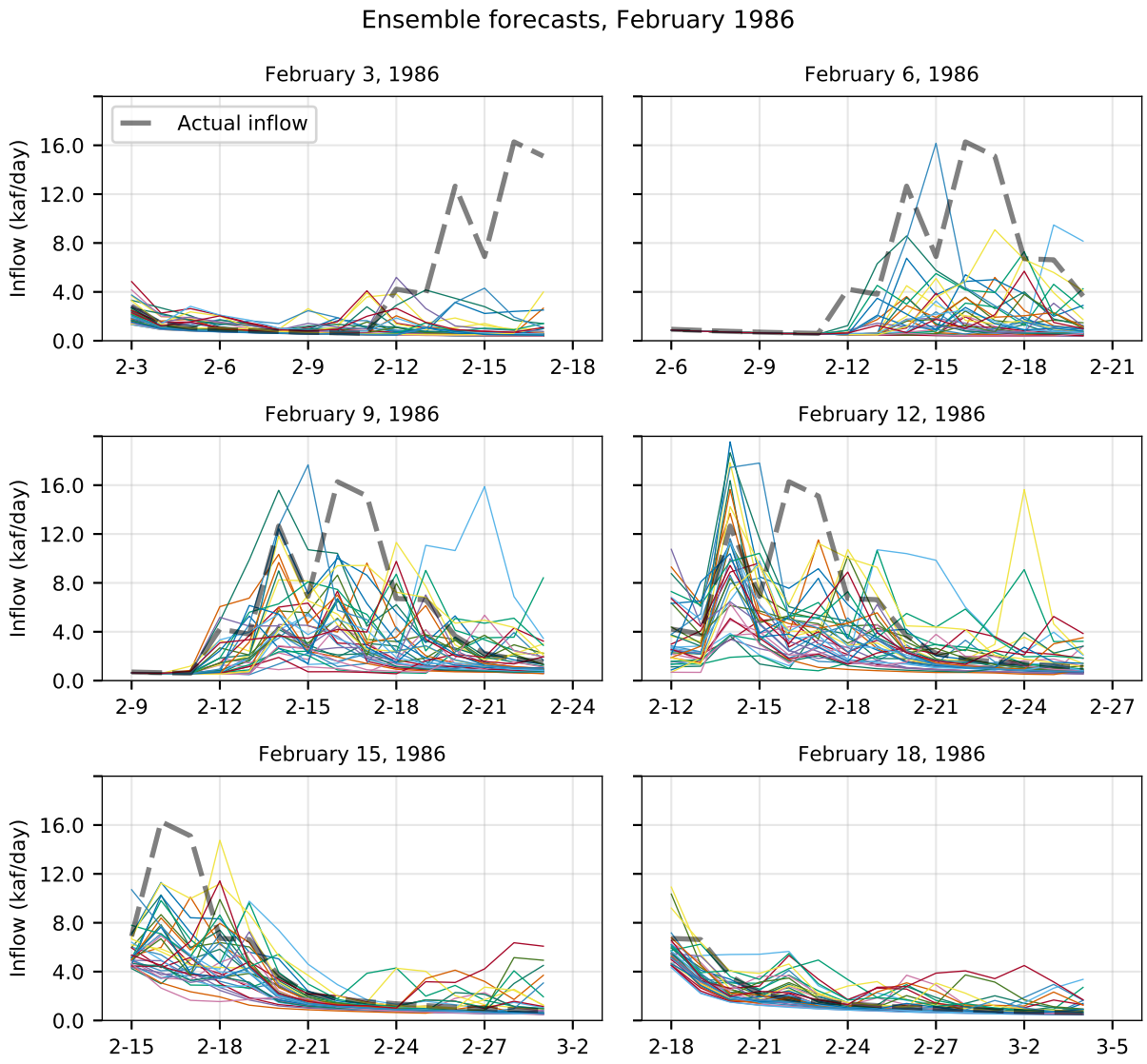

Figure 5.43: Hindcasted inflow to Lake Mendocino for six days during the flood event which reached peak flow on February 16, 1986. Each ensemble consists of 35 scenarios. The dotted line represents the actual inflow.

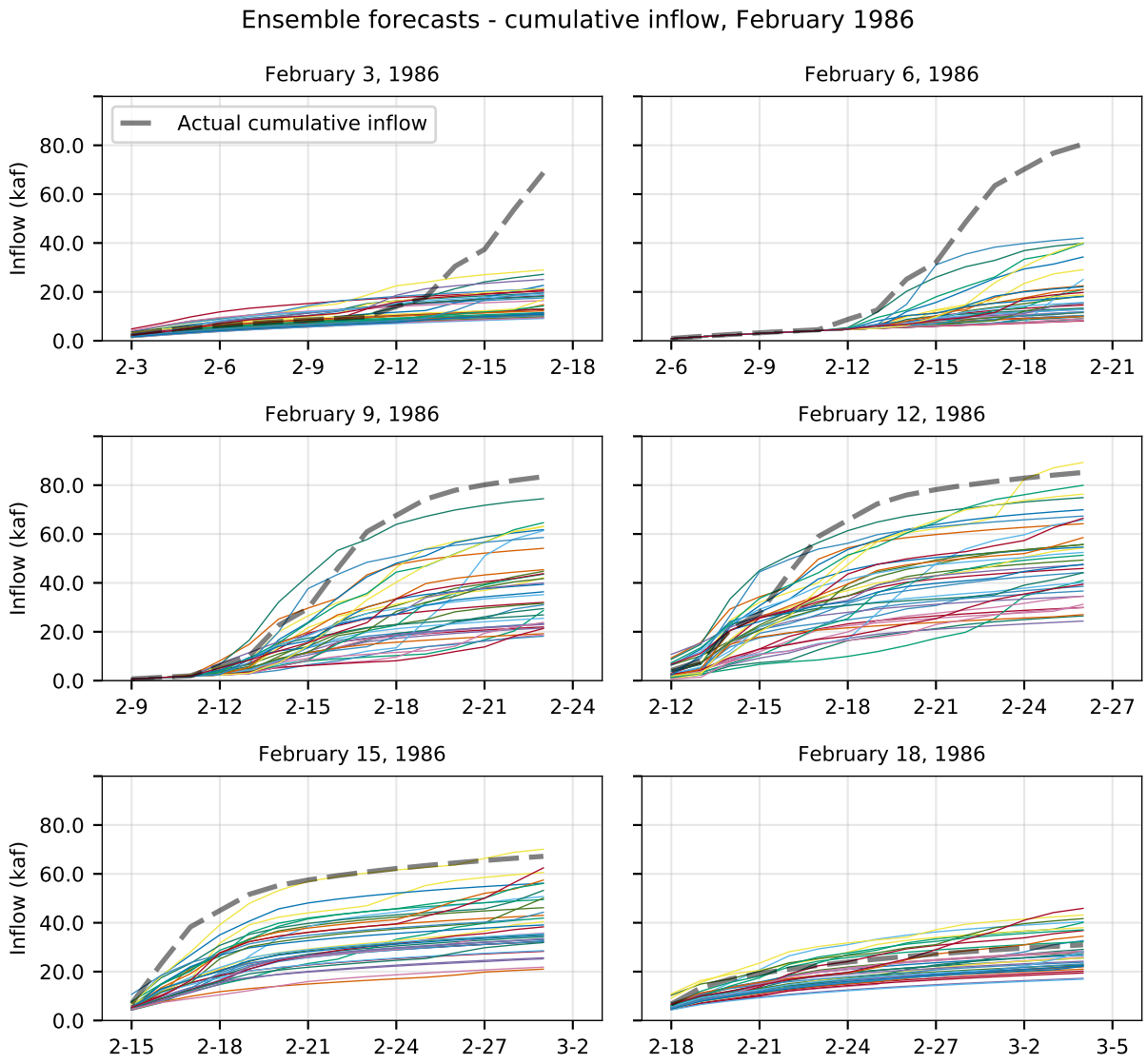

Figure 5.44: Hindcasted cumulative inflow to Lake Mendocino for six days during the flood event which reached peak flow on February 16, 1986. Each ensemble consists of 35 scenarios. The dotted line represents the actual cumulative inflow.

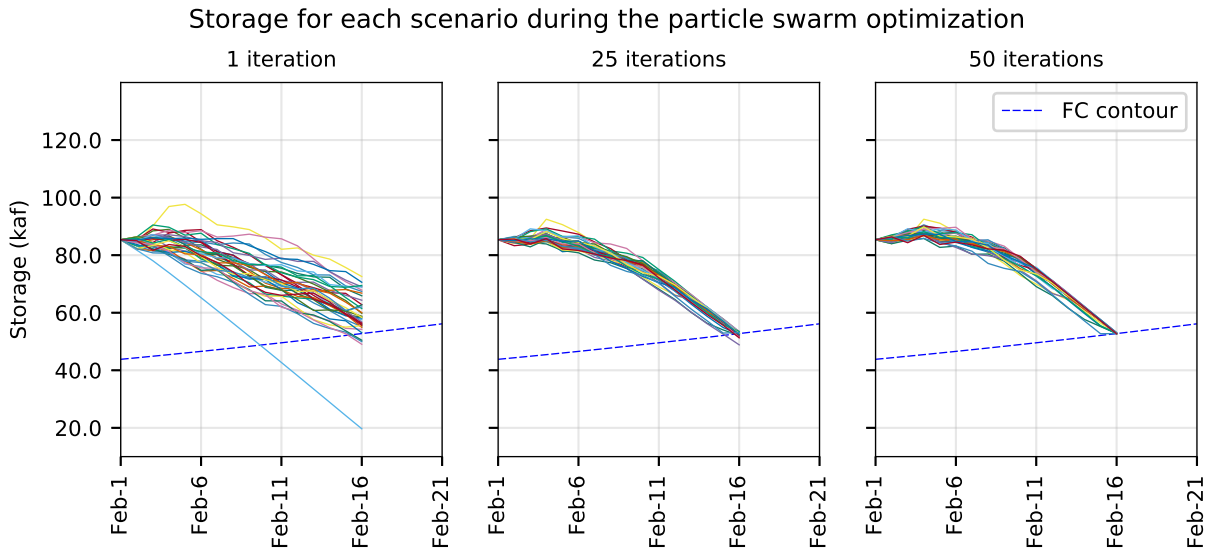

Figure 5.45: Storage plots for each scenario at three steps during the PSO optimization. Initially, decisions are selected randomly and storage levels take on a range of values at the end of the forecast horizon. As the optimization progresses, the decisions seek to meet the target storage at the end of the forecast horizon. Preference is built in to the fitness function to schedule releases later in the forecast horizon.

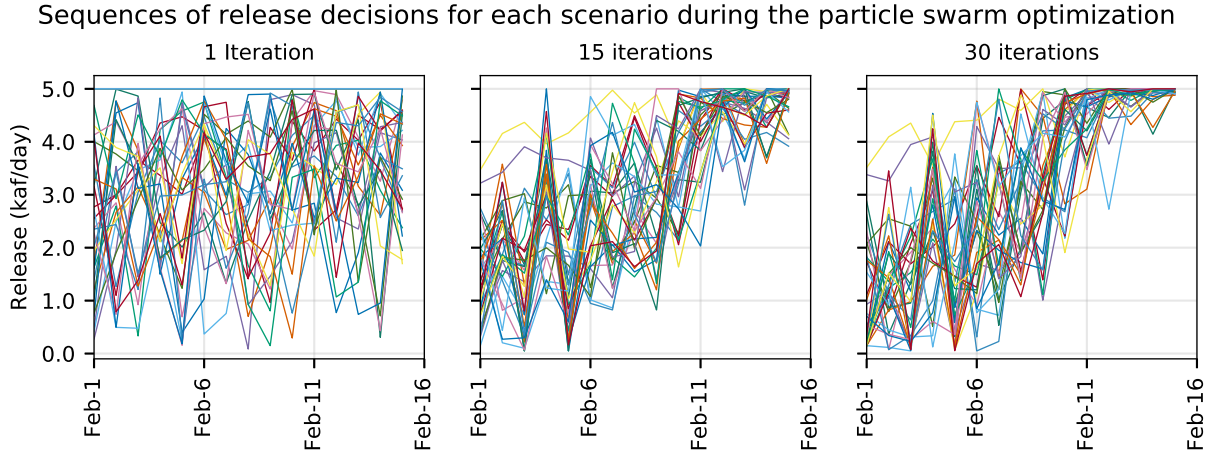

Figure 5.46: Sequences of release decisions for each scenario at three steps during the PSO optimization. Decisions are randomly initialized between the maximum and minimum allowable release. As the PSO progresses, larger releases are made toward the end of the forecast horizon.

$$a_t = \begin{cases} \max_j\{a_{j,0}\} & \text{if } V_{ENV}(s_t) \geq V_{ENV,min} \text{ or } V_{FC}(s_t) < V_{FC,min} \\ \frac{V_{ENV}(s_t)}{V_{ENV,min}} \max_j\{a_{j,0}\} & \text{otherwise.} \end{cases} \tag{5.26}$$

5.6.3 Selecting the Flood Control Contour

The flood control objective is given the highest priority and the first step is to determine which contour will provide the appropriate level of performance. The forecasts are available from 1985-2010 and this period is needed to supply both the training and testing data sets. Because the flood control objective is concerned with events that occur very rarely, the division of the available simulation period could affect the results if the large flood events are not properly represented. To limit the data set aside for training, a single flood event (February 1986) was considered and performance was evaluated for the contours -0.1, -0.08, -0.06, -0.05, -0.04 and -0.03. It is noted here that since the reward which is used in the reinforcement learning process is a binary reward based on success or failure to meet the maximum storage limit, it may be tempting to consider this value as a reliability value, or more accurately, a failure rate. This cannot be assumed as the value is based on a discounted sequence of rewards so it is not a direct computation of the probability of failure. For this analysis, the contours are simply considered levels of constant performance. Also, because of the difficulties encountered during the reinforcement learning process for the FC reward, the state-value function used in the simulation with forecasts was created from the mean of 5 of the state-value functions which did not show the distortion problem in the repeated experiments.

The state-value function represents the value associated with each state that the system could be in, absent any additional information beyond that available in the historical record. To select the contour that represents an acceptable level of performance, we should observe operations using the contour without a forecast. The implementation of this simulation in computer code is configured to require a forecast. In order to run, the simulation was given a 1-day forecast of the mean daily inflow based on the day of the year, which maintains the requirement that no additional information is included. This simulation was run for each value of the flood control contour. The results are shown in Figure 5.48. With a value of -0.1 for the flood control contour, the storage exceeds the maximum during the flood event. As the flood control contour value is increased, the maximum storage decreases. A value of -0.04 maintains the storage just below the maximum level with the 2,000 af buffer, suggesting that this is the value that corresponds to the level of performance that we wish to maintain for the reservoir during flood control operations.

To check the validity of this assumption, the simulation was run with the target flood control contour set to -0.04 and the simulation was given perfect forecasts for varying forecast horizons. The

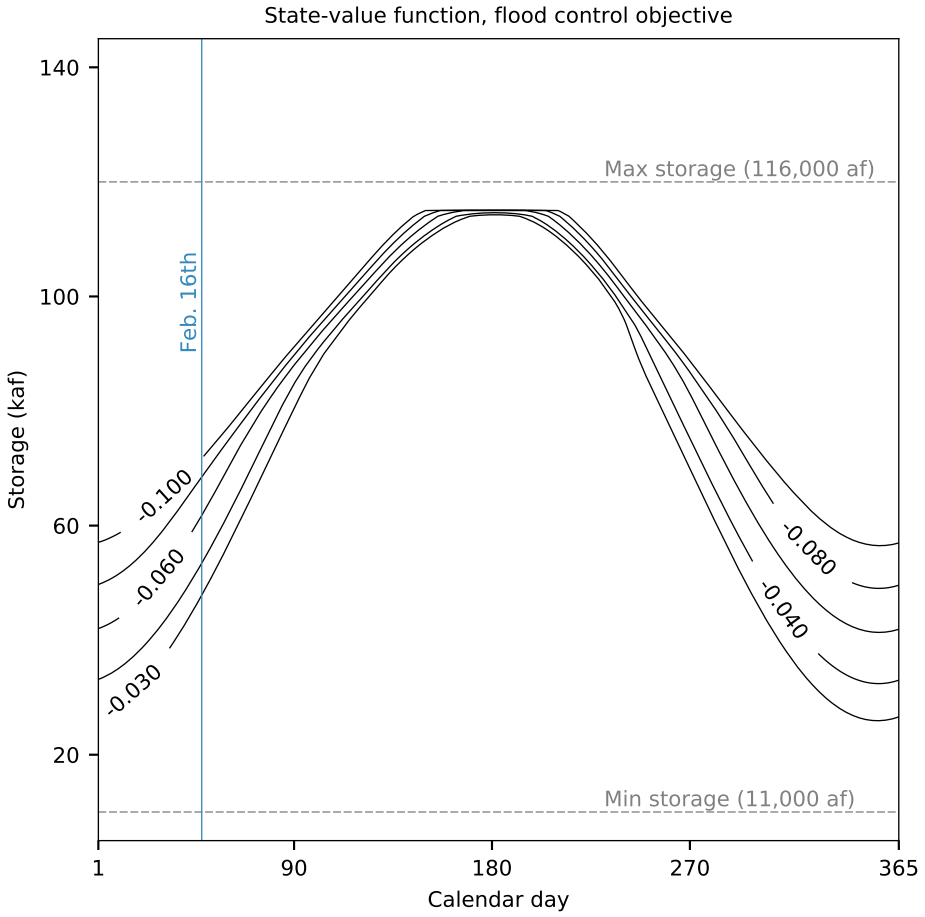

Figure 5.47: Selected contours of the flood control value function used in experiment FBO-01. This is the value function which was obtained by using a discount factor of 0.99 during the learning process. The blue vertical line is included as a reference to the day February 16, which is the day of peak flow during the flood event examined for selecting the flood control contour.

results are shown in Figure 5.49. We would expect that the inclusion of perfect forecasts should not hinder performance and that the system should be able to maintain the storage below the maximum for all forecast horizons. The results confirm that the system is able to maintain the storage below the maximum for all forecast horizons.

The simulation was then run with the ensemble forecasts for a range of forecast horizons using the same contour for the flood control objective (Figure 5.49). For each forecast horizon, the storage is maintained below the maximum storage limit with the buffer.

Another check is to examine the results when using the ensemble forecasts for several lengths of the forecast horizon for multiple values of the flood control contour. Figure 5.51 shows the resulting storage plot for each level of the flood control contour at five different forecast horizons. The storage

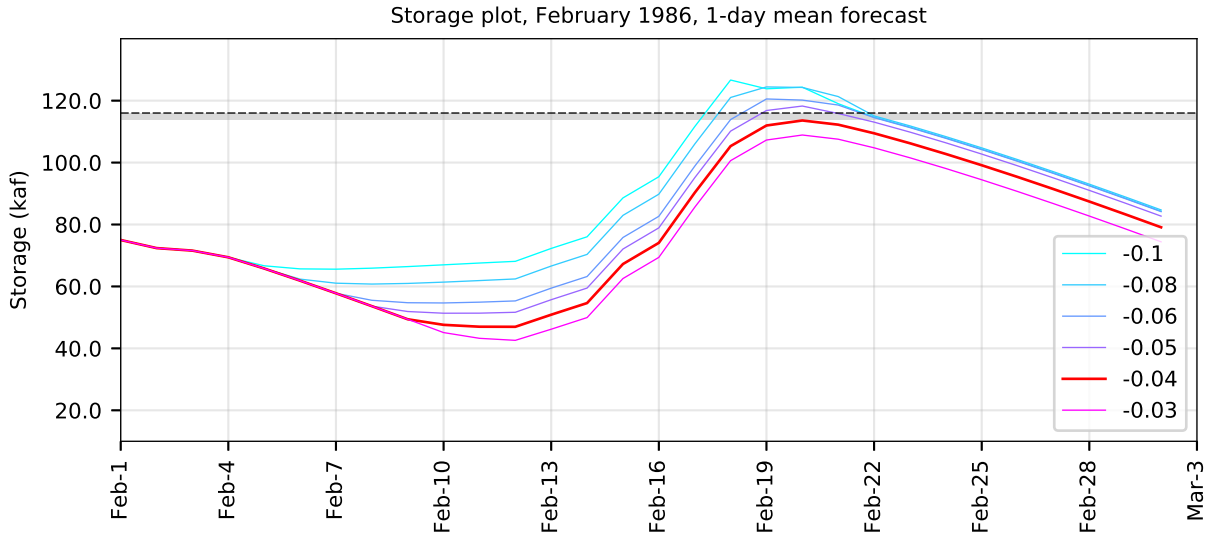

Figure 5.48: Simulation of the operation given a 1-day forecast using the historical mean daily flow for various levels of the contour of the flood control state-value function. When the value of -0.04 is used, the storage is maintained just below the maximum level during the flood event.

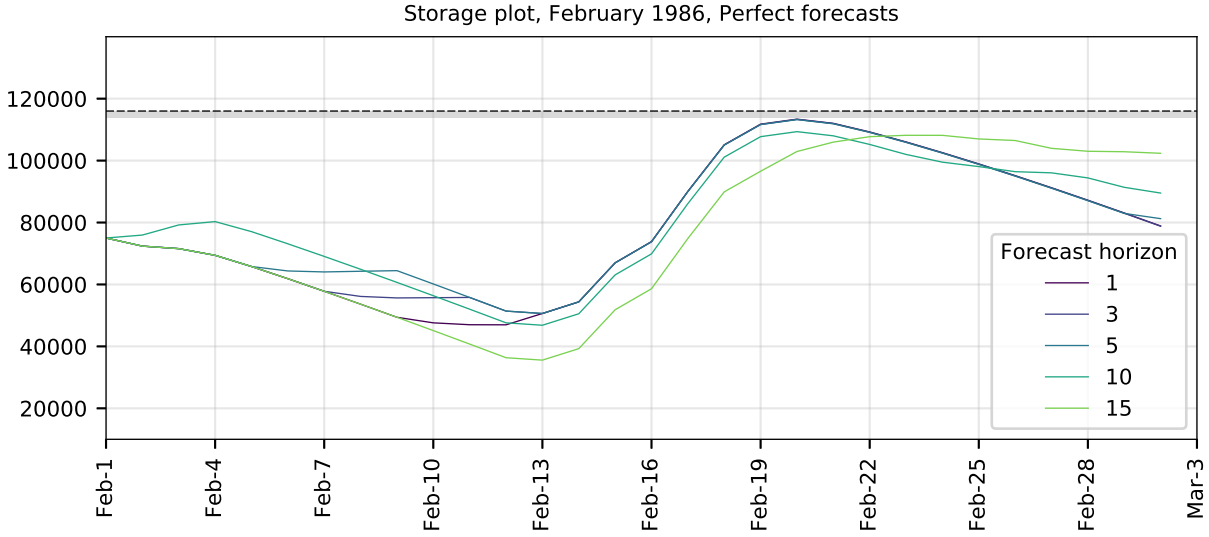

Figure 5.49: Simulation of the operation given a perfect forecast of varying forecast horizon with the value of -0.04 selected as the contour representing the desired level of performance with respect to the flood control objective.

is maintained below or very near the maximum level for all values of the flood control contour at the 10 and 15-day forecast horizons, but exceeds the maximum storage level for the higher values of the flood control contour at shorter forecast horizons. This again suggests that the -0.04 contour is valid for all forecast horizons.

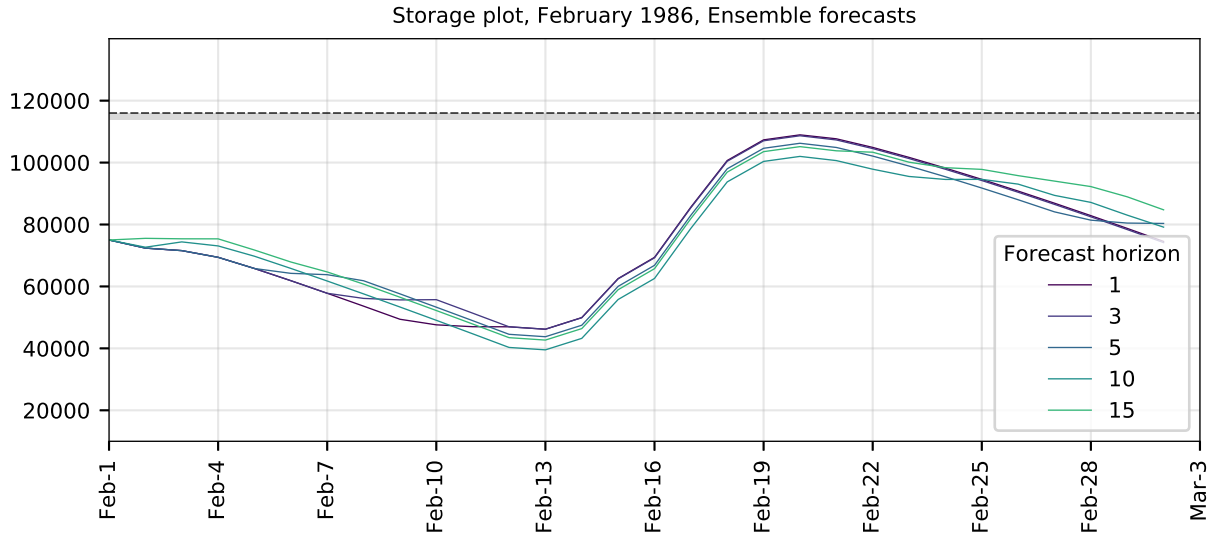

Figure 5.50: Simulation of the operation using the ensemble forecasts with different lengths of the forecast horizon. Storage is maintained below the maximum level for all forecast horizons.

5.6.4 Selecting the Environmental Flow Contour

As with the flood control objective, a level of performance with respect to the environmental flow objective must be selected to guide decision-making during the simulation with forecasts. With the flood control objective, the level of performance was selected as the contour which would avoid storage above the spillway during storm events. This is a simple binary outcome and since the flood control objective was given the highest priority, no trade-off is necessary. The environmental flow criterion is a bit more complicated as it does not follow this binary structure. Because it is given a lower priority, the performance with respect to this criterion can be degraded in order to achieve the desired level of performance with respect to the flood control criterion.

Because of this, four contours from the environmental flow objective state-value function were selected for use in the simulation. These contours are shown in Figure 5.52 for the scenario which maintains the full diversion from the PVP project and has a maximum daily release of 5,000 af. The 1000 contour was selected to represent a strategy in which there is almost no hedging and the environmental flow requirement will be met as long as there is storage in the reservoir. The 8000 contour was selected because this contour represents the lowest level of value placed on carryover storage, that is, contours below this level begin to become indifferent to carryover storage during the

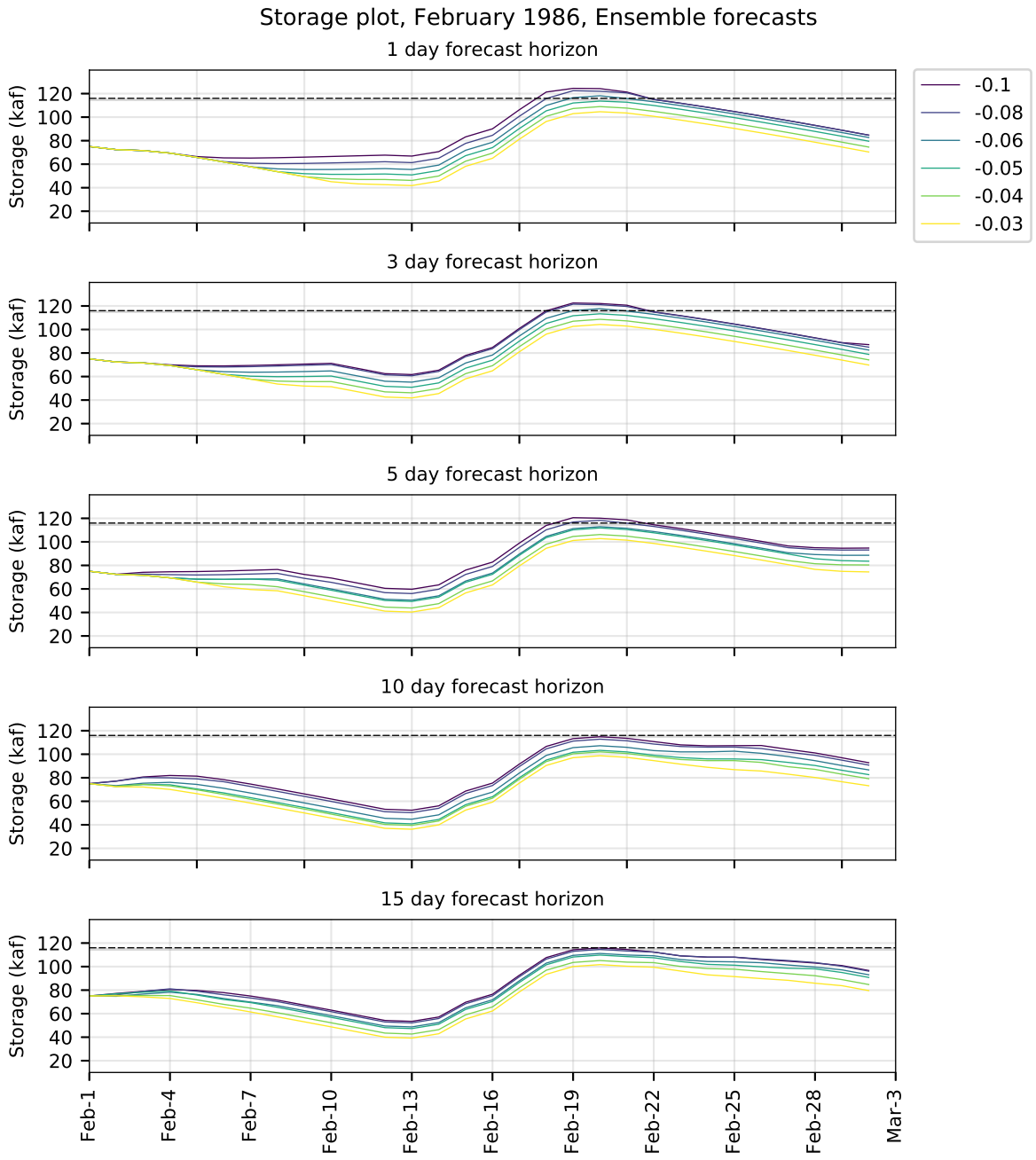

Figure 5.51: Simulation of the operation for various lengths of the forecast horizon for a range of values of the flood control contour. Storage is maintained below or near the maximum for all values of the flood control contour for the 10 and 15-day forecast horizons. With shorter forecast horizons, the maximum storage is exceeded when using the lower values of the flood control contour.

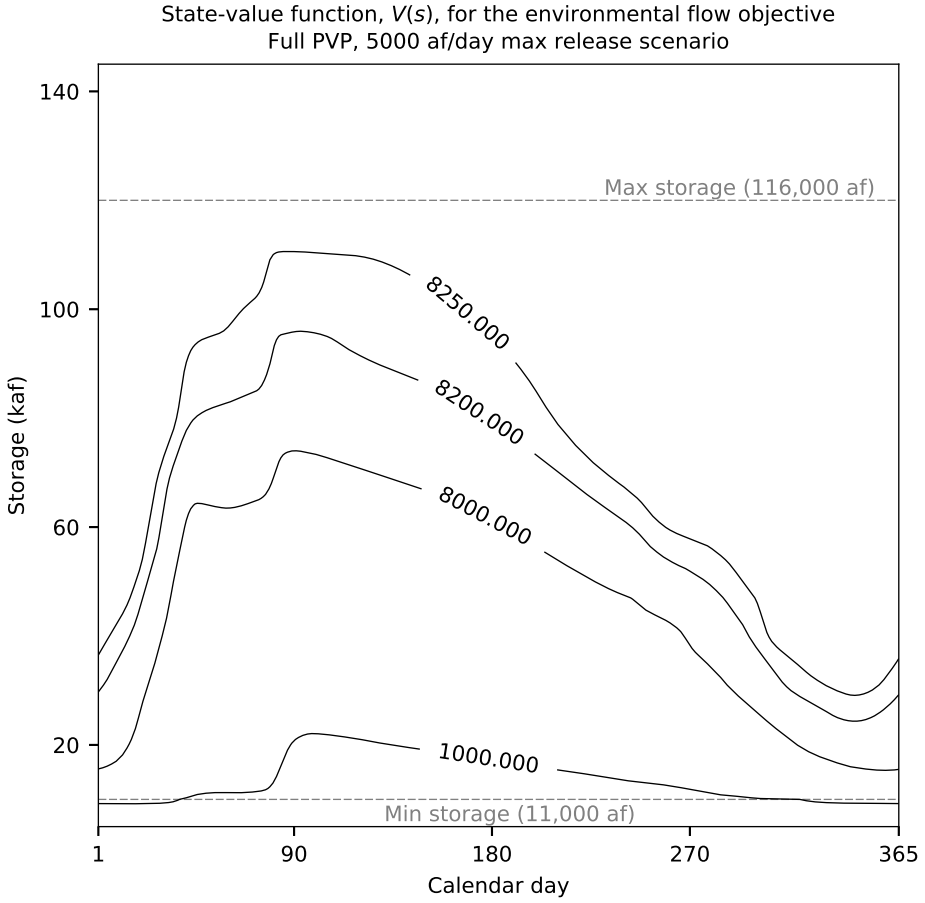

Figure 5.52: Contours of the state-value function, $V(s)$, for the environmental flow criterion for the "Full PVP" scenario with a maximum release of 5000 af/day.

winter. The contour 8250 was selected because it extends nearly to the maximum storage level. The 8200 contour was selected as a middle level of performance between the 8000 and 8250 contours.

The simulation was run for the training period 1985-1989. Figure 5.53 shows the storage plots for each contour when perfect forecasts are used with a 15-day forecast horizon. Figure 5.54 shows the resulting storage when the hindcasts are used with a 15-day forecast horizon.

5.6.5 Issue with the Accuracy of Hindcasts During the Irrigation Season

The performance of the operating strategy was examined by plotting the deficit in the environmental flow at Healdsburg. When using the perfect forecasts, the environmental flow is met at all time steps. When using the hindcasts, deficits appear at multiple times during the training period (Figure 5.55).

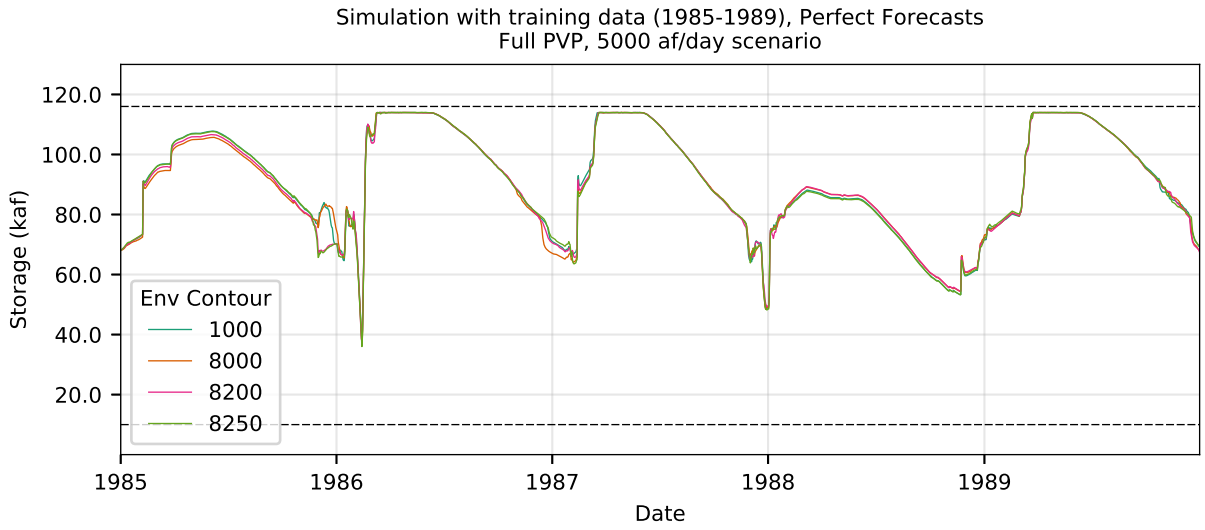

Figure 5.53: Storage plot from the simulation during the training period using perfect forecasts with a 15-day forecast horizon for the Full PVP, 5000 af/day scenario.

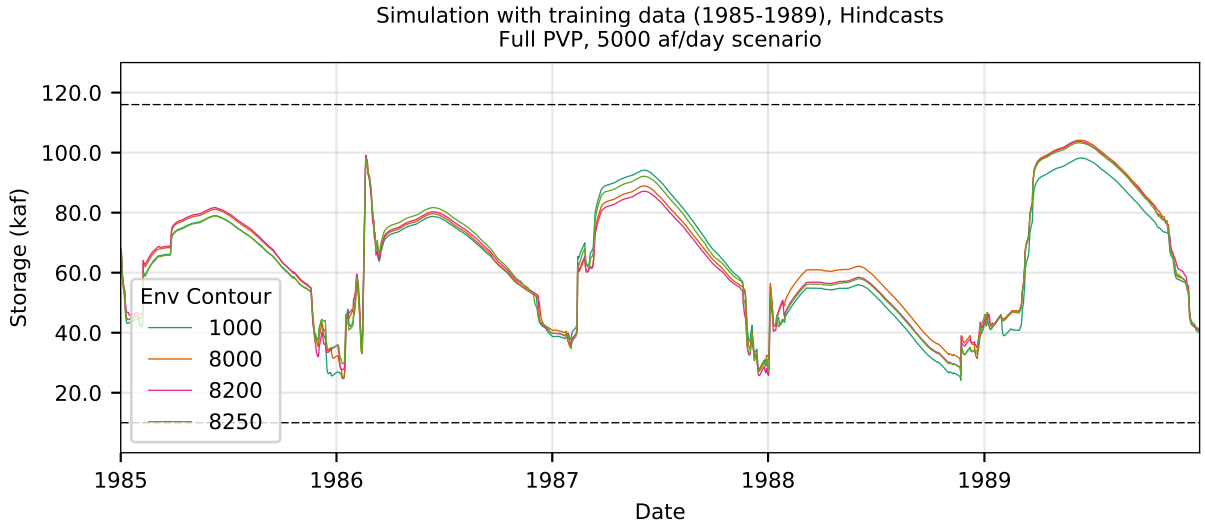

Figure 5.54: Storage plot from the simulation during the training period using hindcasts with a 15-day forecast horizon for the Full PVP, 5000 af/day scenario.

The results show failure to meet the environmental flow requirement at times during the training period when the corresponding storage would suggest that sufficient water is available to meet this demand. Figure 5.56 shows a comparison of the storage plot to the 1000 contour. The storage is above the contour at all times during the training period, however frequent deficits do appear, and appear more often during the irrigation season. It was suspected that the accuracy of the hindcasts

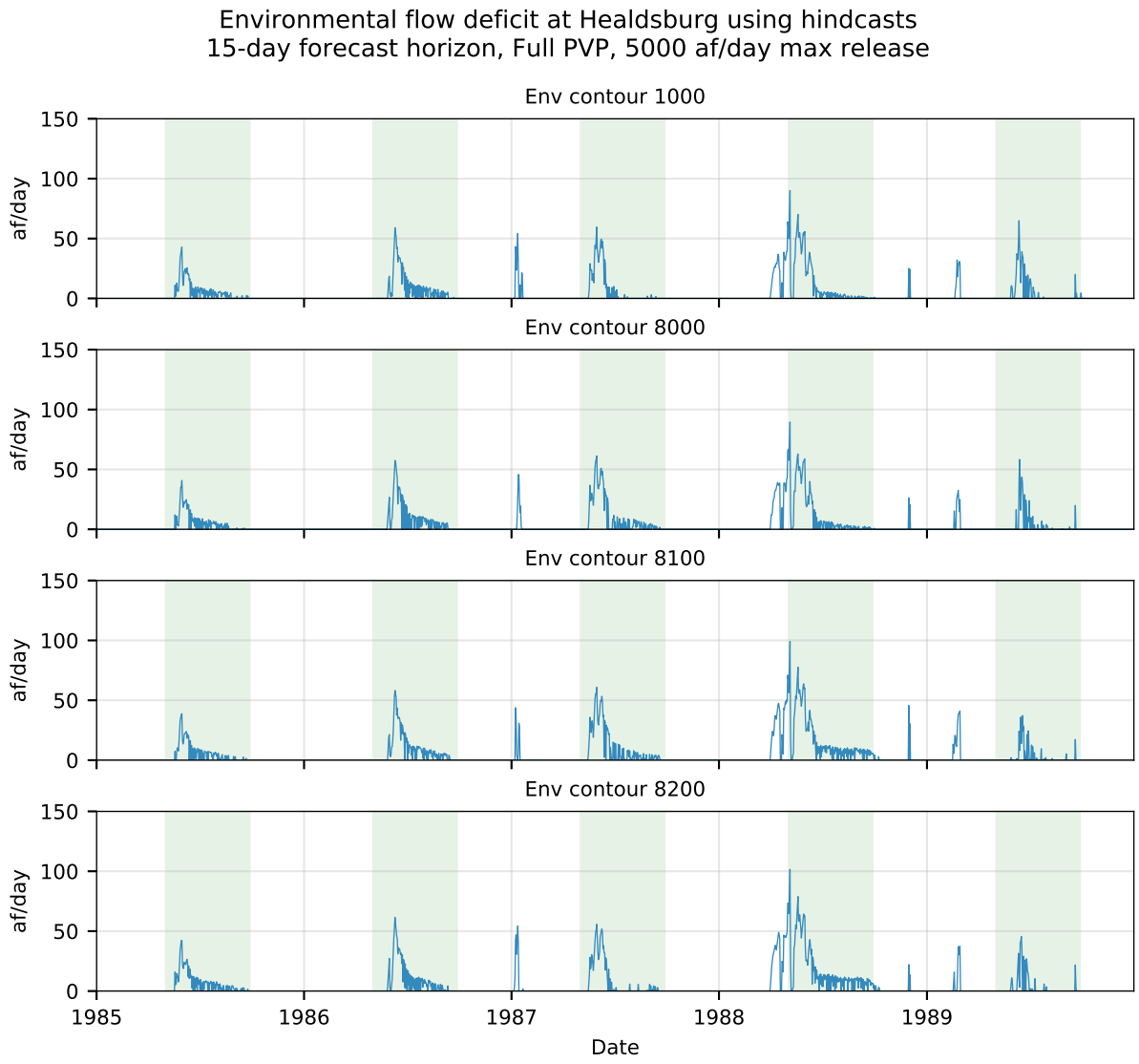

Figure 5.55: Deficits in the environmental flow at Healdsburg for each value of the environmental flow contour with a 15-day forecast horizon for the Full PVP, 5000 af/day max release scenario. The shaded regions indicate the irrigation period May-September.

may be the cause, as the hindcasts frequently over-predict the volume of flow into the network particularly during the irrigation season.

To examine this, the simulation was run again with the hindcasts replaced with perfect forecasts during the period May 1 to September 30 of each year. Figure 5.57 shows the deficits at Healdsburg for each level of the environmental flow contour. The deficits during the irrigation season are greatly reduced. Also, the irrigation season during 1988 does show deficits which increase as the environmental flow contour value is increased suggesting that using the perfect forecasts during the

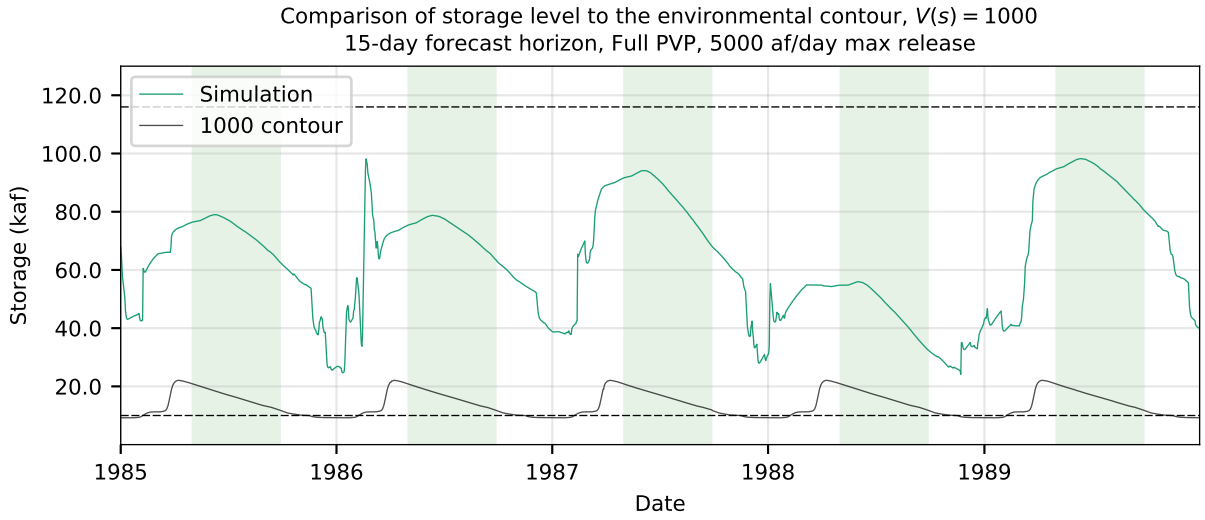

Figure 5.56: Storage levels for the simulation during the training period using a value of $V(s) = 1000$ for the environmental contour with a 15-day forecast horizon for the Full PVP, 5000 af/day max release scenario.

irrigation season did not somehow impose a strict constraint on meeting the environmental flow limit. This scenario of using perfect forecasts during the irrigation season is carried through the rest of the analysis. Although in this system the water managers do not have information about the instantaneous diversions, this scenario is comparable to operation in practice as the flow in the river is monitored and releases adjusted to meet target flow rates. The other periods of deficit such as in January of 1987 also appear to result from this same issue as they are correlated with periods of low flow and overprediction. No additional correction was introduced for this situation when it occurs outside of the irrigation season.

In this simulation, the operation did not require that the release be reduced due to storage below the environmental flow contour during four out of the five irrigation seasons in the training time period. During the irrigation season in 1989, the storage is low enough to require the system to reduce the release when a value of 8200 or 8250 is used for the environmental flow contour. Actual selection of the level of reliability desired for the environmental flow criterion could consider a more thorough analysis including feedback from the stakeholders. For this study, all four contours were simulated during the testing period and compared.

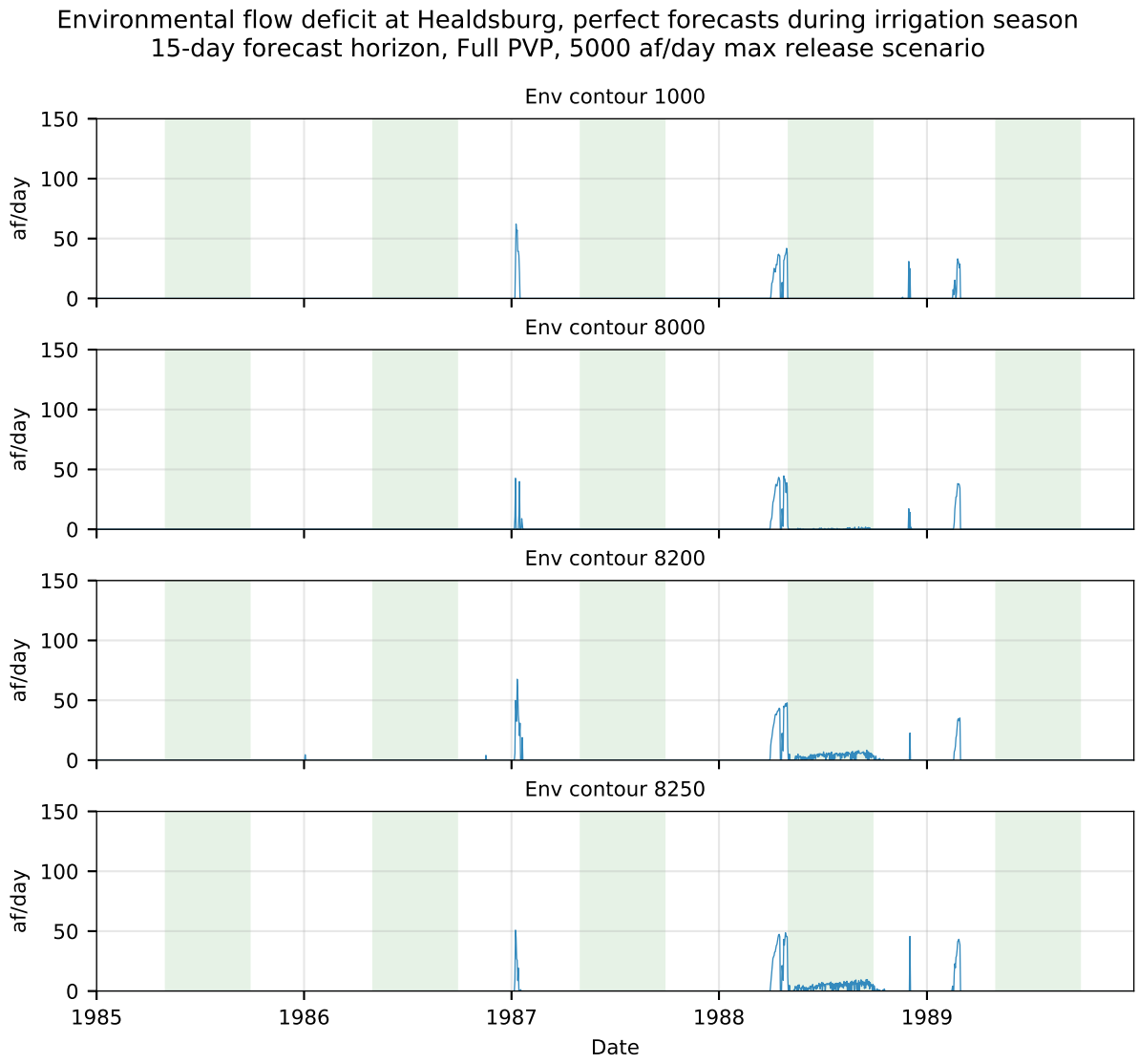

Figure 5.57: Deficits in the environmental flow at Healdsburg for each contour when hindcasts are replaced with perfect forecasts during the irrigation season. This simulation used a 15-day forecast horizon for the Full PVP, 5000 af/day max release scenario.

Discussion of the scenarios examined

The performance of the forecast based operating strategy was examined by varying several conditions and these are introduced here as a preface to the following discussion of the results. The first condition is the amount of diversion into the basin from the Potter Valley Project. The level of diversion from this project in the future is uncertain and so two different levels of flow are considered. The "Full PVP" condition uses the full amount of diversion from the PVP based on the time-series data obtained from the HEC-ResSim model. The "Zero PVP" condition reduces the diversion from

the PVP to zero at all time steps. The state-value functions used in the simulation represent those developed in the reinforcement learning process using the respective level of diversion. That is, in the simulations, the functions used to represent $V_{FC}(s)$ and $V_{ENV}(s)$ were obtained through a reinforcement learning process which used the same level of diversion for the PVP. In both cases, the state-value function for the flood control criterion which was developed using a discount factor of $\gamma = 0.99$ was selected, and the state-value function obtained with a discount factor of $\gamma = 0.9999$ was selected for the environmental flow criterion.

The second condition is the maximum release which can be made from the reservoir during a single time step. A value of 5000 af/day was used in the reinforcement learning portion of the method to obtain the state-value functions for each criterion. This same value is used in the simulation step and an additional option of 12,694 af/day (6400 cfs) is included which is the current maximum release according to the water control manual. When the maximum release from the reservoir is changed to the larger value, the simulation uses the same state-value functions developed with the 5000 af/day maximum. Further investigation could perform the learning process again with the higher value of the maximum release. It is included here to examine the effect of the maximum release on the ability of the system to store water prior to a forecasted storm event.

The third condition is the level of performance used as the contour for the environmental flow criterion. When the "Full PVP" condition is used for the level of PVP diversion, three contours are studied with values of 1000, 8000, 8200, and 8250. When the "Zero PVP" condition is used, the levels are 100, 800, 1000, 1100.

The system was given four different sets of forecasts. The first is the perfect forecast where each scenario is replaced with the actual inflow for the forecast horizon (PF). The second is the hindcasts as they currently exist (HC). The third is the hindcasts with perfect forecasts during the irrigation season (IP) as discussed in Section 5.6.5. In the last set of forecasts, the mean of the flow over all scenarios for each day of the forecast is computed and each scenario is replaced by this sequence of values. This is referred to as the "Mean of Ensemble" scenario (ME). This set of forecasts continues to use the perfect forecasts during the irrigation season. Each of the combinations of these conditions was tested with a 1, 3, 5, 10 and 15-day forecast horizon.

5.6.6 Simulation With Test Data

The operating strategy was tested on the period January 1, 1990 to September 30, 2010. This discussion of the results will begin with the simulations which used a 15-day forecast horizon and a value of 5000 af/day for the maximum release. The results for the scenarios which use the "Full PVP" diversion are presented, followed by the "Zero PVP" condition to examine the effect of water supply stress on the performance. A similar analysis is presented for the scenarios which use a maximum release of 12,694 af/day. The results of the simulations are discussed further with the use of parallel coordinate plots to examine the effect of varying the max release, forecast horizon and environmental contour.

Results - Max release 5000 af/day, 15-day forecast horizon

The resulting storage plot for the "Full PVP" scenario with a maximum release of 5,000 af/day using the 8250 environmental contour with a 15-day forecast horizon is shown in Figure 5.58 and the corresponding deficit in the environmental flow at the Healdsburg node is shown in Figure 5.59. Similar to the training period, when using the perfect forecasts, the storage frequently reaches the maximum storage level during the wet season and the environmental flow target is met at nearly all time steps. When the hindcasts are used, the storage is lower and the deficits appear frequently, though at all times some amount of environmental flow is provided. The storage level does reach the minimum storage limit, though this occurs during the wet season and is due to the release of storage in preparation for a forecasted storm event. The shortages that do occur happen as a result of the system either hedging the release when the storage is below the environmental contour or when low flows combined with overprediction lead to release volumes which are insufficient to meet demand.

When the hindcasts are replaced with perfect forecasts during the irrigation season, the deficits are reduced. The storage is generally higher and the environmental flow is met more frequently when the mean of the ensembles is used as the forecast, however, this leads to occasions where the maximum storage limit is exceeded. With the guide curve operations, the environmental flow demand is met at all time steps and storage is generally higher than when using the hindcasts. The maximum storage limit is exceeded during two storm events during this period.

Table 5.3 shows the performance with respect to each measure for the maximum storage, water supply and the environmental flow criteria for the "Full PVP" scenario with the maximum release

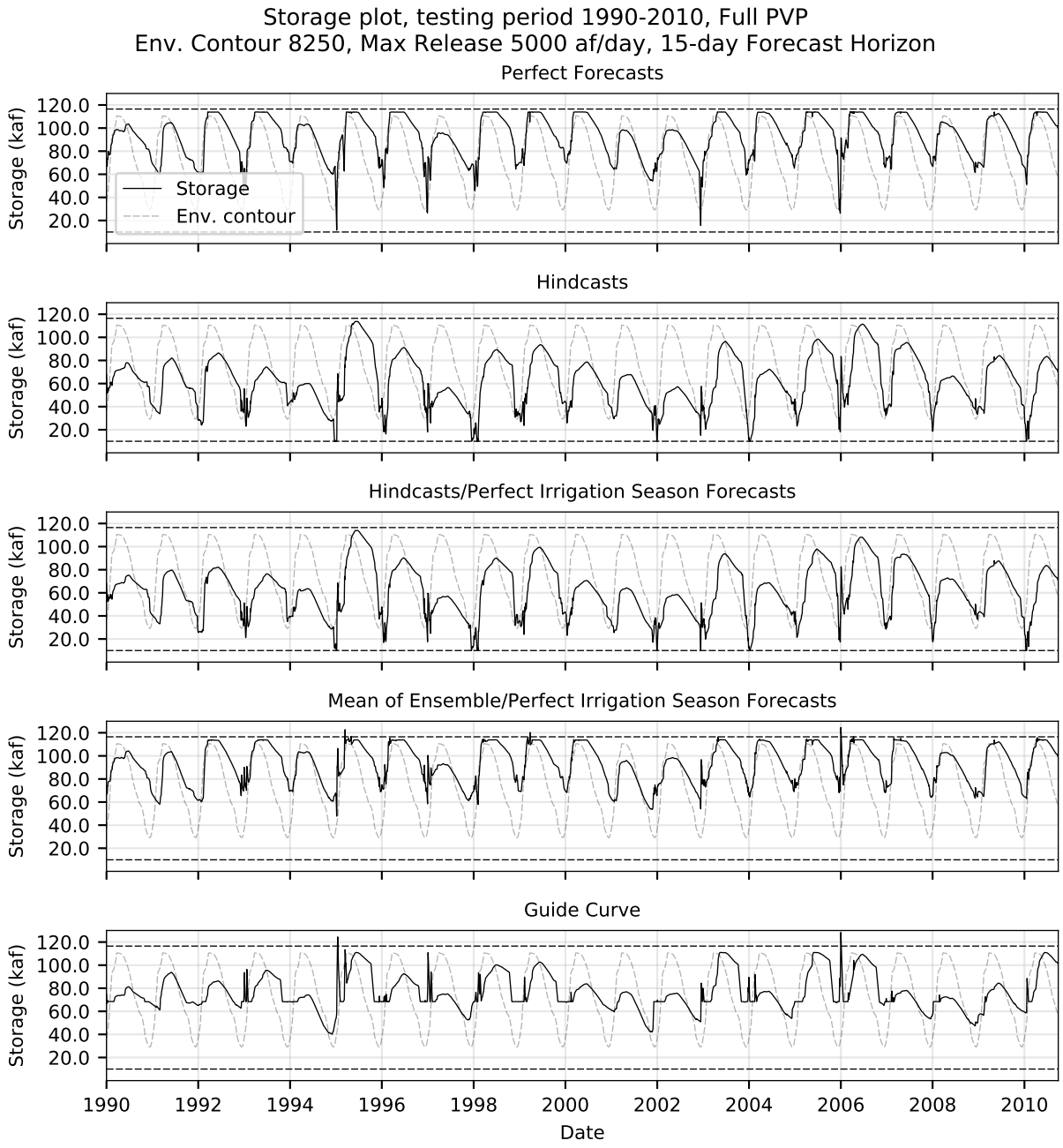

Figure 5.58: Storage during the testing period 1990-2010 for each set of forecasts using a 15-day forecast horizon for the Full PVP, 5000 af/day max release scenario.

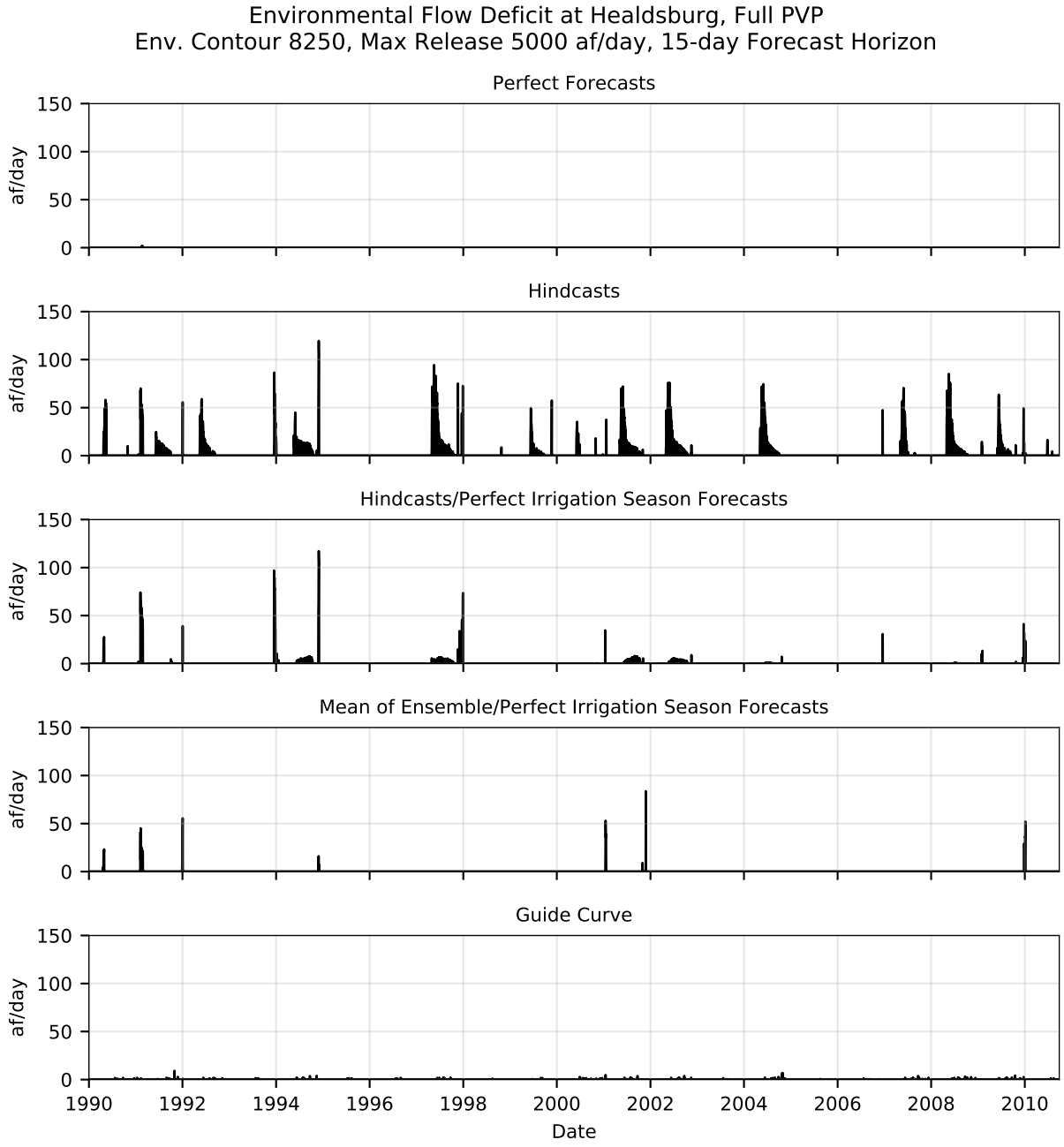

Figure 5.59: Deficit in the environmental flow target at the Healdsburg node during the testing period 1990-2000 for each set of forecasts using a 15-day forecast horizon for the Full PVP, 5000 af/day max release scenario.

set to 5,000 af/day. The number of days when the flow at Hopland exceeds 8,000 cfs (Hopland high flow) is also listed. Each scenario was simulated using perfect forecasts (PF), the hindcasts (HC), the hindcasts with perfect forecasts during the irrigation season (IP) and the mean of the ensemble (ME). The performance is also shown for the simulation in the HEC-ResSim model using the current guide curve. In all cases, the water supply demands are met throughout the testing period. The maximum storage criterion is met in all cases when forecasts are used except for the case where the mean of the ensemble (ME) is used, in which case the reliability drops to 0.9983. The resiliency is 1.86 meaning that the average length of the duration of exceeding the maximum storage is 1.86 days, and the vulnerability ranges from 2800 to 3200, which is an average exceedance of 2800 to 3200 af per time step in which a failure occurs. The guide curve operation has a slightly higher reliability, but with higher resiliency and vulnerability measures as well.

The environmental flow requirement was met at all or nearly all time steps for the guide curve operations and the simulation with perfect forecasts for all values of the environmental contour. Using the hindcasts (HC), the reliability is near 0.92 and with perfect forecasts during the irrigation season (IP), the reliability is near 0.98. In both cases the vulnerability ranges from 36-41 af compared to the vulnerability of 148 af, or very close to the full flow, when the guide curve is used. Also, the operations using forecasts show far fewer days where Hopland experiences high flows, a range of 25-27 days, compared to the guide curve operations which results in 48 days of high flow.

When the flow from the PVP diversion is reduced to zero, the guide curve operations show more dramatic deterioration in performance (Table 5.4). The measures with respect to flood control are similar with a slightly higher reliability which could be attributed to the generally lower storage levels resulting from the lower inflow to the system. The reliability for the environmental flow objective drops to 0.96 with a resiliency of 19.52 and vulnerability of 132, suggesting that failures that do occur are for extended periods of time and for nearly the full amount of the environmental flow. With the inflow from the PVP diversion eliminated, the water supply objective is no longer met at all time steps, with the reliability dropping to 0.989. The resiliency is 13.4 and the vulnerability is 34. Periods of failure with respect to the water supply objective average shorter duration and lower magnitude than the failure with respect to the environmental flow, though the magnitude of the water supply demand should not be directly compared to the environmental flow demand as the

Table 5.3: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 5000 af/day using a 15-day forecast horizon.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	38.84	26
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93430	5.48	36.07	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98815	3.29	36.72	27
ME	0.9983	1.86	2800.04	1.000	0.00	0.00	0.99300	2.10	31.90	29
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	1.00000	0.00	0.00	26
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93400	5.49	36.17	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98501	3.31	37.34	26
ME	0.9983	1.86	3001.21	1.000	0.00	0.00	0.99369	1.95	29.43	27
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	1.00000	0.00	0.00	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92934	5.04	36.87	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98521	3.37	38.84	26
ME	0.9983	1.86	3055.95	1.000	0.00	0.00	0.99373	2.11	29.70	27
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	1.00000	0.00	0.00	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92459	4.80	36.49	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98455	3.44	41.26	26
ME	0.9983	1.86	3196.47	1.000	0.00	0.00	0.99280	2.63	29.31	28
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

daily demand at each node varies as opposed to the environmental flow demand which is constant throughout the year.

The simulations using hindcasts (HC) and perfect irrigation season forecasts (IP) both meet the maximum storage limit at all time steps. When the mean of the ensemble is used (ME), the reliability with respect to this criterion decreases as the system favors higher storages. An odd result is the failure of the perfect forecasts (PF) to meet the maximum storage limit at all time steps. The cause of this is unknown. It is believed to be a problem in the implementation and not in the method in general.

Using the perfect forecasts (PF), the system is able to meet the environmental flow target with high reliability. This is slightly lower when the mean of the ensemble (ME) is used followed by the

Table 5.4: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 5000 af/day using a 15-day forecast horizon.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9980	1.00	268.71	1.000	0.00	0.00	1.00000	0.00	0.00	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92835	5.58	36.09	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98554	3.20	38.30	26
ME	0.9960	1.30	1843.72	1.000	0.00	0.00	0.99634	1.68	30.54	29
Environmental contour 800										
PF	0.9978	1.00	299.57	1.000	0.00	0.00	1.00000	0.00	0.00	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92271	6.50	39.80	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97719	3.86	37.40	26
ME	0.9963	1.33	2136.33	1.000	0.00	0.00	0.99468	2.82	32.24	29
Environmental contour 1000										
PF	0.9976	1.06	348.42	1.000	0.00	0.00	0.99927	1.16	16.90	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91056	5.94	41.64	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95860	4.03	32.07	27
ME	0.9963	1.22	2069.95	1.000	0.00	0.00	0.99412	2.70	29.97	29
Environmental contour 1100										
PF	0.9982	1.00	417.54	1.000	0.00	0.00	0.99917	1.14	21.82	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90508	6.25	42.32	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94797	3.91	30.53	26
ME	0.9966	1.30	2155.13	1.000	0.00	0.00	0.99462	2.23	32.19	28
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

perfect irrigation forecasts (IP) and the year-round hindcasts (HC). The resiliency and vulnerability both decrease when changing from the year-round hindcasts (HC) to the perfect irrigation season forecasts (IP) and again when using the mean of the ensemble (ME).

As in the "Full PVP" scenario, the water supply demands are met at all times when using forecasts. The number of time steps where the flow at the Hopland node exceeds 8,000 cfs is in the range of 25-29 when using forecasts compared to 43 when using the guide curve.

Results - Max release 12,694 af/day, 15-day forecast horizon

Tables 5.5 and 5.6 show the performance measures for the "Full PVP" and "Zero PVP" scenarios when the maximum allowable daily release is increasing to 12,692 af/day (6,400 cfs). In both cases,

the performance of the guide curve operations was not affected by the maximum allowable release. These results show an improvement in most of the performance measures when the maximum allowable release is increased. It is believed that this is because with a high allowable release, the system is able to delay the release of water prior to forecasted storm event which allows time for the uncertainty in the future flows to decrease. Additional tables of performance measures for the other forecast horizons are included in Appendix B.

Table 5.5: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 12,694 af/day using a 15-day forecast horizon.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	0.9995	1.00	1264.35	1.000	0.00	0.00	0.99997	1.00	148.76	32
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.95057	3.89	31.55	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.99142	2.30	34.36	26
ME	0.9951	1.61	4725.01	1.000	0.00	0.00	0.99363	1.93	38.03	31
Environmental contour 8000										
PF	0.9996	1.00	1108.88	1.000	0.00	0.00	0.99993	1.00	27.03	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.94932	4.17	32.48	31
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98983	2.57	40.67	28
ME	0.9946	1.86	5111.66	1.000	0.00	0.00	0.99455	1.68	42.65	34
Environmental contour 8200										
PF	0.9996	1.00	859.82	1.000	0.00	0.00	0.99993	1.00	37.79	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.95038	4.11	33.02	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.99072	2.51	39.30	29
ME	0.9943	1.65	4895.32	1.000	0.00	0.00	0.99356	1.99	30.80	33
Environmental contour 8250										
PF	0.9997	1.00	505.01	1.000	0.00	0.00	0.99993	1.00	35.84	32
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.94291	3.89	33.48	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98907	2.57	38.30	31
ME	0.9941	1.73	4576.42	1.000	0.00	0.00	0.99369	1.87	35.62	33
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

5.6.7 Parallel Coordinate Plots

We can examine these results further using parallel coordinate plots of selected groups of simulation results. In these figures, the reliability measure is replaced with the failure rate, $P_f = 1 - P_{rel}$, so that on all coordinates, lower values are preferred. Because the water supply

Table 5.6: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 12,694 af/day using a 15-day forecast horizon.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9956	1.00	404.91	1.000	0.00	0.00	1.00000	0.00	0.00	33
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.94318	4.49	33.00	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.99013	2.65	31.56	29
ME	0.9927	1.31	3180.04	1.000	0.00	0.00	0.99323	2.30	36.66	34
Environmental contour 800										
PF	0.9964	1.00	406.32	1.000	0.00	0.00	1.00000	0.00	0.00	31
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.94024	5.08	37.09	32
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98940	2.59	29.96	28
ME	0.9937	1.23	2964.28	1.000	0.00	0.00	0.99518	1.90	43.19	31
Environmental contour 1000										
PF	0.9959	1.00	377.96	1.000	0.00	0.00	0.99997	1.00	17.52	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93793	5.12	37.80	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98072	2.86	27.93	29
ME	0.9933	1.21	2787.84	1.000	0.00	0.00	0.99389	1.97	38.29	31
Environmental contour 1100										
PF	0.9964	1.00	379.96	1.000	0.00	0.00	0.99980	1.00	16.64	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93004	5.84	39.74	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97610	3.19	27.33	30
ME	0.9934	1.22	3056.64	1.000	0.00	0.00	0.99426	2.85	42.28	32
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

demand is frequently fully met, the performance measures for the water supply objective are omitted in the following plots.

Effect of the Length of the Forecast Horizon

The forecast horizon is an important parameter and we can look at the effect of this parameter on the performance. Consider first the "Full PVP" scenario with a maximum daily release of 5000 af/day. Figure 5.60 shows the performance measures of all of the simulations with the "Full PVP" diversion and with a maximum daily release of 5,000 af/day. There are a total of 81 simulations represented, one for each combination of 5 forecast horizons (15, 10, 5, 3, 1) , 4 environmental contours (8250, 8200, 8000, 1000), and 4 sets of forecasts (PF, HC, IP, ME), plus an additional

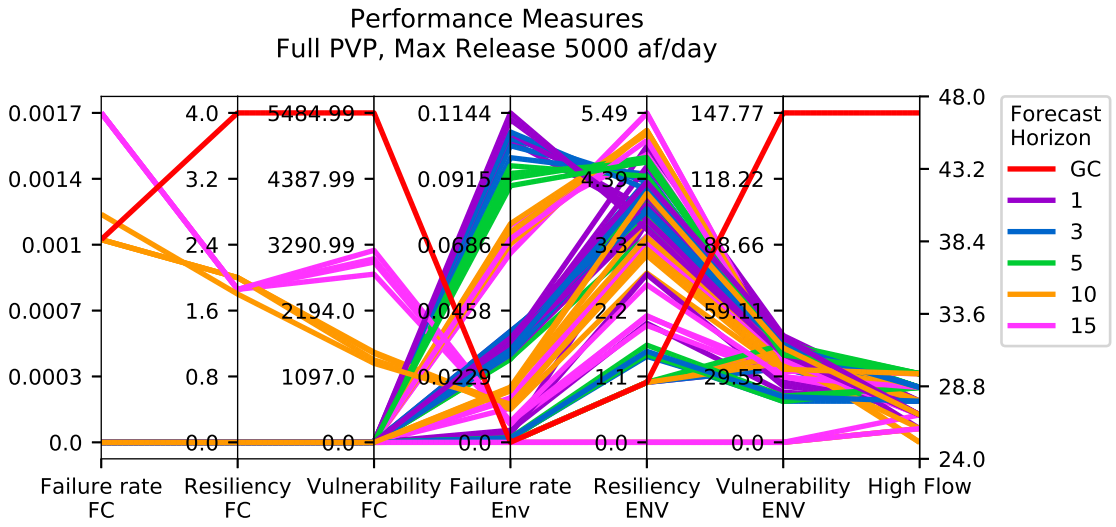

Figure 5.60: Parallel coordinate plot of the simulations with the "Full PVP" diversion with a maximum daily release of 5,000 af/day. There are 81 simulations represented, one for each combination of 5 forecast horizons (15, 10, 5, 3, 1) , 4 environmental contours (8250, 8200, 8000, 1000), and 4 sets of forecasts (PF, HC, IP, ME), plus an additional simulation using the guide curve operations (GC). Axes are scaled to the maximum and minimum values of each performance measure over this set of simulations.

simulation using the guide curve operations (GC). Each axis is scaled to the maximum and minimum values for each performance measure for this set of simulations. The rest of this section will examine subsets of this collection of simulations in more detail.

Figure 5.61 highlights the set of simulations which used the hindcasts/perfect irrigation (IP) forecasts. This includes all four values for the environmental contour and each is colored according to the length of the forecast horizon. The set of simulations which use forecasts meet the maximum storage limit at all times during the simulation, compared to the guide curve operations which show a positive failure rate. The guide curve operations failed less frequently than the operations with forecasts, though the magnitude of the failure was much greater. There is a general trend of improving performance with increasing forecast horizon with respect to the environmental flow criterion.

We can examine this further by conditioning on the environmental contour, Figure 5.62 highlights the simulations which use a value of $V_{ENV}(s) = 1000$ for the environmental contour, which is the contour with the least conservative hedging policy. This shows the trend of improvement with forecast horizon more clearly. When a different value of the contour is selected, the behavior is not as clear. Figure 5.63 shows the simulations which used a value of 8250 for the environmental

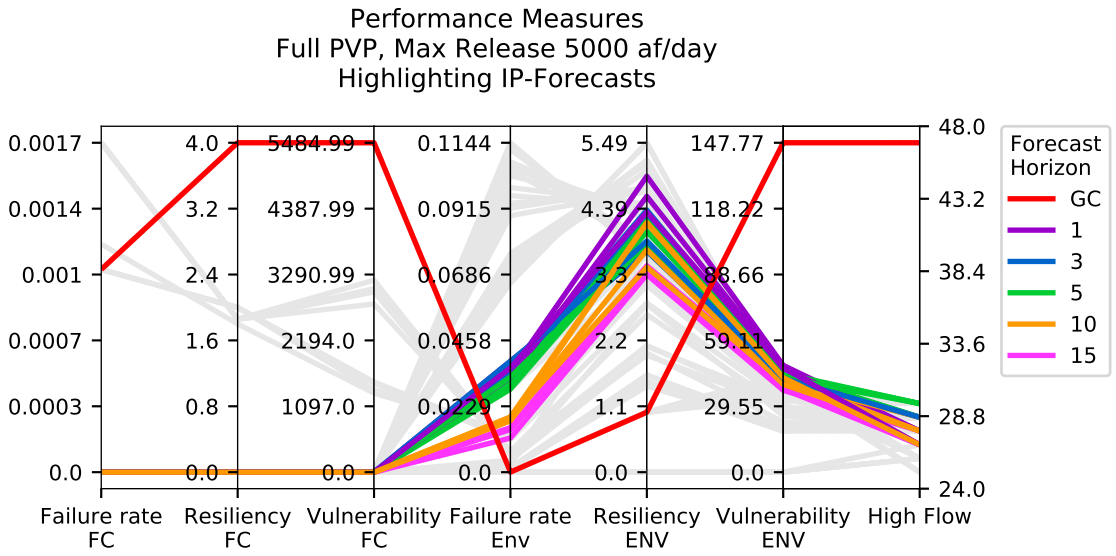

Figure 5.61: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a maximum daily release of 5,000 af/day. There are 21 simulations represented one for each combination of 5 forecast horizons and 4 environmental contours plus one simulation for the guide curve operations. The simulations which use the forecasts show a general trend of improving performance with extended forecast horizon.

flow contour. These measures do not show as strong of a relationship between the order of the performance value and the forecast horizon.

The improvement provided by the use of the IP forecasts is more noticeable in the scenarios with higher water supply stress. Figure 5.64 shows the results for the simulations with a "Zero PVP" diversion and a maximum release of 5000 af/day. Figure 5.66 highlights the simulations which use a value of $V_{ENV}(s) = 100$ for the environmental contour with the IP forecasts, which is the least conservative contour level for this value of the PVP diversion. In this case, the simulations which used forecasts outperform the guide curve operations in all performance measures. Also, the simulations show a general trend of improved performance with increasing forecast horizon in all measures except for the number of high flow days at the Hopland node.

Effect of the Maximum Daily Release

It was expected that the use of a higher maximum daily release would improve performance by allowing the system to store more water prior to a forecasted storm event and to delay the release of that storage until closer to the arrival of the event. The system would then benefit from the reduction in uncertainty in the forecasts at shorter lead-times. Figure 5.67 shows the results

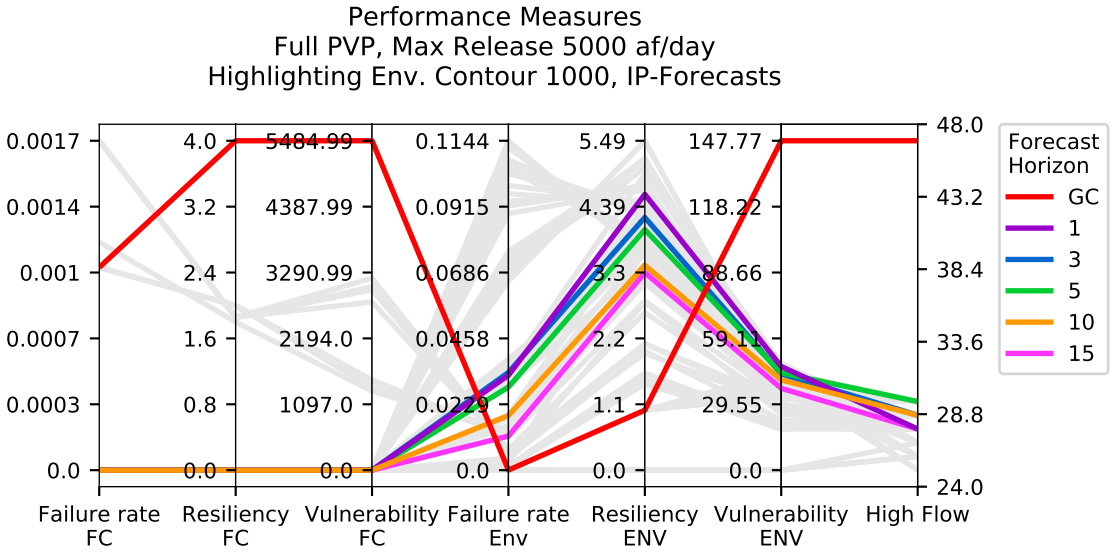

Figure 5.62: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a maximum daily release of 5,000 af/day further conditioned on using a value of $V_{ENV}(s) = 1000$ for the environmental contour. The simulations show a general trend of improving performance with extended forecast horizon.

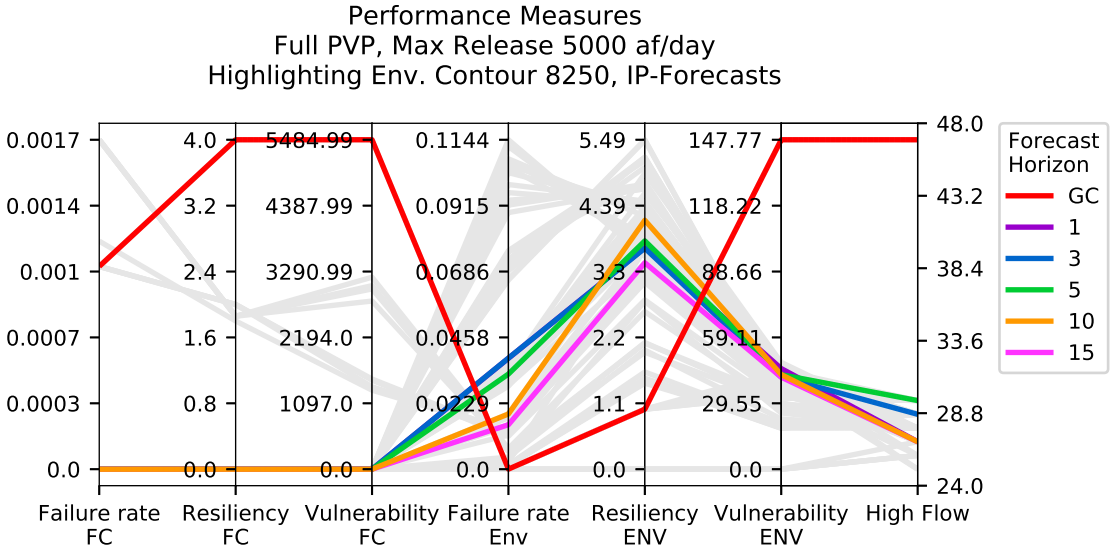

Figure 5.63: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a maximum daily release of 5,000 af/day further conditioned on using a value of $V_{ENV}(s) = 8250$ for the environmental contour. The simulations show some relationship between performance and forecast horizon for the failure rate for the environmental flow criterion, but do not show the same trend for the resiliency, vulnerability, and high flow measures.

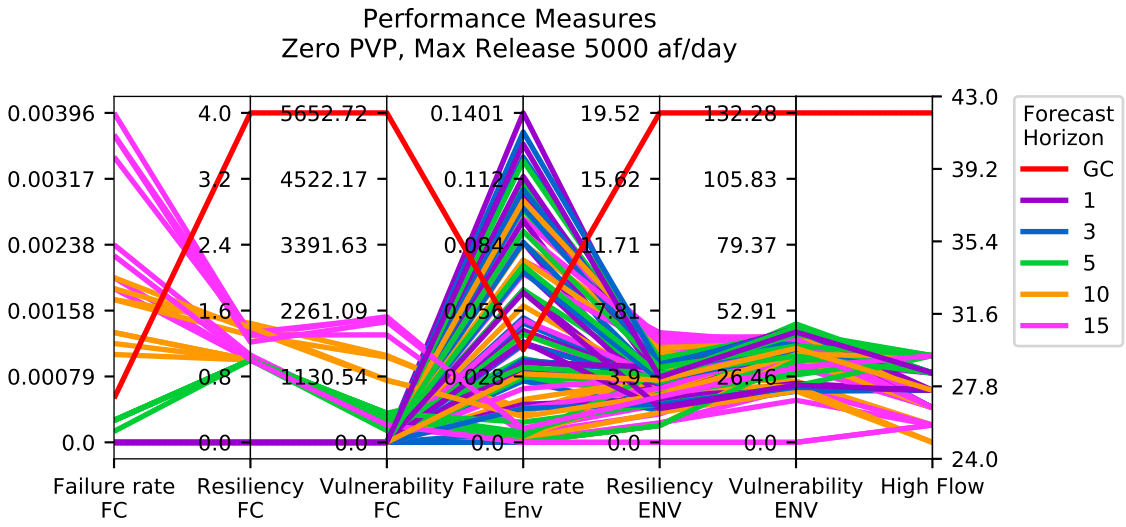

Figure 5.64: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Zero PVP" scenario with a maximum daily release of 5,000 af/day.

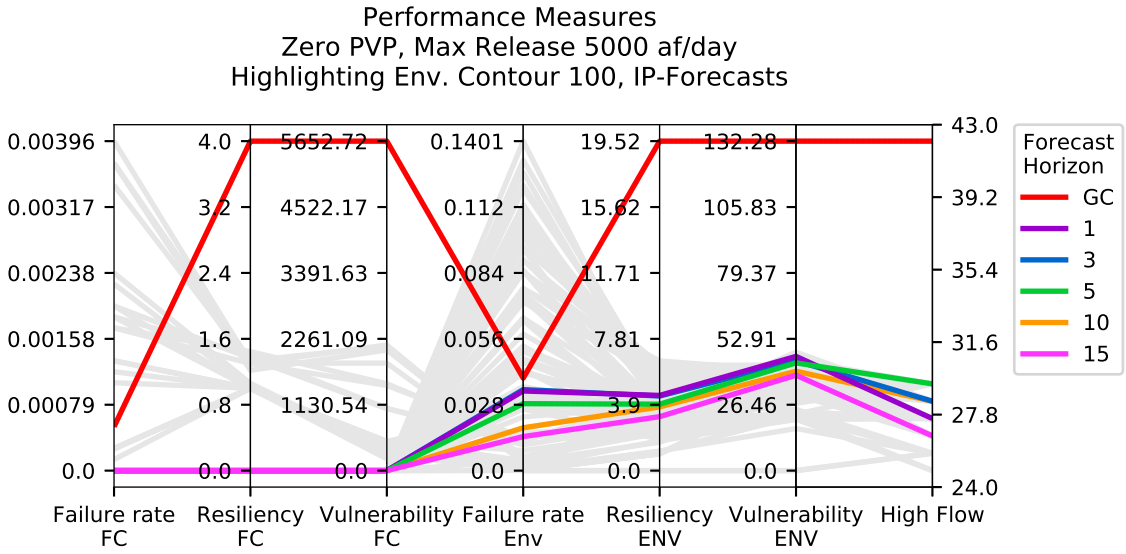

Figure 5.65: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Zero PVP" scenario with a maximum daily release of 5,000 af/day further conditioned on using a value of $V_{ENV}(s) = 100$ for the environmental contour, which is the least conservative level for this value of the PVP diversion. The simulations which used forecasts outperform the guide curve operations in all measures and generally show a trend of improving performance with increasing forecast horizon for all measures except for the Hopland high flow.

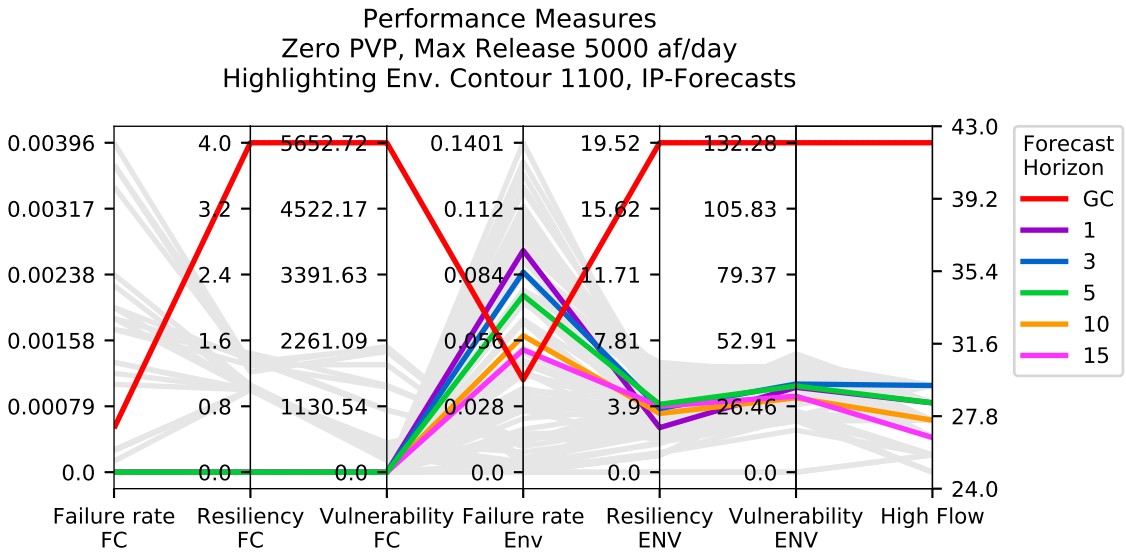

Figure 5.66: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Zero PVP" scenario with a maximum daily release of 5,000 af/day further conditioned on using a value of $V_{ENV}(s) = 1100$ for the environmental contour, which is the most conservative level. The failure rates with respect to the environmental flow criterion are higher than the guide curve and those simulations which used a value of 100 for the environmental contour, though the resiliency and vulnerability are similar.

for the "Full PVP" scenario using the hindcasts/perfect irrigation (IP) forecasts. It is difficult to identify any trends in this representation. If we select the subset of simulations that also use the 15-day forecast horizon, we can begin to see the benefit of increasing the maximum daily release (Figure 5.68). These simulations outperform the guide curve operations with respect to the flood control objective. The guide curve operation has a lower failure rate and resiliency value for the environmental flow objective but with a much higher value for the average magnitude of a failure (vulnerability). The number of days with high flows at the Hopland node are higher as well. By removing the simulation with the guide curve operations, we can observe the effect on performance in more detail (Figure 5.69).

The change in performance due to increasing the maximum daily release is more pronounced in the condition of higher water stress. In the "Zero PVP" scenario, the performance again improves for the failure rate and resiliency measures for the environmental flow criterion (Figure 5.70). In this case, the performance also improves for the vulnerability measure and increasing the maximum release consistently results in a larger number of time steps with high flows at the Hopland node. The effect of increasing the maximum release is more pronounced in this case as the failure rate for each contour is reduced by approximately half when the maximum release is increased.

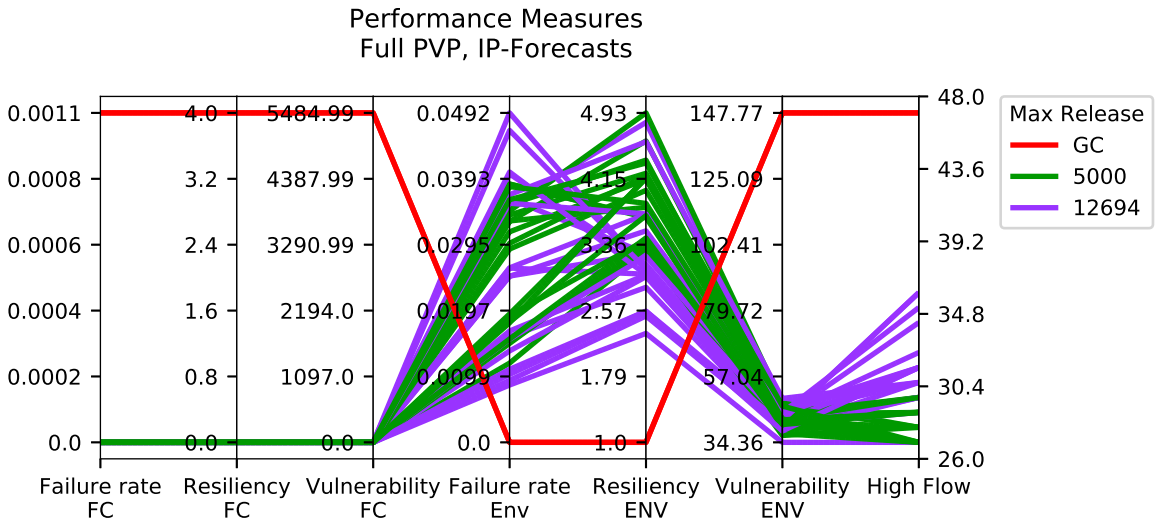

Figure 5.67: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario. The simulations which use forecasts are colored according to the maximum release used in the simulation. The performance of the guide curve operation was the same for both levels of the maximum release and is colored in red.

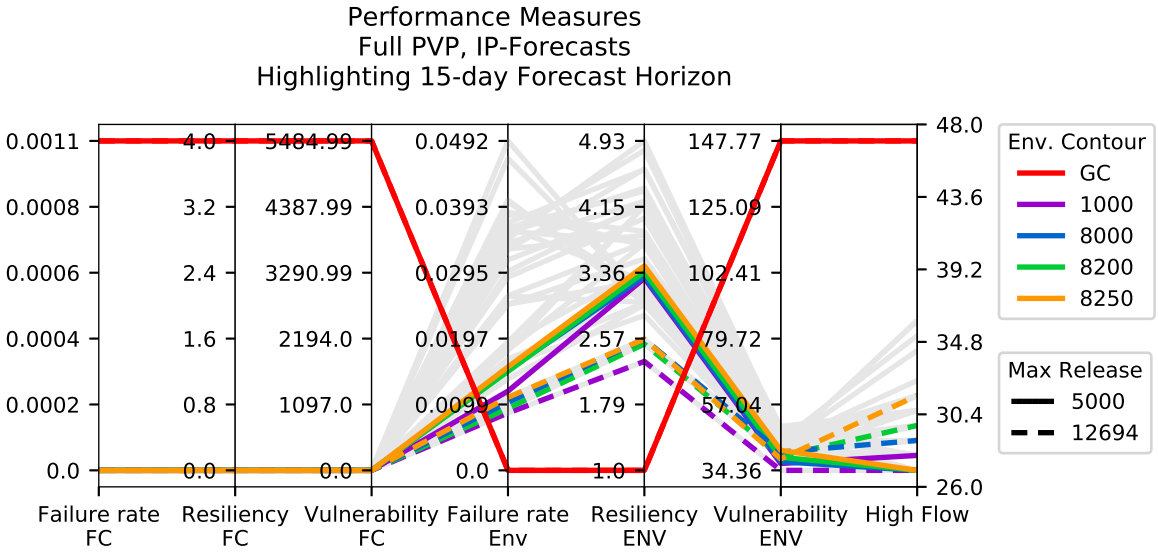

Figure 5.68: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a 15-day forecast horizon. The simulations which use forecasts are colored according to the environmental contour used in the simulation. The simulations which used a 5000 af/day max release are shown with a solid line, and those that used a 12,694 af/day max release a shown with a dashed line. The performance of the guide curve operation was the same for both levels of the maximum release and is colored in red.

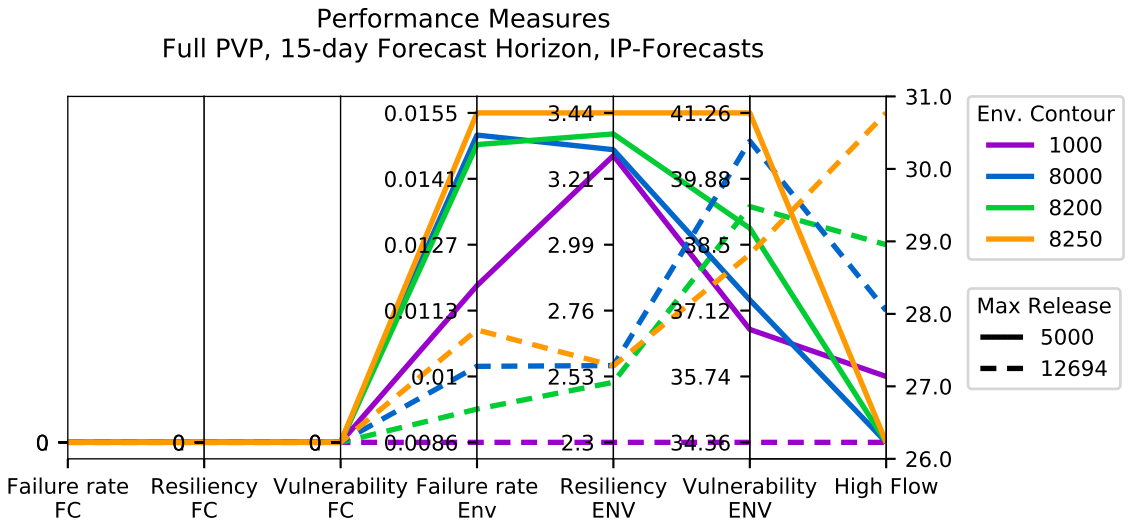

Figure 5.69: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a 15-day forecast horizon. The simulations which use forecasts are colored according to the environmental contour used in the simulation. The simulations which used a 5000 af/day max release are shown with a solid line, and those that used a 12,694 af/day max release a shown with a dashed line. Increasing the max release improves the performance measures for the failure rate and resiliency for the environmental flow objective. There are some simulations which improve with respect to the vulnerability and high flow measures, though others do not.

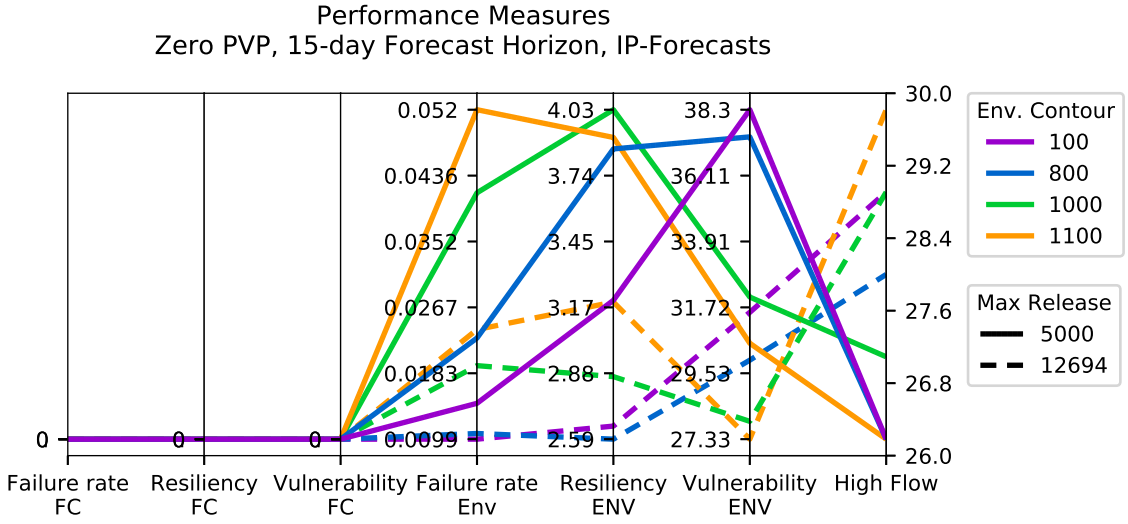

Figure 5.70: Parallel coordinate plot of the simulations which used hindcasts/perfect irrigation (IP) forecasts in the "Full PVP" scenario with a 15-day forecast horizon. The simulations which use forecasts are colored according to the environmental contour used in the simulation. The simulations which used a 5000 af/day max release are shown with a solid line, and those that used a 12,694 af/day max release a shown with a dashed line. Increasing the max release improves the performance measures for the failure rate and resiliency for the environmental flow objective. There are some simulations which improve with respect to the vulnerability and high flow measures, though others do not.

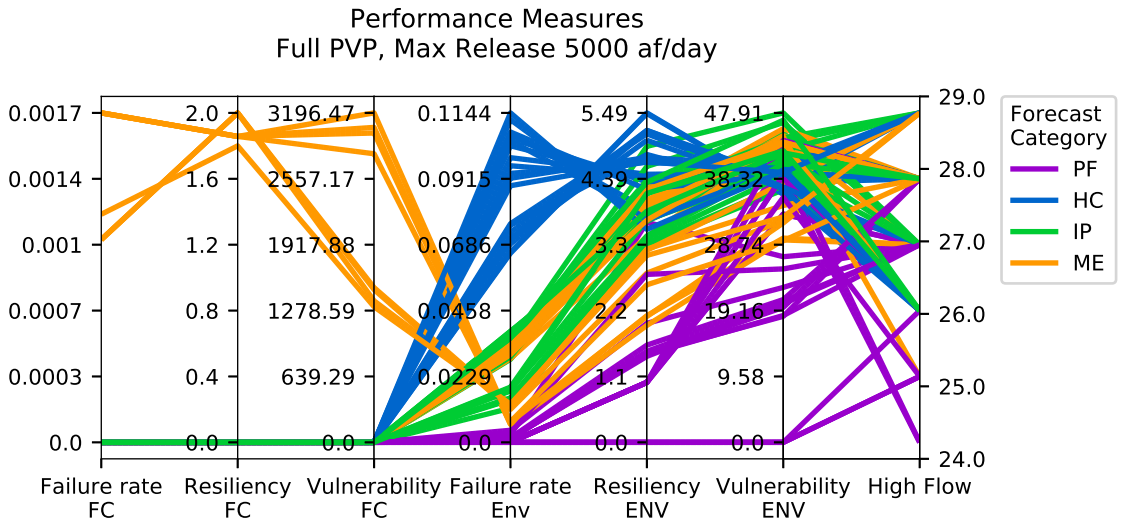

Figure 5.71: Parallel coordinate plot of the simulations which used the "Full PVP" diversion with a 5000 af/day max release. The simulations are colored by the type of forecast used. The guide curve operations are not included. The perfect forecasts (PF) generally perform better overall and the year-round hindcasts (HC) the worst. The hindcasts/perfect irrigation forecasts (IP) and the mean of ensemble (ME) forecasts show somewhat similar results except for the group of simulations with ME forecasts which did not meet the maximum storage level at all times during the simulation period.

Performance of the Mean of the Ensemble

One of the variations examined was to use a forecast created by taking the daily mean over the set of scenarios in the hindcast ensemble. This was included to as a first step in determining if there is a benefit to using the scenarios intact rather than a forecast which represents a statistic of the scenarios. Figure 5.71 shows the set of simulations which used the "Full PVP" diversion with a maximum release of 5000 af/day. The perfect forecasts (PF) generally perform better overall and the year-round hindcasts (HC) the worst. The hindcasts/perfect irrigation forecasts (IP) and the mean of ensemble (ME) forecasts show somewhat similar results except for the group of simulations with ME forecasts which did not meet the maximum storage level at all times during the simulation period.

We can examine these results in more detail by selecting the simulations which use the ME forecasts out of this set and categorizing them by the forecast horizon (Figure 5.72). Here we can see that the simulations which do not meet the maximum storage limit are the ones which use a forecast horizon of longer than 5 days. With a shorter forecast horizon, there is a stricter limit on the ability of the system to deviate above the flood control contour. With a longer forecast horizon, the system is able to maintain a greater storage prior to a storm event leaving it vulnerable to

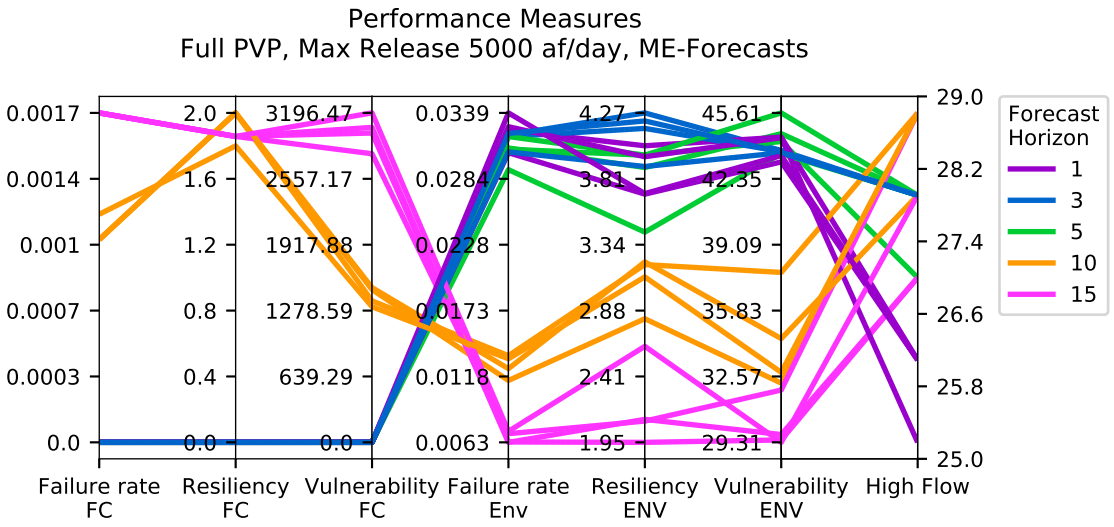

Figure 5.72: Parallel coordinate plot of the simulations which used the "Full PVP" diversion with a 5000 af/day max release and the mean of the ensemble (ME) forecasts. The simulations are colored by the forecast horizon. The guide curve operations are not included. The longer forecast horizons of 10 and 15 days provide better performance with respect to the reliability, resiliency and vulnerability measures for the environmental flow objective. However, this is at the expense of leaving the system vulnerable to exceeding the maximum storage limit. With a forecast horizon lower than 10 days, the system is required to maintain storage closer to the flood control contour. With the longer forecast horizon, the system is able to maintain storage at higher levels above the guide curve which leaves it vulnerable to exceeding the maximum storage when the mean of the ensemble underpredicts the actual inflow.

exceeding the maximum storage limit when the mean of the ensemble underpredicts the volume of inflow to the reservoir.

Chapter 6

Summary and Conclusions

Several methods for using ensemble forecasts in reservoir operations in the past have shown promise as a means for improving the performance obtained from a reservoir system. Significant challenges arise in both methods which use either a dynamic programming or scenario-tree approach and limit their applicability. Both methods suffer from the curse of dimensionality which limits the resolution of the analysis, each requires a probabilistic model of the random variants which are part of the environment and both are limited to problems which do not consider routing effects.

This research presented a method to address these limitations through the use of continuous action deep reinforcement learning to compute the state-value function of a system in the no-forecast condition and then using this state-value function to inform decision-making by representing the value of the states of the system at the end of the forecast horizon. Decision-making during the forecast horizon uses a particle swarm optimization technique to determine optimal sequences of decisions for each scenario individually based on a fitness function constructed to represent the priority given to each objective. The release decision for the current time step is determined by selecting the maximum of the first day releases from the set of optimal sequences.

6.1 Conclusions on the Performance of the Hindcasts

The performance of the hindcasts was analyzed through several measures. These show the difficulty in making an assessment of the value of an ensemble forecast and the importance of considering multiple performance measures along with the application in which the forecast is used when making this assessment. Consider the percent bias of the winter forecasts. This measure suggested that the winter forecasts underpredict the actual inflow when using either the mean or median of the ensemble as the representative value (Figure 5.3). However, the rank histogram (Figure 5.8) showed that the most common position of the actual flow was the first position, indicating that overprediction is the most common condition. Additionally, Figure 5.10 shows the percentage of ensembles which overpredict is greater than those that underpredict. When we examine the deviations from the ensemble range in Figure 5.11, we can see that the mean deviation outside of the ensemble range is similar for overprediction and underprediction, however, the magnitude

of the underprediction can be much greater. Indeed, when examining the percent bias further we can see that the frequency of underprediction (36%) is less than overprediction (64%), however, the mean magnitude of an underprediction is 8326 af while the mean overprediction is 3327 af. Cumulatively this returns a percent bias value which indicates underprediction on average, even though overprediction is a more common occurrence.

In this application, winter forecasts which overpredict are beneficial to the flood control operations as the system will ensure that there is sufficient capacity for the actual inflow. The less frequent underpredictions, however, are characterized by much larger magnitude in the error and could leave the system susceptible to exceeding the maximum storage when there is insufficient capacity prior to a storm event. At the same time, frequent overprediction will reserve storage capacity for inflow that is not realized, leaving less supply available to meet the environmental demands during the dry season.

The assessment of whether the forecasts are "good" or not depends on the benefit to performance that can be achieved by using them, which will be discussed in Section 6.4. For the purposes of this research, one of the intentions of the hindcast performance analysis is to understand the relationship between the accuracy of the forecasts and the performance of the strategy that employs them. Based on these performance measures, it does not appear that the forecasts are unreasonably accurate which suggests that the performance of the forecast-based operations is not dependent on the availability of highly accurate forecasts.

Other aspects of the characteristics of ensemble forecasts could be studied. The effect of timing and magnitude of overprediction or underprediction could be studied in more detail and with reference to the objectives of the problem. The results showed that the ensembles frequently overpredicted during the winter wet season where flood control is a dominant concern. The results also showed that the magnitude of the underpredictions that occurred were large in comparison with the magnitudes of the overprediction. This could perhaps be studied by eliminating either overprediction or underprediction and studying the effect to the forecast based operations.

6.2 Conclusions on the Synthetic Data

The method of synthetic data generation used in this application was developed because of the desire to preserve the short-duration, high intensity inflow patterns of the winter storms that

bring the majority of the precipitation to the basin. Other methods such as the autoregressive-moving-average (ARMA) model and the parameterized autoregressive-moving-average (PARMA) model were studied. While these methods were able to generate data sets with similar mean values, they could not recreate the high peak flow values associated with storm events. This is a critical feature when developing operational strategies for this basin because of the high ratio of storm size to reservoir storage and release capacity.

The synthetic data generator was able to produce a data set that does recreate the peak flow events because it is essentially resampling the historical data intact. The addition of a temporal shift and random variation in the magnitude result in a data set with maximum values which exceed the original data set at all dates throughout the year.

There are several parameters which affect the synthetic data generator. The frequency of switching the sequence which is resampled, the standard deviation of the temporal shift, and the standard deviation of the magnitude shift can all be modified to affect the synthetic data that is generated. The affect of these values on the final performance of the operational strategies was not examined and could be an area of future research.

6.3 Conclusions on the Reinforcement Learning Method

The continuous action reinforcement learning method was applied to estimate the state-value function and the policy function for a reward function which combined the three main objectives of flood control, water supply and environmental flow. This was used to show the ability of the algorithm to learn to operate the system given conflicting objectives. Reinforcement learning was then used to estimate state-value functions for reward functions which represented individual objectives.

The learning process for the combined reward appeared stable except for the "Full PVP" scenario when a value of $\gamma = 0.9999$ was used as the discount factor. In this case there was noticeably more variation of the mean of the $V(s)$ function through the end of the learning process. Though when the corresponding policies were simulated, the results of this scenario were not necessarily out of line with what may be expected based on the results of the other scenarios.

The simulation using the guide curve provides a reference level of performance which can be used to compare the policies output by the learning process. The guide curve operations use a

target maximum level of storage which varies throughout the year to provide storage capacity to meet flood control objectives. When storage is below the guide curve, the operation is then given the ability to iteratively adjust the releases so that the water which is stored is used most efficiently. Given the fixed guide curve then, the performance resulting from this operation sets an upper level of performance for reliability with regard to water supply and environmental flow objectives. The question remains as to the optimality of the current guide curve itself and whether modifications to this limit may improve the performance. But asking this question then brings us back to the original problem of finding optimal operating policies and how to do so in a manner which can accommodate the complexity of this problem.

This complexity arises from the attempt to develop an approach which is applicable to a general set of reservoir operation problems which ultimately requires that the method be able to include routing effects, eliminate discretization and operate without a probabilistic model of the environment. As of now, there is no other known method beyond continuous action reinforcement learning which can handle these factors. The use of the guide curve operations provides a good reference level of performance, but beyond this, assessing the performance depends on comparison to another continuous action reinforcement learning algorithm. There are other methods such as the *asynchronous advantage actor-critic* (A3C) method (Mnih et al., 2016), which directly computes the $V(s)$ function, and could be applied to this problem. This is left for future work.

In comparison to the guide curve operations, the policies obtained from the learning algorithm performed well. Without any knowledge of the downstream diversions, the system is able to find policies which are pareto-optimal when comparing the performance measures with those of the guide curve operations. This is more noticeable for the scenarios with lower water supply availability as the guide curve operations are not able to reduce flows to avoid failure. For the "Zero PVP" scenario, the learned policy with a discount factor of 0.99 outperformed the guide curve with respect to the water supply and flood control measures. The reliability with respect to flood control was lower for the learned policy, though with similar vulnerability and lower resiliency measures compared to the guide curve. These learned policies were based on a simple combination of rewards, without the use of weights to show preference for an objective. The performance of the learned policies likely could be improved for any of the objectives by implementing a weighting strategy to show preference for that objective.

Optimizing the operations in the no-forecast condition was not the objective of this portion of the research, so no work was done to try to find a combination of weights that would match a desired performance outcome. Indeed, the position taken in this research is that the method of using a linear combination of objectives is one to be avoided if possible, as the utility or value function associated with any single objective generally does not provide a value that is commensurate with those returned by any of the other objectives. Creating a value function or utility function is already one level of abstraction, and a weighted linear combination of these functions loses meaning. The forecast-based method presented in this research uses state-value functions based on single objectives. Because the simulation of a single-objective policy does not provide any valuable information about the validity of the policy function or the associated state-value function, the reward function which is a combination of the objectives was used to generate policy functions which could then be used in a simulation to assess the validity of the reinforcement learning method. The resulting performance of the system, when operated according to the policies obtained from the reinforcement learning process, was comparable to the performance using the guide curve operation. The guide curve operation has knowledge of the downstream diversions and the learned policies do not, which makes the performance of the reinforcement learning policies more impressive. Because the learned policies showed this level of performance, the reinforcement learning method proposed could then be relied upon once the focus shifted to developing the single objective state-value functions.

The results obtained using single-objective reward functions showed good repeatability for the environmental objective but did not show the same for the flood control objective. The environmental flow objective implicitly contains a trade-off between storage and release arising from the potential for future rewards which comes with storing water in the near-term versus the release of water to meet current demands. There is no trade-off between storage and release for the flood control objective as there is no future benefit to be realized by storing water in the near-term. It is believed that this is the source of the instability in the learning process and could be an area for future research. It may be possible to include a minimal benefit to storing water in the reward function to create this trade-off between storage and release.

This modification was not pursued in this research as it was desired to maintain the intuitiveness of the state-value functions. With only a penalty for failure to meet the maximum storage limit, the state-value function has some resemblance to a reliability calculation. This link to reliability may

not be necessary, and without this requirement there is the opportunity to explore other reward functions. The relationship to the reliability calculation is intriguing and future research could examine methods to make learning the state-value function of a trivial policy more stable, perhaps through understanding the cause of the distortion in the state-value function noted in Section 5.5.5, Figure 5.40.

Effect of the Discount Factor, γ

Gamma is an important parameter in this process, and its role should be distinguished from other parameters which are used to "tune" the reinforcement learning process. Some parameters, such as the step size of a stochastic gradient update, affect the performance of the learning process. In this example of the step size, a larger value may lead to a decrease in the training time required, but may also lead to instability in the learning process. But this is strictly a computational performance parameter. There is nothing about the nature of the problem that is being analyzed, or the decision-maker, that suggests that a higher or lower value would be more appropriate. The discount factor is different. This parameter changes the way the system makes decisions.

As we saw in the results from the simulation of the policy which was trained on the reward function that combined the objectives (Section 5.5.4), a higher discount factor places higher value on future rewards which has the effect of creating a more long-term oriented policy. It can have an effect on the performance of the learning process as we saw in Section 5.5.6, where the state-value functions which used the higher discount factor for the flood control objective became quite irregular. But the effect on the nature of the decision-making is unique and because of it, the discount factor should be considered carefully with respect to the problem being analyzed. The value may not be the same for different objectives (e.g. flood control vs. environmental flow) which adds a new level of complexity to the multi-objective problem. The discount factor also may not be the same for different sets of decision-makers. The performance of the policy for the combined objectives showed that there are trade-offs to the different performance measures, with no single value of the discount factor dominating the set of all performance measures. The discount factor then should be selected which provides the overall performance which aligns with the priorities of the decision-makers.

In this application, the same discount factor is used at all time steps suggesting that rewards are uniformly discounted at each stage. This may not necessarily be the case and would depend on the decision-maker. Further research could examine the effect of eliminating this assumption.

To summarize this section on the reinforcement learning process, the results show the trade-offs between the three objectives as well as the three measures of performance. Improving the vulnerability for the environmental flow or water supply objectives in some cases will decrease the performance with respect to flood control. Improving reliability for any objective often comes with a decrease in performance with respect to resiliency and vulnerability.

The policies obtained from the reinforcement learning process performed very well compared to a the guide curve operation which was used as a reference. The guide curve operation had the advantage of possessing knowledge of the downstream diversions and the ability to adjust releases to meet the environmental flow demand without excess flow leaving the network. These policies were learned for state and action spaces that were fully continuous, something that has not been done before in reservoir operation applications. And, this learning process was completed in only 15 hours of training. Additionally, research into continuous action reinforcement learning methods in robotics have successfully applied the DDPG and A3C methods to state and action spaces with up to 37 state and 12 action dimensions when using physics-based models (Lillicrap et al., 2015; Mnih et al., 2016). This research has also begun to explore learning directly from the pixel input from video of the models which suggests states with hundreds of dimensions. All of this suggests that the application of continuous action reinforcement learning methods is not limited to single reservoir system, and the use of these methods should be explored for complex multi-reservoir systems.

6.4 Conclusions on the Forecast Based Operation

The proposed operating strategy which used the ensemble forecasts created a fitness function to be used in an optimization which would determine optimal sequences of decisions for each scenario in the ensemble. The fitness function was developed to prefer storage over release, so the process of action selection is then choosing the maximum of the first stage decisions over the set of scenarios. This method did create a single objective optimization problem but was able to do so without requiring much in the way of subjective selection of weights. The weights that are used essentially provide a means of prioritizing flood control over the environmental flow objective. And, the fitness

function is able to use the state-value functions which were developed through reinforcement learning processes which considered each objective individually. The hedging mechanism which reduces the release when storage lies below the desired level of performance defined by the environmental contour is one suggestion and other methods could certainly be developed to suit a particular set of decision-makers. The benefit of using the state-value function based method is that the decision-maker is supplied with a quantified return at each state which can then support a hedging decision.

The selection of the flood control contour could certainly involve a more rigorous approach. To preserve data for testing, a single large flood event was used to select the contour which provided the desired level of protection. It may be that this flood event happens to be a worst-case scenario which would lead to a sufficiently conservative selection of the flood control contour. There is another large flood event in the data set which occurred in January of 2006. This event is larger in magnitude in the single day and 10-day total inflow to Lake Mendocino, however, the majority of the inflow occurs over a very short time frame, allowing time for the reservoir to be drawn down prior to the peak inflow, without being constrained by the Hopland rule. If the same process of selecting the flood control contour is repeated for this flood event, a lower value of the flood contour would be selected which is a less conservative level of protection for flood control. Future research could examine methods for performing this part of the analysis. Also, it was desired to select a flood event that occurred during the period 1985-2010 when hindcasts are available in order to examine the performance with and without forecasts. The selection of the contour is made based on the performance of the system without forecasts so the selection process could be conducted based on flood events that occurred at other times in the historical record, or it could be conducted with synthetic hypothetical flood events.

It would certainly be useful if the contour values could be related directly to reliability. Then the selection of the contour would simply be a process of matching the contour value with the desired level of reliability. The value of the contour is a discounted sum of binary values, as opposed to a reliability calculation in which success or failure is not discounted. This leaves the value of the state-value function dependent both on the number of failures as well as the timing of those failures. A sequence of outcomes with more failures that are further in the future could result in the same state-value as one with fewer failures in the near-term. Perhaps it could be shown that, with the

right conditions on the synthetic data set, the state-value function for the flood control objective can provide a good approximation of reliability.

In this process of selecting the flood control contour, we can observe one of the important features of the overall value-function based strategy. This feature is that it does not rely on a sufficiently long forecast horizon. Because the target storage at the end of the forecast horizon is defined by the flood control contour, any length of forecast horizon can be used. If we were to simply state that at the end of the forecast horizon, the storage should below the maximum, then shorter forecast horizons may not be able to provide enough time for the storage to be evacuated from the reservoir prior to a flood event. As we can see in Figure 5.51, storage is maintained below the maximum limit for all contours with a 15-day forecast horizon. As the forecast horizon is reduced, failure occurs for the lower values of the contours. The maximum storage level is equivalent to a contour value much lower than the values simulated suggesting that if this level was used as the target storage at the end of the forecast horizon, failure would frequently occur as the forecast horizon is shortened. If the operator decided to change the forecast horizon used for making decisions, the proposed method would still be applicable. And, this same method could be applied to systems where long-horizon forecasts are not available.

Four sets of forecasts were used to study the performance of the forecast based operations strategy. The simulation was run first with perfect forecasts (PF) which provides an upper limit to the performance of the value-function based strategy. The simulation was also run using the hindcasts with 35 scenarios in each ensemble (HC). Overprediction during periods of low flow negatively impacted the performance as the system would not release sufficient volume to meet downstream demands. Because this occurred mostly during the irrigation season, an additional set of forecasts was used which was made up of hindcasts during the wet season and perfect forecasts during the irrigation season (IP). This set of forecasts is considered the closest comparison to operations in practice as operators are able to monitor downstream flows and make releases to maintain minimum flow levels during the summer irrigation season. This condition of overprediction combined with low flows leading to insufficient releases also occurred outside of the irrigation season. This was not corrected, and the performance measures reflect these failures. A correction for this could be included and the effect on the performance measures examined again.

A set of forecasts which represented the mean of the ensemble was also included (ME). These forecasts outperformed the IP forecasts with respect to the environmental flow criteria. This arises because the mean of the ensemble is always less than the highest flow scenario in the ensemble, which then causes the system to release less water prior to a storm event. This maintains more storage to meet environmental flow demands during the dry season though also at the cost of leaving the system vulnerable to exceeding the maximum storage limit. Should this trade-off be considered acceptable, it may be tempting to use the mean of the ensemble instead of the full set of scenarios. This is not recommended as the calculation of the performance measures with respect to the flood control objective is then more closely related to the statistical characteristics of the forecasts. Should the quality degrade or shift from overprediction to under prediction, the analysis would need to be repeated.

An ensemble forecast provides a set of scenarios which are all possible outcomes and the ensemble forecasting process itself provides no information about which scenarios are more or less likely to occur. Flood control operations are conservative by nature and the method proposed provides a means of decision-making which protects the system from exceeding the maximum storage limit by considering all scenarios as possible, including those which differ from the mean significantly, without requiring any assumptions about the relative likelihood of any scenario. This may result in water begin released from storage which is not replaced when a storm event is overpredicted, however, this approach provides for reliable performance with respect to flood control without requiring high precision forecasts. Any improvement in the forecasting system will then go to benefit the water supply and environmental flow objectives, rather than requiring improvement in the forecasting accuracy in order to meet the desired level of performance with respect to flood control. This approach is based on the interpretation of the maximum storage limit as critical to the integrity of the dam leading to the desire to have high confidence in avoiding spilling at all times.

When using the hindcasts with the perfect forecasts during irrigation season (IP), the maximum storage limit was met in all cases and extending the forecast horizon generally led to lower failure rates with respect to the environmental flow objective. The resiliency and vulnerability measures of the environmental flow objective did not consistently improve with longer forecast horizons, though the relative ranges of the magnitude of these measures was lower. The improvement in the

failure rate then might be worth the smaller trade-off in performance with respect to resiliency and vulnerability.

With the higher maximum daily release of 12,694 af/day, the performance measures were consistently better compared to the maximum daily release of 5,000 af/day. In the "Zero PVP" scenario, the failure rate for meeting the environmental flow requirement was reduced by nearly half when using the IP set of forecasts. The larger maximum release allows the system to delay making releases necessary to provide storage capacity prior to a forecasted storm event. The performance benefits from this as the system does not need to react to forecasted events which are further in the future when uncertainty is greater. This avoids preemptive releases ahead of forecasted events which may not subsequently realize the predicted inflow volume, with the preserved water then being available to meet water supply and environmental flow needs during the dry season.

Establishing the ability to make a larger daily release from the reservoir is a means of recuperating most of the loss of performance associated with the lower water supply availability that would result from the reduction or elimination of the PVP diversion. And, the forecast based strategy showed better overall performance compared to the guide curve operations when using the less conservative environmental contours (100, 800) when the PVP diversion was eliminated. The guide curve operations showed failures in all three objectives under this condition. Using the forecasts, the maximum storage level and water supply objectives were met at all time steps. The reliability, resiliency and vulnerability measures for the environmental flow objective were all improved over the guide curve operations. When the more conservative contours (1000, 1100) were used, the reliability is slightly lower than the guide curve operations, though with the benefit of lower vulnerability.

These performance measures are somewhat degraded by the inability of the system to correct for overprediction during periods of low flows outside of the irrigation season. Some of the failures to meet the environmental flow objectives did occur when there was sufficient storage in the reservoir to meet the demand. This was corrected during the irrigation season by replacing the hindcasts with perfect forecasts which is believed to be appropriate as operators have the ability to monitor downstream flows and adjust the release to maintain minimum flow levels. This same correction is not possible during the wet season when operations are concerned with flood control. Replacing the hindcasts with perfect forecasts in this case would be inappropriate as there is no comparable practice in which operators are able to gain information more reliable than the hindcasts.

Overall this approach appears to be promising as it is able to address many of the limitations of previous methods. There is no need to assume probability distributions which may be difficult to obtain. The method is able to handle delayed rewards associated with routing effects. There was no need to rely on a target storage to represent the desired state of the system at the end of the forecast horizon. The reinforcement learning process provides this information the the form of the state-value function. A multi-objective problem can be analyzed without the need to employ a weighted linear combination which risks losing meaningfulness to the decision-maker. In the reinforcement learning step, the objectives are analyzed individually which allows the appropriate discount factor to be applied to each objective. The method does not rely on sufficient release capacity or forecast horizon length to be effective. The continuous nature of the reinforcement learning process mitigates the curse of dimensionality associated with discretization and allows for very fine resolution of analysis of both the state and action spaces. This also will be a benefit when the analysis is expanded to include multiple reservoirs.

Also, the method of determining optimal sequences of decisions for each scenario in the ensemble forecast allows for the scenarios to be used intact which preserves the temporal and spatial correlation. Using the graph-based computing technique, the optimization for each scenario is done concurrently. The computing time for the optimization is approximately linear with the length of the forecast horizon and rather insensitive to the number forecasts (there are a total of 6 forecasts used in the Lake Mendocino case study). The computing time for scenario-tree based methods are exponential in the forecast horizon and require an additional scenario reduction step when the number of scenarios becomes large. In one experiment using the method presented in this study, the time to perform the simulation for 5 years using 35 scenarios and a 15-day forecast horizon was 219 minutes. When the number of scenarios was increased to 61, the simulation time was 264 minutes. This is an increase of only 21% in the computing time from an increase of 74% in the number of scenarios, far below an exponential increase.

And finally, the method presented in this research overcomes the limitations of methods which provide operating policies without providing information about the consequences of deviating from the policy. With the method presented, the operator is able to incorporate any information that is available at time the decision is made. The state-value functions provide the operator with information about the performance with respect to each objective for every point in the state-space.

The operator is then able to consider the trade-off between each objective that results from any decision.

Suggestions for possible areas of future research have been noted throughout this dissertation and are collected here for convenience.

- The deep reinforcement learning process uses a synthetic data generator to provide the realizations of the random variants in the embedded simulation step. One possible area for further study is to examine the effect of the parameters of this synthetic data generator on the value-functions returned by the deep reinforcement learning process and the performance of the corresponding policies.

- Other methods are available for computing the value-function over a continuous state and action space. The DDPG method presented in this research requires an additional step to compute the $V(s)$ function. The asynchronous advantage actor-critic (A3C) method (Mnih et al., 2016) computes the state-value function directly which could decrease the overall computation time. The research presented by (Mnih et al., 2016) also provides methods of parallelizing the algorithm for additional gains in computing speed.

- Further research could examine different constructions of the reward functions for each objective. For the flood control objective, this work could explore variations which provide more stability to the learning process to produce results with repeatability comparable to that of the ENV and COMBO objectives shown in this research.

- The link between the state-value function for the flood control objective and the reliability with respect to this goal is intriguing and may be a useful area to explore.

- The flood control contour used in operations with forecasts was selected with reference to a single flood event. This selection process could be examined in more detail. Alternative flood events or synthetic flood events could be used.

- It was noted that the discount factor, γ, is an important parameter in the reinforcement learning process. The usual assumption is that this parameter is constant at all stages. This assumption facilitates computation but may not necessarily be true and future work may want to consider the consequences of removing this assumption.

Bibliography

Alfieri, L. et al. (2014). "Evaluation of ensemble streamflow predictions in Europe". *Journal of Hydrology* 517, pp. 913–922. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2014.06.035.

Anvari, S., Mousavi, S. J., and Morid, S. (2014). "Sampling/stochastic dynamic programming for optimal operation of multi-purpose reservoirs using artificial neural network-based ensemble stream ow predictions". *Journal of Hydroinformatics* 16.4, pp. 907–921. ISSN: 1464-7141, 1465-1734. DOI: 10.2166/hydro.2013.236.

Araujo, A. d. and Terry, L. A. (1974). "Operação de sistema hidrotérmico usando Programação Dinâmica Determinística". *Rev. Bras. Energia Elétrica* 29. (in portuguese), pp. 44–45.

Askew, A. J., Yeh, W. W.-G., and Hall, W. A. (1971). "Use of Monte Carlo Techniques in the Design and Operation of a Multipurpose Reservoir System". *Water Resources Research* 7.4, pp. 819–826. ISSN: 1944-7973. DOI: 10.1029/WR007i004p00819.

Bellman, R. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Bouchart, F. J. C. and Chkam, H. (1998). "A reinforcement learning model for the operation of conjunctive use schemes". *WIT Transactions on Ecology and the Environment* 26.

Boucher, M. .-.-A. et al. (2012). "Hydro-economic assessment of hydrological forecasting systems". *Journal of Hydrology* 416417, pp. 133–144. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2011.11.042.

Brier, G. W. (1950). "Verification of Forecasts Expressed In Terms of Probability". *Monthly Weather Review* 78.1, pp. 1–3.

Bryce, D., Cushing, W., and Kambhampati, S. (2007). "Probabilistic planning is multi-objective". *ASU CSE TR*, pp. 07–006.

Castelletti, A. et al. (2002). "Reinforcement learning in the operational management of a water system". *IFAC workshop on modeling and control in environmental issues, Keio University, Yokohama, Japan*, pp. 325–330.

Castelletti, A. et al. (2010). "Tree-based reinforcement learning for optimal water reservoir operation". *Water Resources Research* 46.9, W09507. ISSN: 1944-7973. DOI: 10.1029/2009WR008898.

Charnes, A. and Cooper, W. W. (1963). "Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints". *Operations Research* 11.1, pp. 18–39. ISSN: 0030-364X. DOI: 10.1287/opre.11.1.18.

Charnes, A., Cooper, W. W., and Symonds, G. H. (1958). "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil". *Management Science* 4.3, pp. 235–263.

Colorni, A. and Fronza, G. (1976). "Reservoir management via reliability programing". *Water Resources Research* 12.1, pp. 85–88. ISSN: 1944-7973. DOI: 10.1029/WR012i001p00085.

Côté, P. and Leconte, R. (2016). "Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction". *Journal of Water Resources Planning and Management* 142.2, p. 04015046. ISSN: 0733-9496. DOI: 10.1061/(ASCE)WR.1943-5452.0000575.

Defourny, B., Ernst, D., and Wehenkel, L. (2008). "Risk-Aware Decision Making and Dynamic Programming". *NIPS 2008 Workshop on Model Uncertainty and Risk in RL*.

Delaney, C. and Mendoza, J. (2016). *Forecast Informed Reservoir Operations, Lake Mendocino Demonstration Project, Evaluation of Ensemble Forecast Operations*. Santa Rosa, CA: Sonoma County Water Agency.

Dias, N., Pereira, M., and Kelman, J. (1985). "Optimization of flood control and power generation requirements in a multipurpose reservoir". Symposium on Planning and Operation of Electric Energy Sysetms. Rio de Janero, Brazil: International Federation of Automatic Controls.

Dupaová, J., Gröwe-Kuska, N., and Römisch, W. (2003). "Scenario reduction in stochastic programming". *Mathematical Programming* 95.3, pp. 493–511. ISSN: 0025-5610, 1436-4646. DOI: 10.1007/s10107-002-0331-0.

Dupaová, J., Consigli, G., and Wallace, S. W. (2000). "Scenarios for Multistage Stochastic Programs". *Annals of Operations Research* 100, pp. 25–53.

Eisel, L. M. (1972). "Chance constrained reservoir model". *Water Resources Research* 8.2, pp. 339–347. ISSN: 1944-7973. DOI: 10.1029/WR008i002p00339.

Etkin, D. et al. (2015). "Stochastic Programming for Improved Multiuse Reservoir Operation in Burkina Faso, West Africa". *Journal of Water Resources Planning and Management* 141.3, p. 04014056. ISSN: 0733-9496, 1943-5452. DOI: 10.1061/(ASCE)WR.1943-5452.0000396.

Eum, H.-I., Kim, Y.-O., and Palmer, R. N. (2011). "Optimal Drought Management Using Sampling Stochastic Dynamic Programming with a Hedging Rule". *Journal of Water Resources Planning and Management* 137.1, pp. 113–122. ISSN: 0733-9496. DOI: 10.1061/(ASCE)WR.1943-5452.0000095.

Faber, B. A. and Stedinger, J. R. (2001). "Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts". *Journal of Hydrology* 249.1, pp. 113–133. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(01)00419-X.

Faber, M. H. and Stewart, M. G. (2003). "Risk assessment for civil engineering facilities: critical overview and discussion". *Reliability Engineering & System Safety* 80.2, pp. 173–184. ISSN: 0951-8320. DOI: 10.1016/S0951-8320(03)00027-9.

Fan, F. M. et al. (2014). "Ensemble streamflow forecasting experiments in a tropical basin: The São Francisco river case study". *Journal of Hydrology* 519, Part D, pp. 2906–2919. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2014.04.038.

Fan, F. M. et al. (2016). "Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir". *Water Resources Management* 30.10, pp. 3609–3625. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-016-1377-8.

Ficchì, A. et al. (2016). "Optimal Operation of the Multireservoir System in the Seine River Basin Using Deterministic and Ensemble Forecasts". *Journal of Water Resources Planning and Management* 142.1, p. 05015005. ISSN: 0733-9496. DOI: 10.1061/(ASCE)WR.1943-5452.0000571.

FIRO Steering Committee (2017). *Preliminary Viability Assessment of Lake Mendocino Forecast Informed Reservoir Operations*. Forecast Informed Reservoir Operations Steering Committee.

Fleming, M. (2012). *Determination of a Hydrologic Index for the Russian River Watershed Using HEC-ResSim*. PR-85. USACE Hydrologic Engineering Center.

Glorot, X. and Bengio, Y. (2010). "Understanding the difculty of training deep feedforward neural networks". The 13th International Conference on Artificial Intelligence and Statistics (AISTATS). Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, p. 8.

Gröwe-Kuska, N., Heitsch, H., and Romisch, W. (2003). "Scenario reduction and scenario tree construction for power management problems". *Power Tech Conference Proceedings, 2003 IEEE Bologna*. Vol. 3. IEEE, 7–pp.

Hashimoto, T., Stedinger, J. R., and Loucks, D. P. (1982). "Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation". *Water Resources Research* 18.1, pp. 14–20. ISSN: 1944-7973. DOI: 10.1029/WR018i001p00014.

Hausknecht, M. and Stone, P. (2015). "Deep Reinforcement Learning in Parameterized Action Space". *arXiv:1511.04143 [cs]*. arXiv: 1511.04143.

He, K. et al. (2015). "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". *arXiv:1502.01852 [cs]*. arXiv: 1502.01852.

Houck, M. H. (1979). "A Chance Constrained Optimization Model for reservoir design and operation". *Water Resources Research* 15.5, pp. 1011–1016. ISSN: 1944-7973. DOI: 10.1029/WR015i005p01011.

Houck, M. H. and Datta, B. (1981). "Performance evaluation of a stochastic optimization model for reservoir design and management with explicit reliability criteria". *Water Resources Research* 17.4, pp. 827–832. ISSN: 1944-7973. DOI: 10.1029/WR017i004p00827.

Joeres, E. F., Seus, G., and Engelmann, H. M. (1981). "The Linear Decision Rule (LDR) reservoir problem with correlated inflows: 1. Model development". *Water Resources Research* 17.1, pp. 18–24. ISSN: 1944-7973. DOI: 10.1029/WR017i001p00018.

Keeney, R. L. (1973). "Risk independence and multiattributed utility functions". *Econometrica: Journal of the Econometric Society*, pp. 27–34.

Keeney, R. L. and Wood, E. F. (1977). "An illustrative example of the use of multiattribute utility theory for water resource planning". *Water Resources Research* 13.4, pp. 705–712. ISSN: 1944-7973. DOI: 10.1029/WR013i004p00705.

Kelman, J. et al. (1990). "Sampling stochastic dynamic programming applied to reservoir operation". *Water Resources Research* 26.3, pp. 447–454. ISSN: 1944-7973. DOI: 10.1029/WR026i003p00447.

Kennedy, J. and Eberhart, R. (1995). "Particle swarm optimization". *Proceedings of ICNN'95 - International Conference on Neural Networks*. Proceedings of ICNN'95 - International Conference on Neural Networks. Vol. 4, 1942–1948 vol.4. DOI: 10.1109/ICNN.1995.488968.

Kim, Y.-O. et al. (2007). "Optimizing Operational Policies of a Korean Multireservoir System Using Sampling Stochastic Dynamic Programming with Ensemble Streamflow Prediction". *Journal of Water Resources Planning and Management* 133.1, pp. 4–14. ISSN: 0733-9496. DOI: 10.1061/(ASCE)0733-9496(2007)133:1(4).

Kingma, D. P. and Ba, J. (2014). "Adam: A Method for Stochastic Optimization". *arXiv:1412.6980 [cs]*. arXiv: 1412.6980.

Kovner, G. (2019). *PG&E drops effort to sell Potter Valley Project*. Santa Rosa Press Democrat. URL: http://www.pressdemocrat.com/news/9211996-181/pge-abandoning-water-power-project-in (visited on 04/18/2019).

Labadie, J. W. (2004). "Optimal Operation of Multireservoir Systems: State-of-the-Art Review". *Journal of Water Resources Planning and Management* 130.2, pp. 93–111. ISSN: 0733-9496. DOI: 10.1061/(ASCE)0733-9496(2004)130:2(93).

LeClerc, G. and Marks, D. H. (1973). "Determination of the discharge policy for existing reservoir networks under differing objectives". *Water Resources Research* 9.5, pp. 1155–1165. ISSN: 1944-7973. DOI: 10.1029/WR009i005p01155.

Lee, J.-H. and Labadie, J. W. (2007). "Stochastic optimization of multireservoir systems via reinforcement learning". *Water Resources Research* 43.11. ISSN: 00431397. DOI: 10.1029/2006WR005627.

Lillicrap, T. P. et al. (2015). "Continuous control with deep reinforcement learning". *arXiv:1509.02971 [cs, stat]*. arXiv: 1509.02971.

Loucks, D. P. and Dorfman, P. J. (1975). "An evaluation of some linear decision rules in chance-constrained models for reservoir planning and operation". *Water Resources Research* 11.6, pp. 777–782. ISSN: 1944-7973. DOI: 10.1029/WR011i006p00777.

Lovejoy, W. S. (1991). "A Survey of Algorithmic Methods for Partially Observed Markov Decision Processes". *Annals of Operations Research* 28.1, pp. 47–65. ISSN: 02545330. DOI: 10.1007/BF02055574.

Maurer, E. P. and Lettenmaier, D. P. (2004). "Potential Effects of Long-Lead Hydrologic Predictability on Missouri River Main-Stem Reservoirs". *Journal of Climate* 17.1, pp. 174–186. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2004)017<0174:PEOLHP>2.0.CO;2.

McMahon, T. A., Adeloye, A. J., and Zhou, S.-L. (2006). "Understanding performance measures of reservoirs". *Journal of Hydrology* 324.1, pp. 359–382. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2005.09.030.

Mnih, V. et al. (2015). "Human-level control through deep reinforcement learning". *Nature* 518.7540, pp. 529–533. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14236.

Mnih, V. et al. (2016). "Asynchronous Methods for Deep Reinforcement Learning". *arXiv:1602.01783 [cs]*. arXiv: 1602.01783.

Moy, W.-S., Cohon, J. L., and ReVelle, C. S. (1986). "A Programming Model for Analysis of the Reliability, Resilience, and Vulnerability of a Water Supply Reservoir". *Water Resources Research* 22.4, pp. 489–498. ISSN: 1944-7973. DOI: 10.1029/WR022i004p00489.

Narasimhan, K., Kulkarni, T., and Barzilay, R. (2015). "Language Understanding for Text-based Games Using Deep Reinforcement Learning". *arXiv:1506.08941 [cs]*. arXiv: 1506.08941.

Nash, J. and Sutcliffe, J. (1970). "River flow forecasting through conceptual models part I A discussion of principles". *Journal of Hydrology* 10.3, pp. 282–290. ISSN: 00221694. DOI: 10.1016/0022-1694(70)90255-6.

Nayak, M. A., Herman, J. D., and Steinschneider, S. (2018). "Balancing Flood Risk and Water Supply in California: Policy Search Integrating Short-Term Forecast Ensembles With Conjunctive Use". *Water Resources Research* 54.10, pp. 7557–7576. ISSN: 1944-7973. DOI: 10.1029/2018WR023177.

Pappenberger, F. et al. (2015). "How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction". *Journal of Hydrology* 522, pp. 697–713. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2015.01.024.

Pianosi, F., Castelletti, A., and Restelli, M. (2013). "Tree-based fitted Q-iteration for multi-objective Markov decision processes in water resource management". *Journal of Hydroinformatics* 15.2, p. 258. ISSN: 1464-7141. DOI: 10.2166/hydro.2013.169.

Raso, L. et al. (2013). "Tree structure generation from ensemble forecasts for real time control". *Hydrological Processes* 27.1, pp. 75–82. ISSN: 1099-1085. DOI: 10.1002/hyp.9473.

Renner, M. et al. (2009). "Verification of ensemble flow forecasts for the River Rhine". *Journal of Hydrology* 376.3, pp. 463–475. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2009.07.059.

Revelle, C., Joeres, E., and Kirby, W. (1969). "The Linear Decision Rule in Reservoir Management and Design: 1, Development of the Stochastic Model". *Water Resources Research* 5.4, pp. 767–777. ISSN: 1944-7973. DOI: 10.1029/WR005i004p00767.

Rieker, J. D. and Labadie, J. W. (2012). "An intelligent agent for optimal river-reservoir system management". *Water Resources Research* 48.9. ISSN: 00431397. DOI: 10.1029/2012WR011958.

Roijers, D. M. et al. (2013). "A survey of multi-objective sequential decision-making". *Journal of Artificial Intelligence Research*.

Roulin, E. and Vannitsem, S. (2005). "Skill of Medium-Range Hydrological Ensemble Predictions". *Journal of Hydrometeorology* 6.5, pp. 729–744. ISSN: 1525-755X, 1525-7541. DOI: 10.1175/JHM436.1.

Schaul, T. et al. (2015). "Prioritized Experience Replay". *arXiv:1511.05952 [cs]*. arXiv: 1511.05952.

Schwanenberg, D. et al. (2014). "Short-Term Reservoir Optimization By Stochastic Optimization To Mitigate Downstream Flood Risks".

Schwanenberg, D. et al. (2015). "Short-Term Reservoir Optimization for Flood Mitigation under Meteorological and Hydrological Forecast Uncertainty". *Water Resources Management* 29.5, pp. 1635–1651. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-014-0899-1.

Shi, Y. and Eberhart, R. (1998). "A modified particle swarm optimizer". *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*. 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), pp. 69–73. DOI: 10.1109/ICEC.1998.699146.

Silver, D. et al. (2014). "Deterministic policy gradient algorithms". *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 387–395.

Simonovic, S. P. and Burn, D. H. (1989). "An improved methodology for short-term operation of a single multipurpose reservoir". *Water Resources Research* 25.1, pp. 1–8. ISSN: 1944-7973. DOI: 10.1029/WR025i001p00001.

Simonovic, S. P. and Mariño, M. A. (1980). "Reliability programing in reservoir management: 1. Single multipurpose reservoir". *Water Resources Research* 16.5, pp. 844–848. ISSN: 1944-7973. DOI: 10.1029/WR016i005p00844.

Sonoma County Water Agency (2015). *Lake Mendocino Water Supply Reliability Evaluation Report*.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Second edition. Adaptive computation and machine learning series. Cambridge, Massachusetts: The MIT Press. 526 pp. ISBN: 978-0-262-03924-6.

Takeuchi, K. and Sivaarthitkul, V. (1995). "Assessment of effectiveness of the use of inflow forecasts to reservoir management". *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences* 231, pp. 299–310.

Tejada-Guibert, J. A., Johnson, S. A., and Stedinger, J. R. (1993). "Comparison of two approaches for implementing multireservoir operating policies derived using stochastic dynamic programming". *Water Resources Research* 29.12, pp. 3969–3980. ISSN: 1944-7973. DOI: 10.1029/93WR02277.

Tsitsiklis, J. N. and Roy, B. V. (1997). "An analysis of temporal-difference learning with function approximation". *IEEE Transactions on Automatic Control* 42.5, pp. 674–690. ISSN: 0018-9286. DOI: 10.1109/9.580874.

USACE (1986). *United States Army Corps of Engineers, Coyote Valley Dam and Lake Mendocino Russian River, California Water Control Manual, Appendix I to Master Water Control Manual Russian River Basin, California*. US Army Corp of Engineers, Sacramento District.

— (2007). *HEC-ResSim Reservoir System Simulation, User's manual. Version 3.1*. U.S. Army Corp of Engineers, Hydrologic Engineering Center (USACE-HEC).

Watkins, C. J. C. H. (1989). "Learning from Delayed Rewards". Doctoral dissertation. Cambridge, UK: King's College.

Watkins, C. J. C. H. and Dayan, P. (1992). "Q-learning". *Machine Learning*, pp. 279–292.

World Commission on Dams, ed. (2001). *Dams and development: a new framework for decision-making ; the report of the World Commission on Dams*. 1. ed., repr. OCLC: 257030108. London: Earthscan Publ. 404 pp. ISBN: 978-1-85383-798-2.

You, J.-Y. and Cai, X. (2008). "Determining forecast and decision horizons for reservoir operations under hedging policies". *Water Resources Research* 44.11, W11430. ISSN: 1944-7973. DOI: 10.1029/2008WR006978.

Appendix A

Hindcast Performance

A.1 NSE

NSE values for the ensemble forecasts using the mean or median as representative of the ensemble. The forecasted value is the cumulative flow over the forecast horizon. The top two plots in each figure use the mean of the seasonal values as the reference (Equation 5.1), the bottom figures use the persistence value (Equation 5.2). The winter forecasts show a similar pattern for all locations with a largest value falling in the middle of the forecast horizon. CDLC1 has the overall highest score. CDLC1 and HEAC1 outperform the nodes upstream. The summer forecasts downstream all outperform the forecasts for the inflow to Lake Mendocino with the scores at all other nodes staying positive at all lead times. The summer forecasts show a generally decreasing trend with increasing lead time when compared to the median and mean, and an increasing trend with lead time compared to the persistence reference.

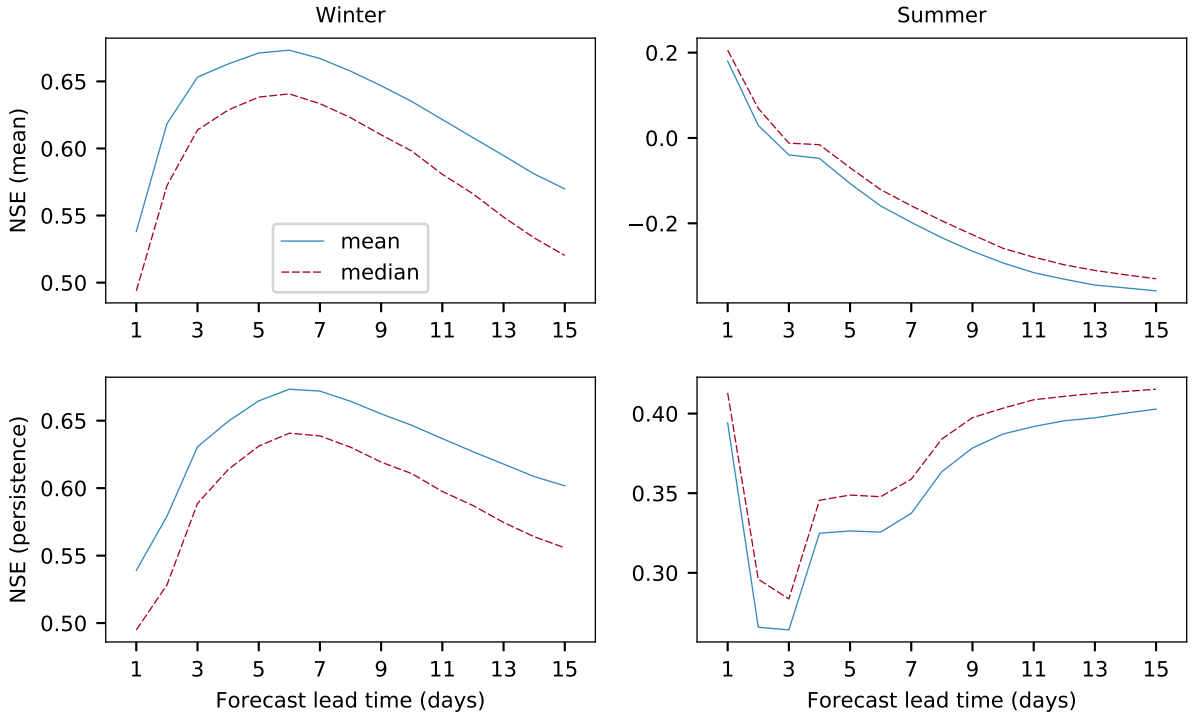

Figure A.1: Nash-Sutcliffe efficiency for the forecast LM_INFLOW (Calpella)

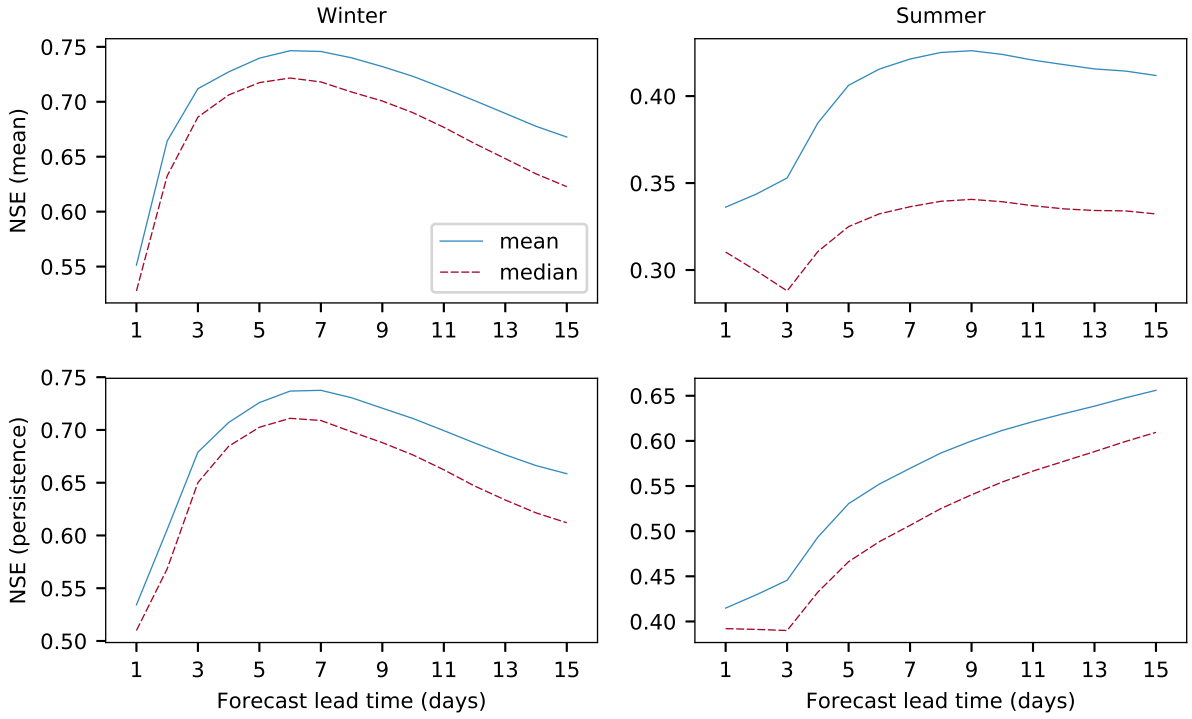

Figure A.2: Nash-Sutcliffe efficiency for the forecast UKAC1 (West Fork)

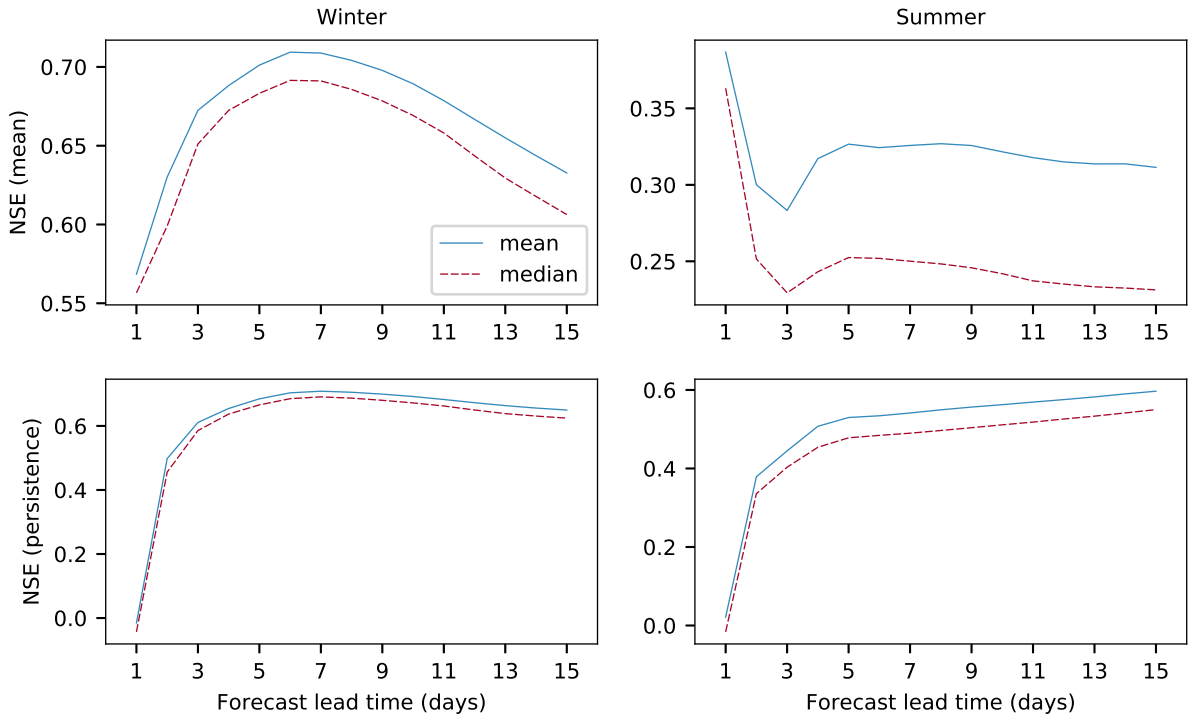

Figure A.3: Nash-Sutcliffe efficiency for the forecast HOPC1 (Ukiah)

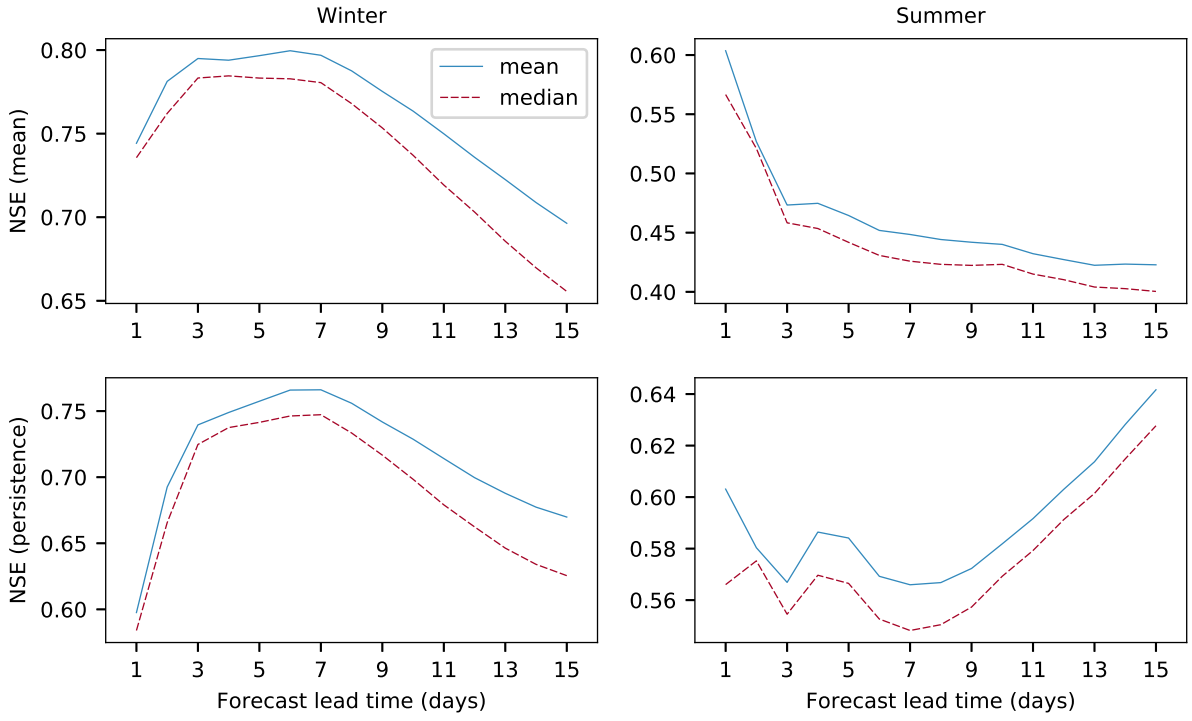

Figure A.4: Nash-Sutcliffe efficiency for the forecast CDLC1 (Cloverdale)

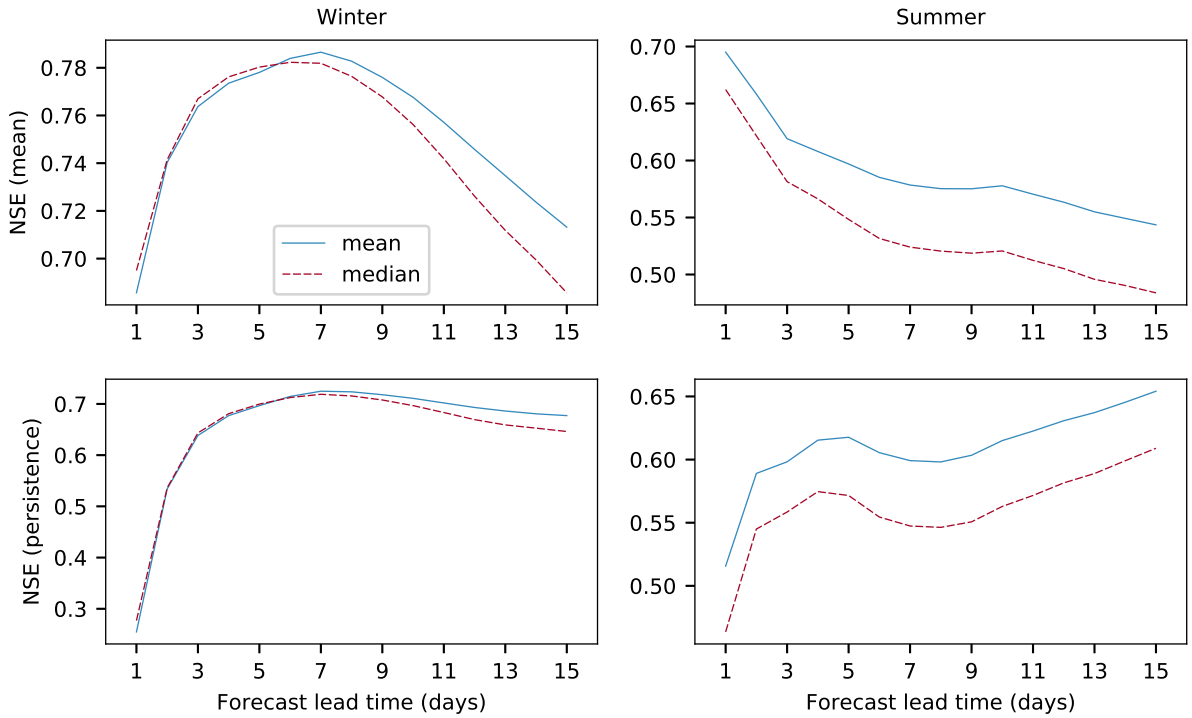

Figure A.5: Nash-Sutcliffe efficiency for the forecast HEAC1 (Big Sulphur Springs)

A.2 Percent Bias

Percent bias values for the ensemble forecasts are shown in Figures A.6 - A.10 using the mean or median cumulative flow as representative of the ensemble, with the forecasts grouped by season. The mean of the ensemble shows less bias for all locations for the winter forecasts, with the bias near zero at HOPC1 and HEAC1. The summer forecasts show much larger bias at all locations with some locations having positive bias and others a negative bias.

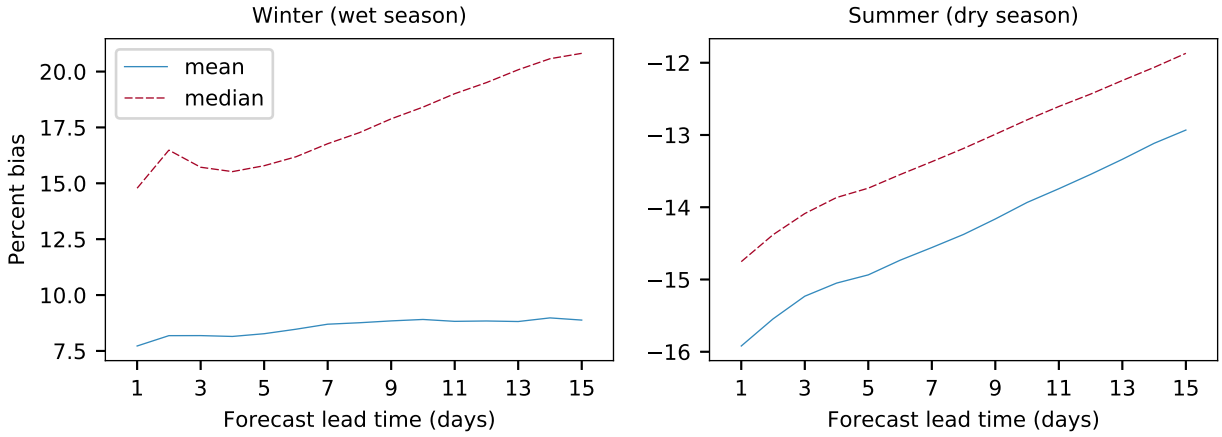

Figure A.6: Percent bias for the forecasts LM_INFLOW (Calpella)

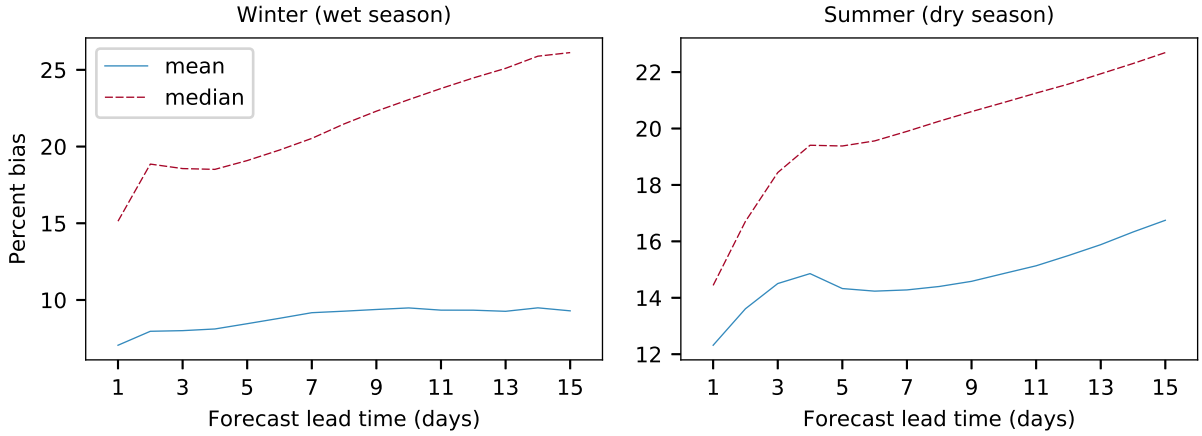

Figure A.7: Percent bias for the forecasts UKAC1 (West Fork)

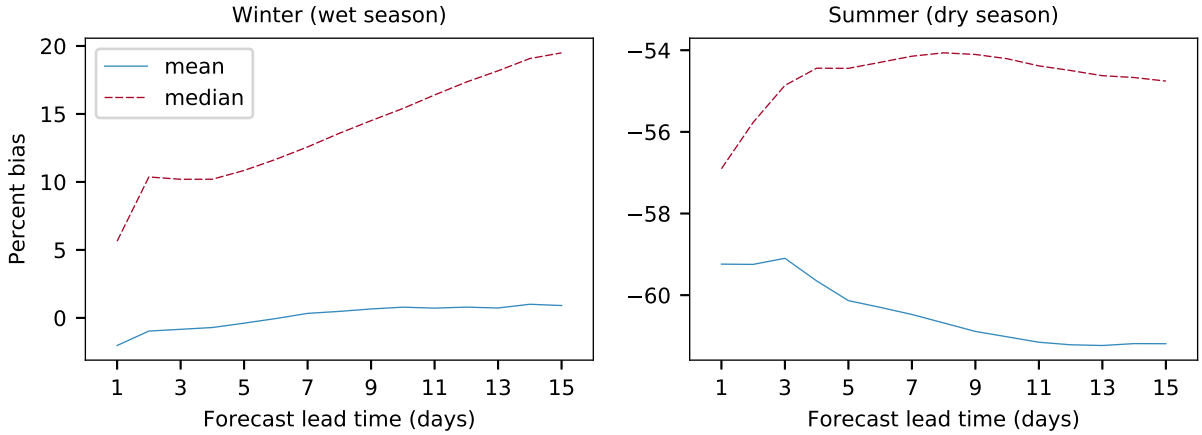

Figure A.8: Percent bias for the forecasts HOPC1 (West Fork)

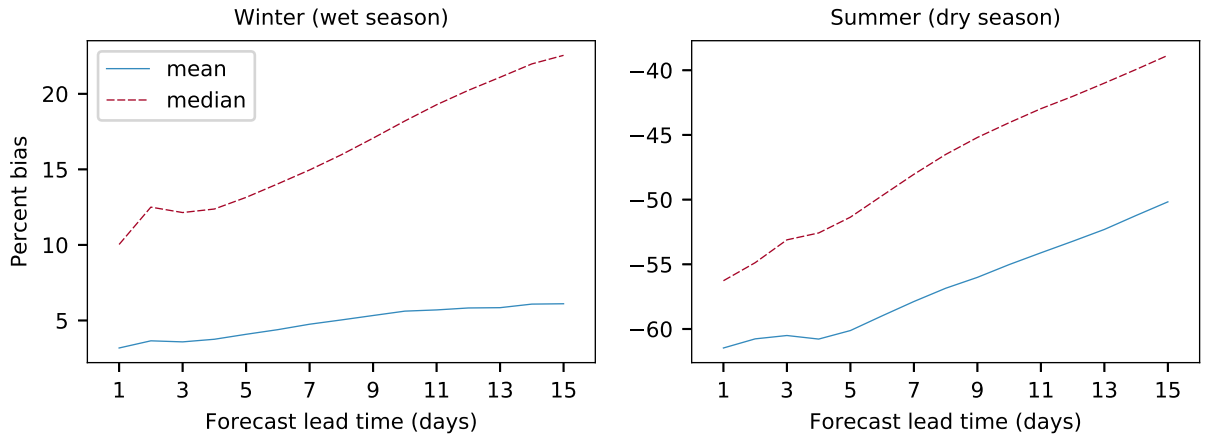

Figure A.9: Percent bias for the forecasts CDLC1 (West Fork)

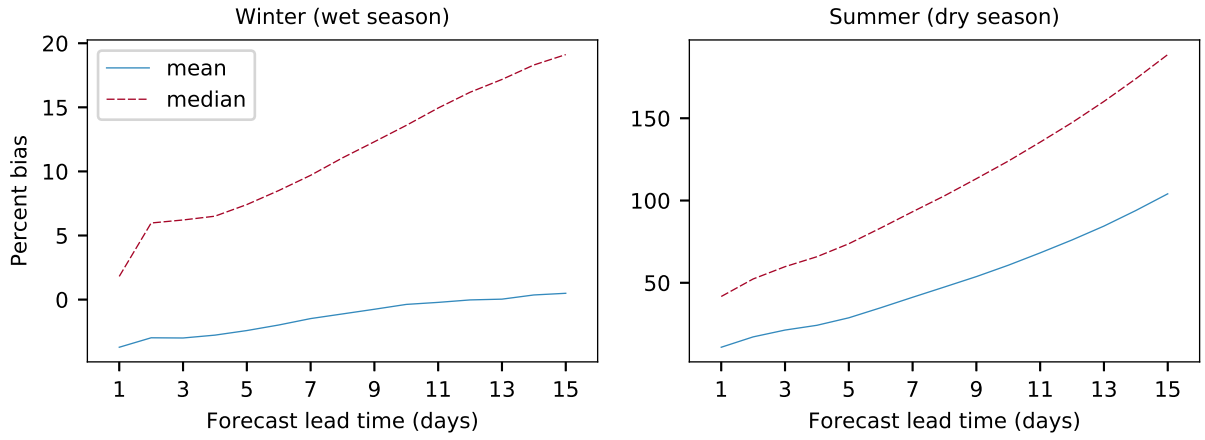

Figure A.10: Percent bias for the forecasts HEAC1 (West Fork)

A.3 CRPSS

Continuous ranked probability skill score for the ensemble forecasts 1985 - 2010 grouped by season with three different values used as the reference forecast. The winter forecasts show positive scores at all locations with the lowest values occurring at the shortest lead times. The persistence forecast is a more challenging forecast to outperform in this case, as CRPSS scores are consistently lower when using the persistence reference. The summer forecasts show a wide variety of scores with some locations having positive and others negative scores.

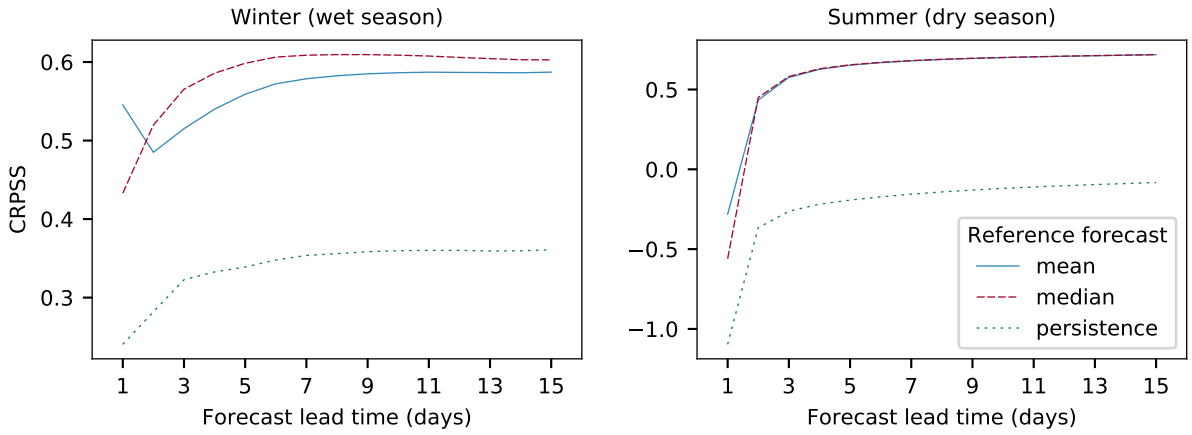

Figure A.11: Continuous Ranked Probability Skill Score (CRPSS) for forecasts at LM_Inflow (Calpella)

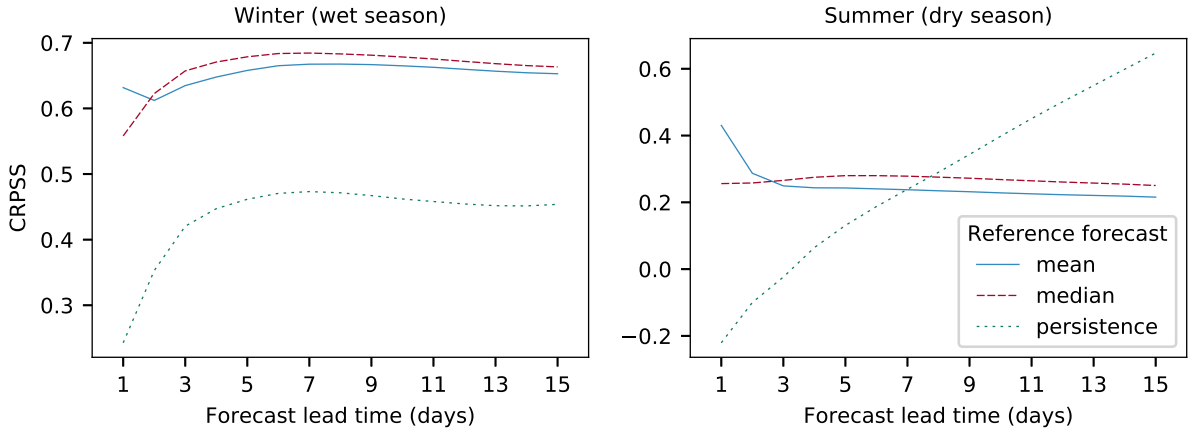

Figure A.12: Continuous Ranked Probability Skill Score (CRPSS) for forecasts at UKAC1 (West Fork)

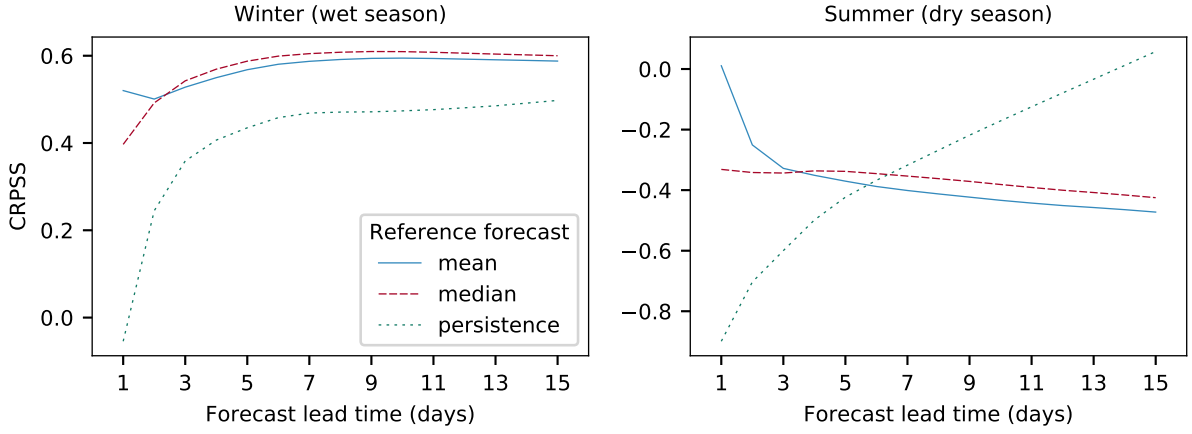

Figure A.13: Continuous Ranked Probability Skill Score (CRPSS) for forecasts at HOPC1 (Ukiah)

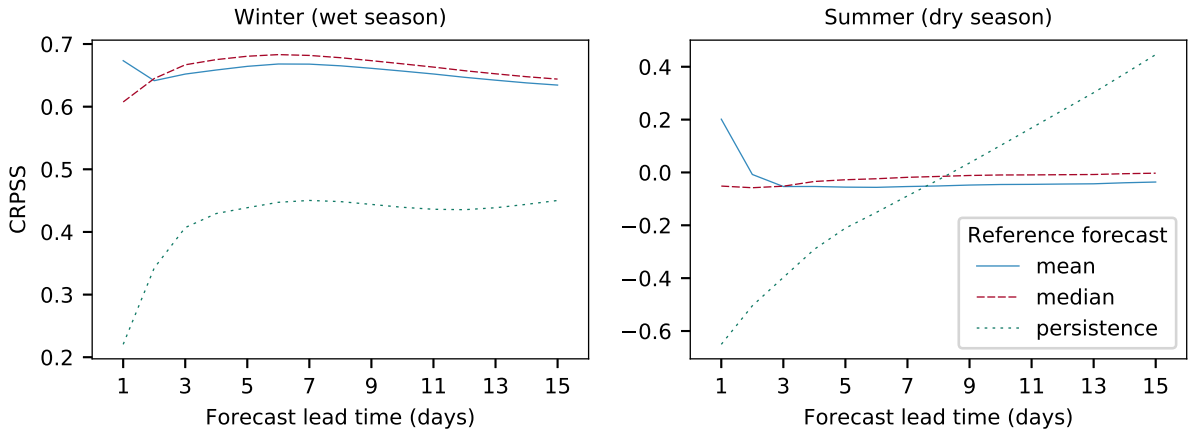

Figure A.14: Continuous Ranked Probability Skill Score (CRPSS) for forecasts at CDLC1 (Cloverdale)

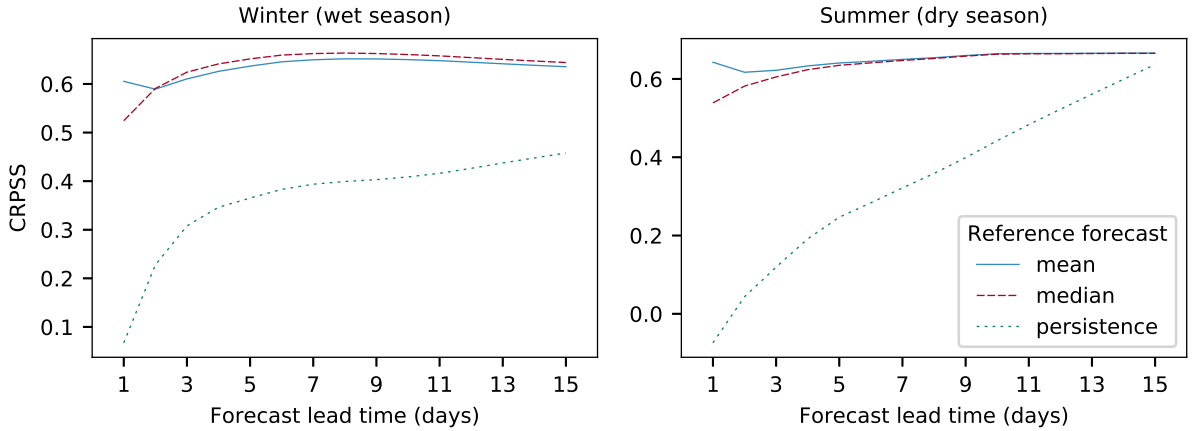

Figure A.15: Continuous Ranked Probability Skill Score (CRPSS) for forecasts at HEAC1 (Big Sulphur Springs)

A.4 Coefficient of Variation of the RMSE

The Root Mean Squared Error (RMSE) and Coefficient of Variation (CV) for each lead time in the forecast horizon grouped by season. The CV for the winter forecasts at all locations initially decreases with increased lead time, then maintaining an approximately constant value through the remainder of the forecast horizon. This value is near 1 for each of the locations. The CV is much larger for the summer forecasts.

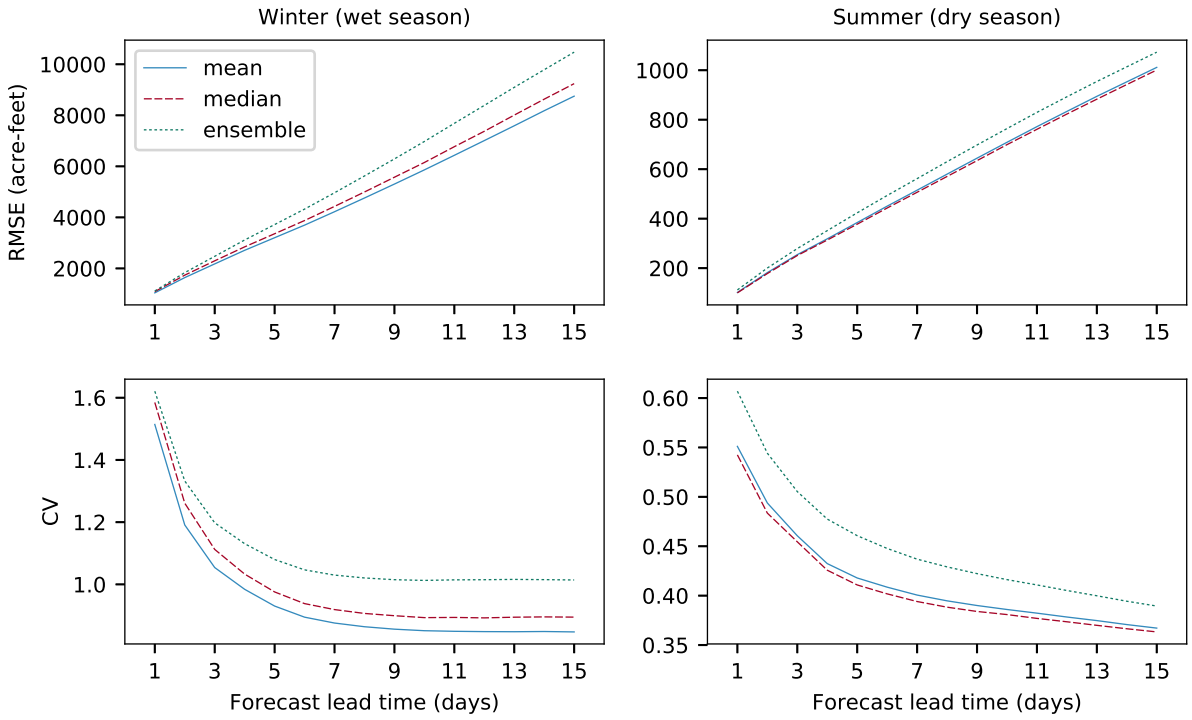

Figure A.16: RMSE and CV of the forecasts at LM_Inflow (Calpella)

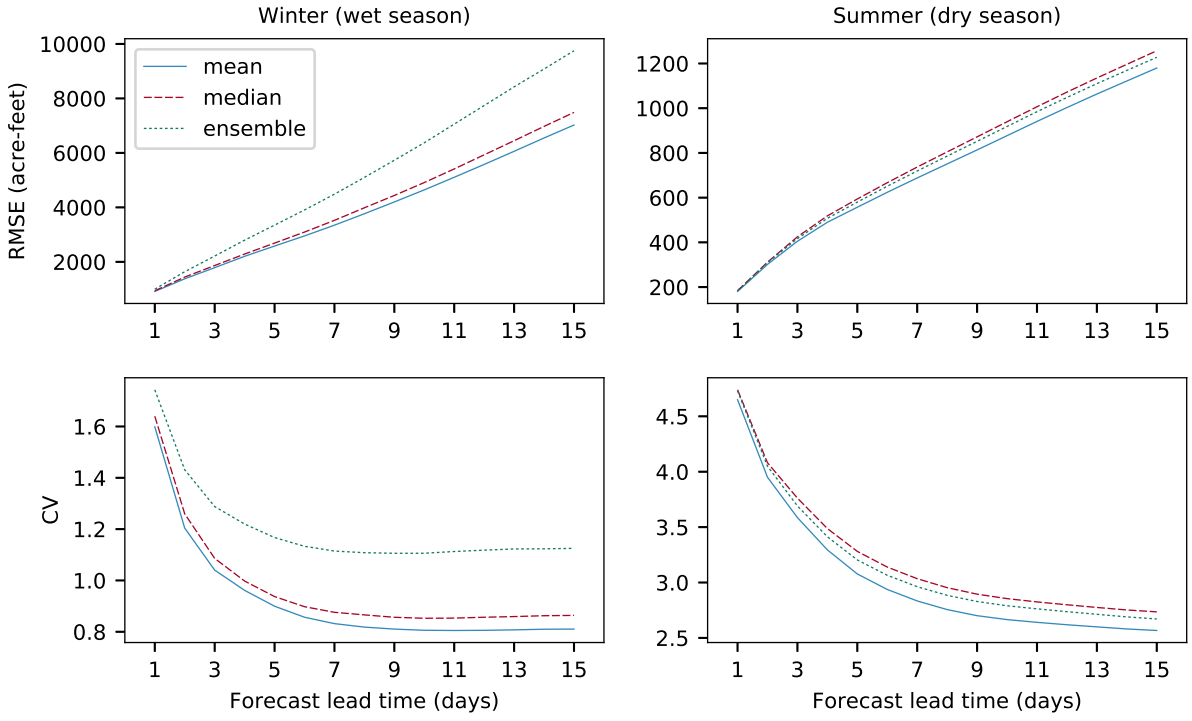

Figure A.17: RMSE and CV of the forecasts at UKAC1 (West Fork)

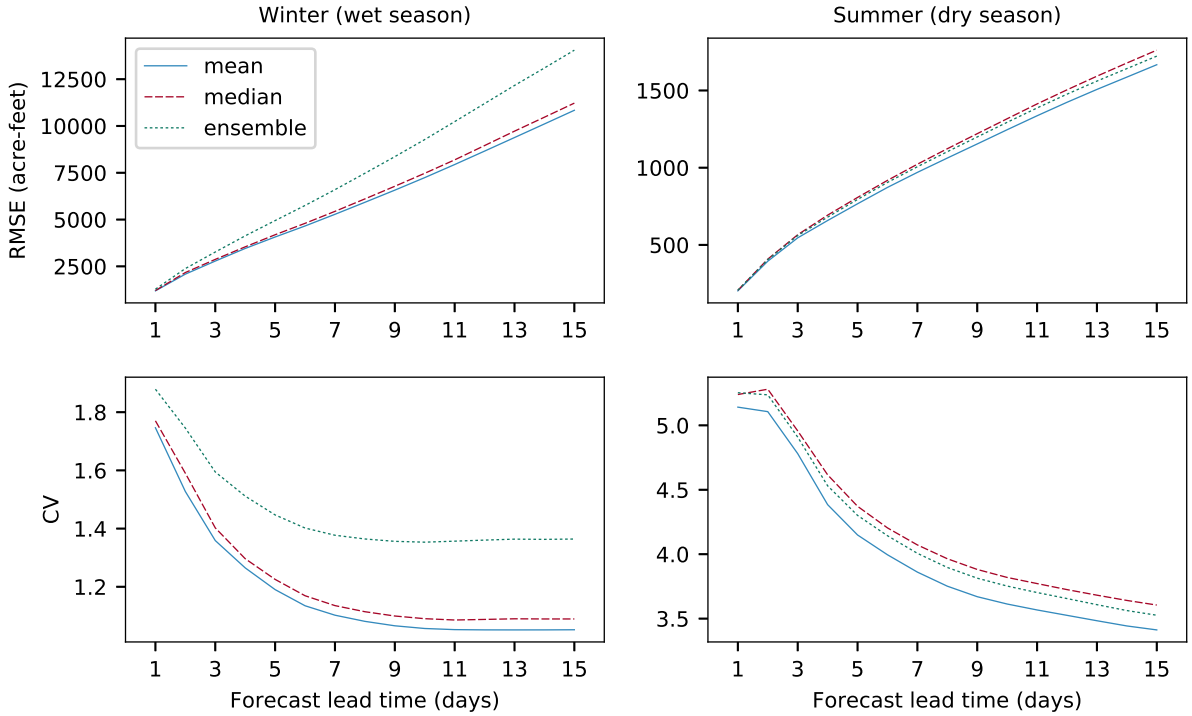

Figure A.18: RMSE and CV of the forecasts at HOPC1 (Ukiah)

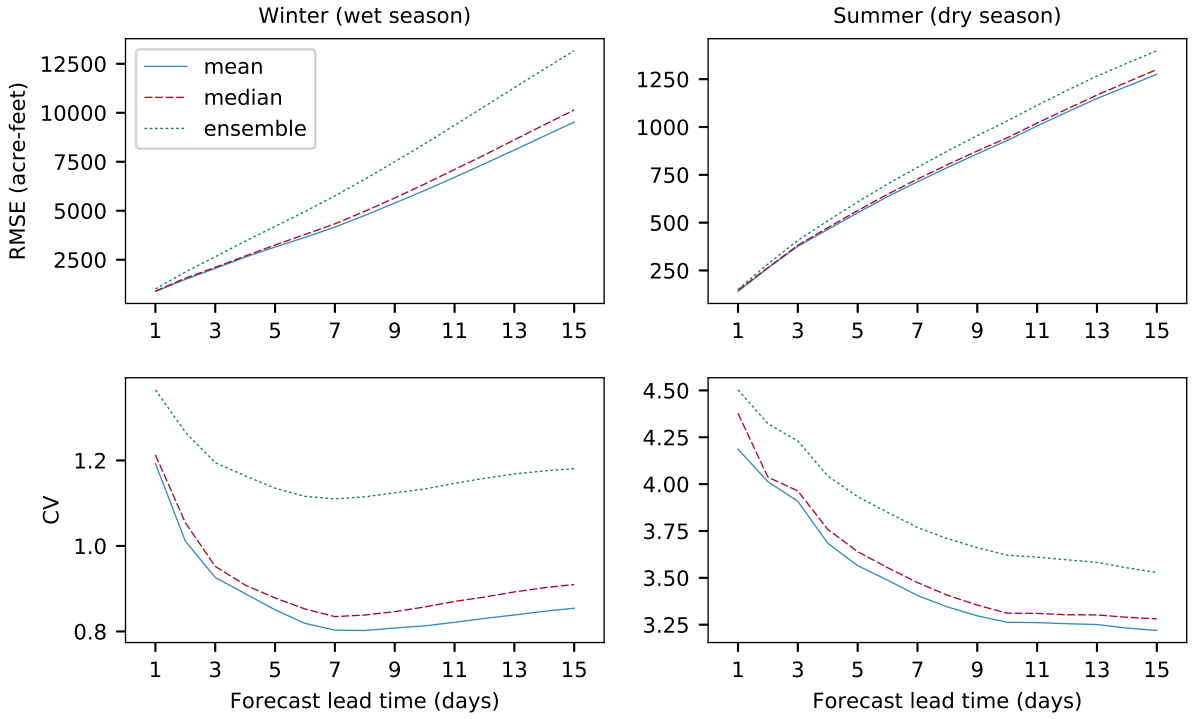

Figure A.19: RMSE and CV of the forecasts at CDLC1 (Cloverdale)

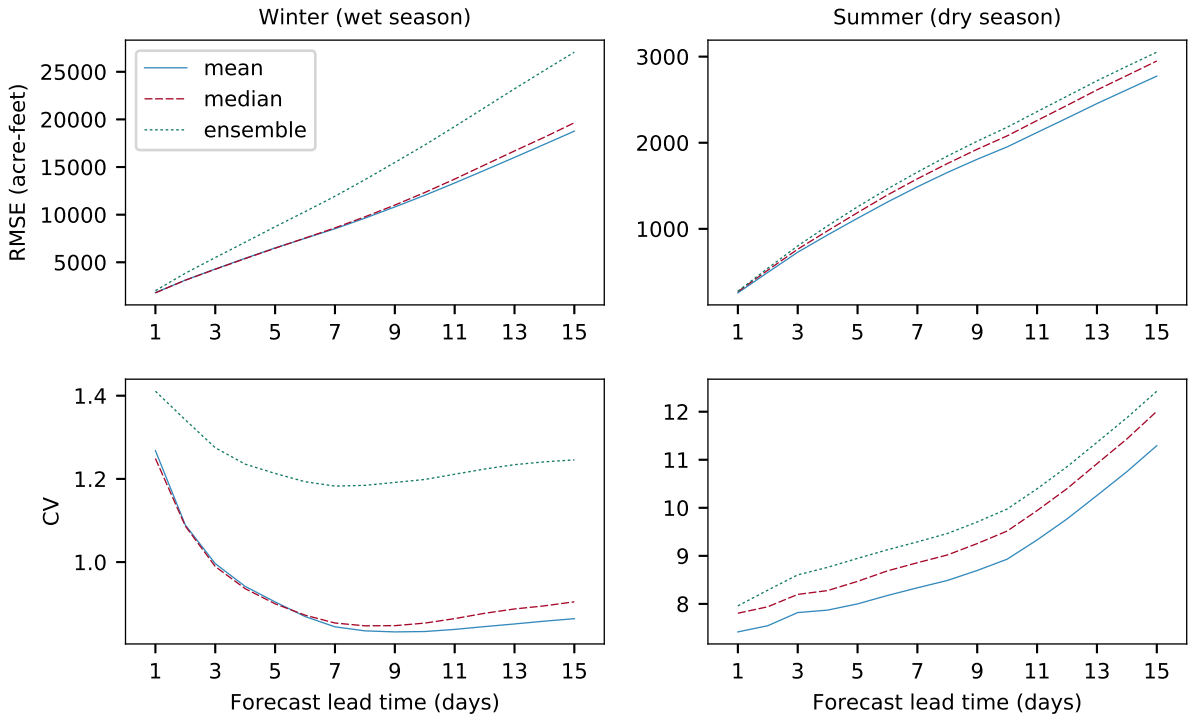

Figure A.20: RMSE and CV of the forecasts at HEAC1 (Big Sulphur Springs)

A.5 Rank Histogram

Rank histograms for each lead time in the available forecast horizon for the ensemble forecasts grouped by season. All of the locations show a similar pattern for the winter forecasts. The first and last position are the first and second most frequent locations respectively, with the other positions showing an increasing trend with higher rank. The summer forecasts at LM_Inflow, HOPC1, and UKAC1 show this same pattern. The summer forecasts at CDLC1 show the last position with higher frequency than the first for the lead times up to 12 days. The summer forecasts at HEAC1 consistently show position 25 as the most frequent followed by the last position and the first position as the second and third most frequent respectively. All lead times show the most frequent position of the observed flow to be at the first position indicating that the observed flow is lower than all of the forecasted values. The second most common position is at the other end of the list (position 36) indicating that the observed value was higher than all of the forecasted values. These two positions occur with an overwhelming majority.

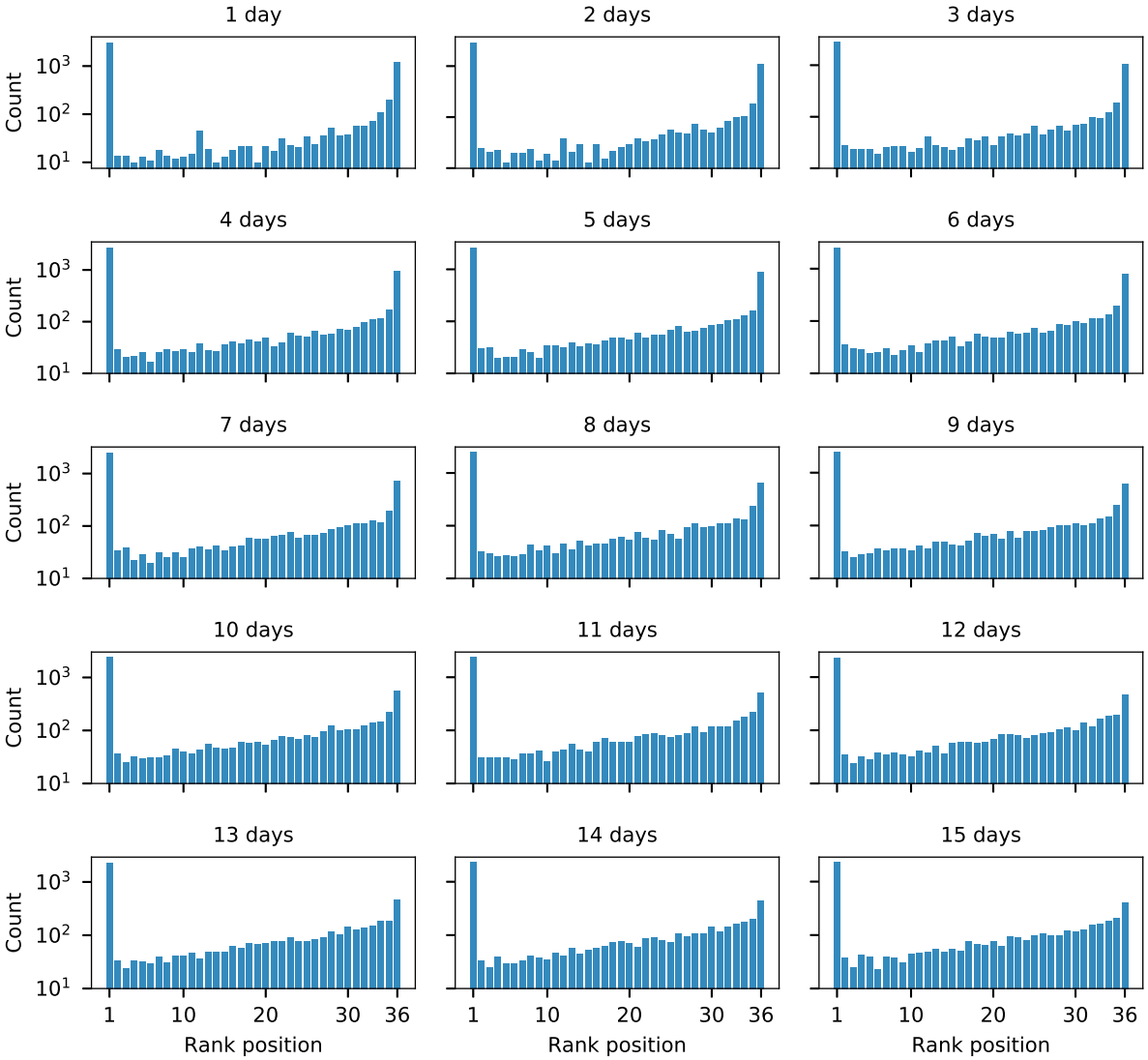

Figure A.21: Rank histogram for all lead times for the winter forecasts at LM_Inflow (Calpella), shown with a log scale.

Figure A.22: Rank histogram for all lead times for the summer forecasts at LM_Inflow (Calpella), shown with a log scale.

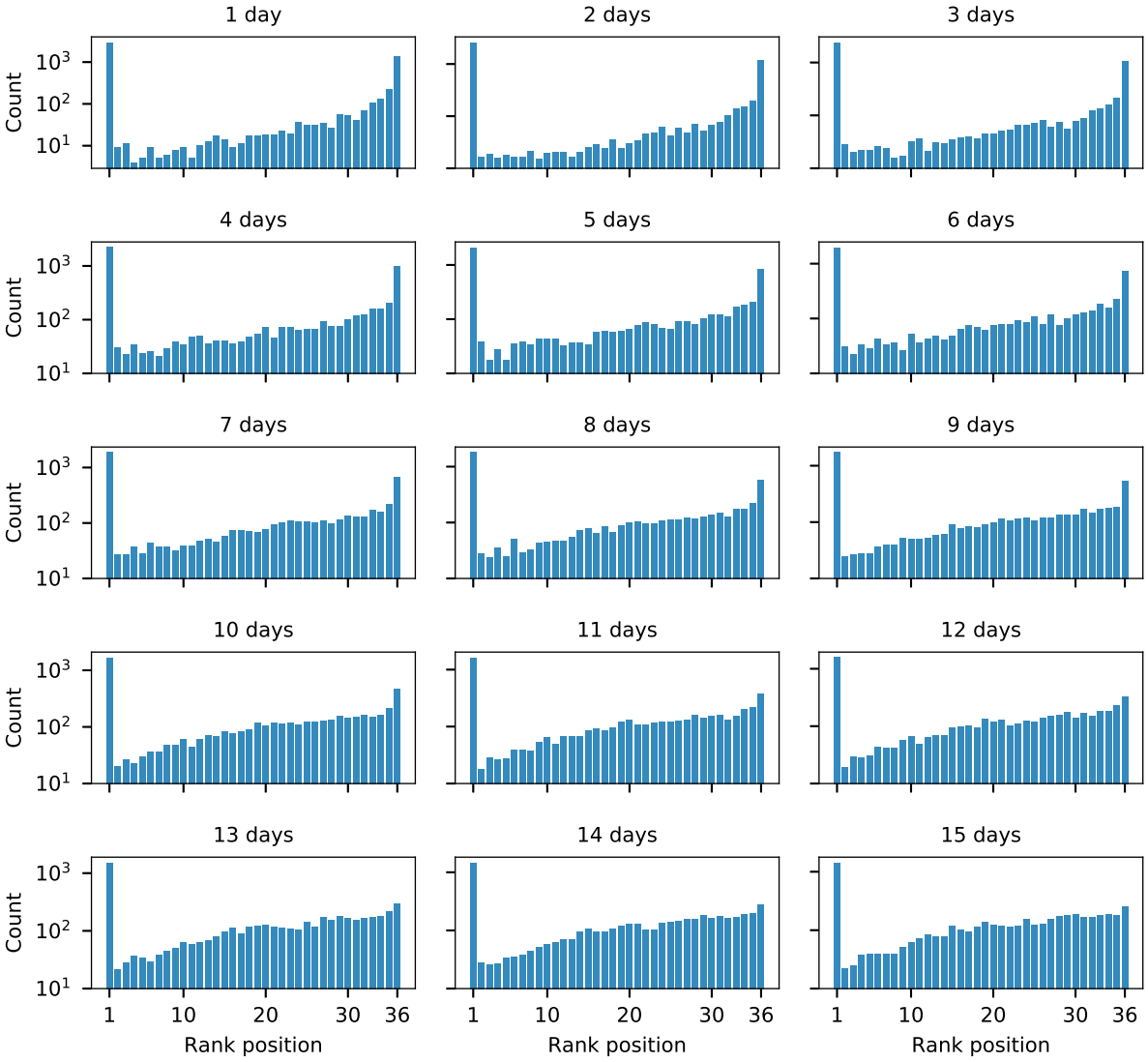

Figure A.23: Rank histogram for all lead times for the winter forecasts at UKAC1 (West Fork), shown with a log scale.

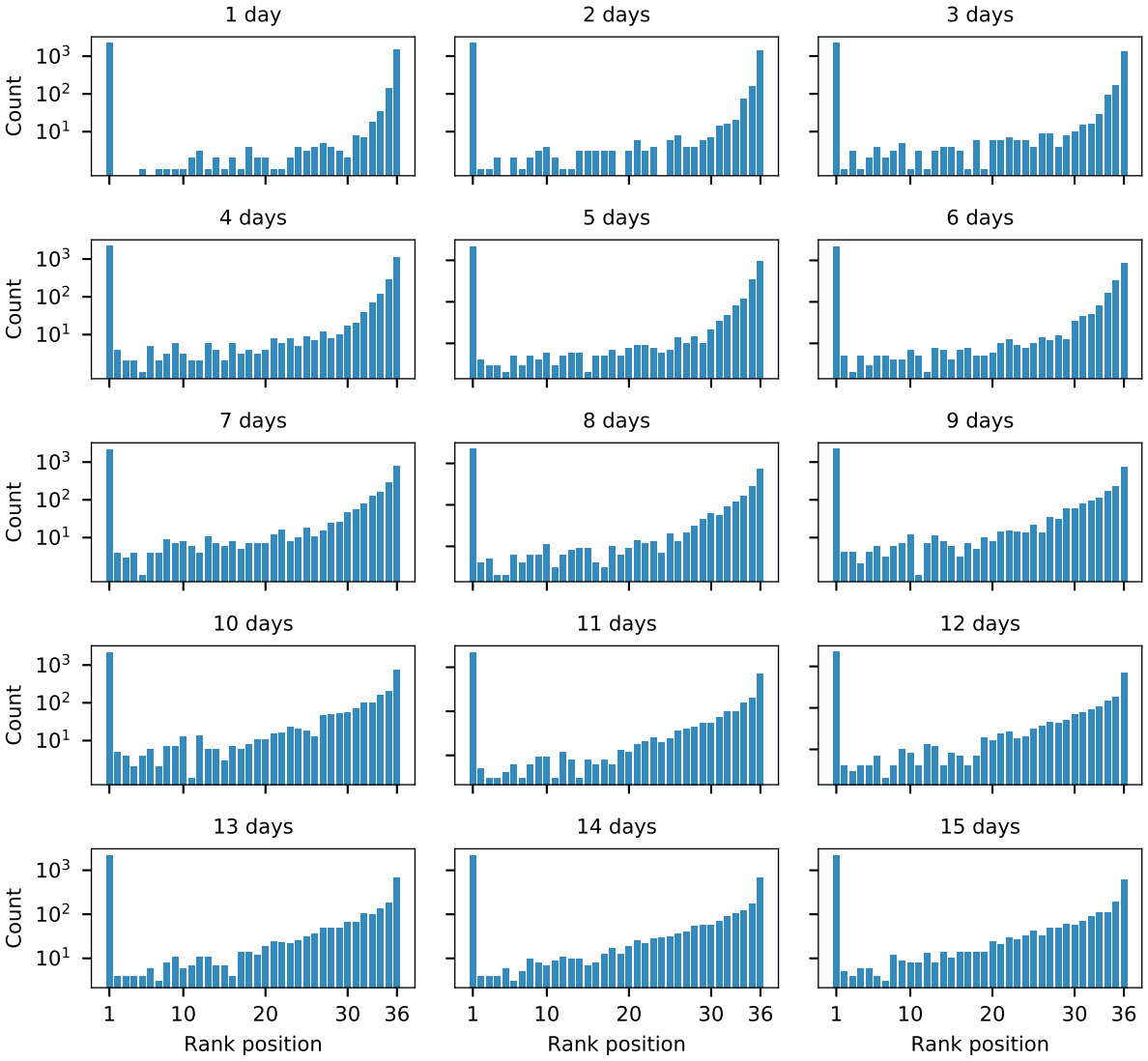

Figure A.24: Rank histogram for all lead times for the summer forecasts at UKAC1 (West Fork), shown with a log scale.

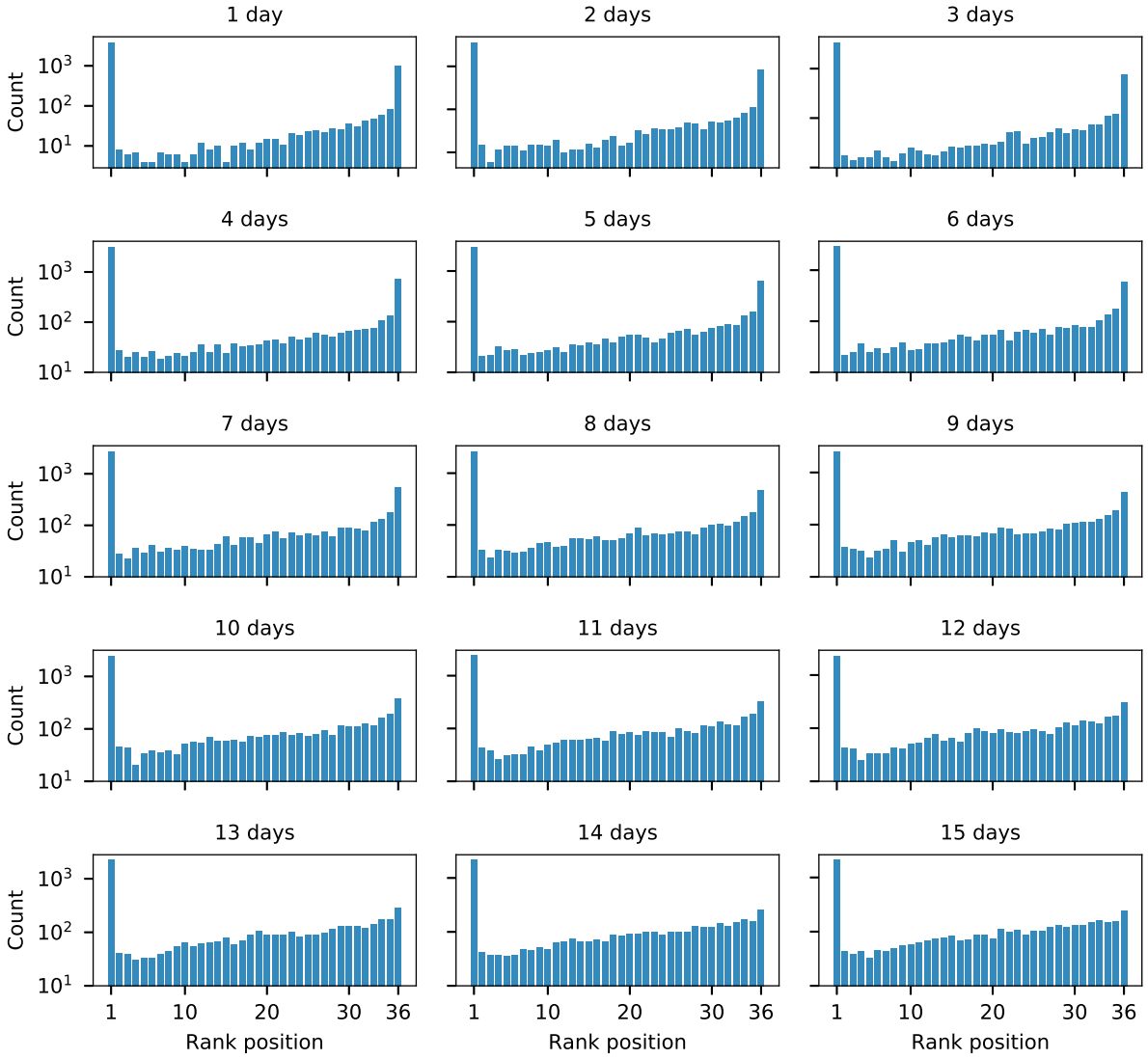

Figure A.25: Rank histogram for all lead times for the winter forecasts at HOPC1 (Ukiah), shown with a log scale.

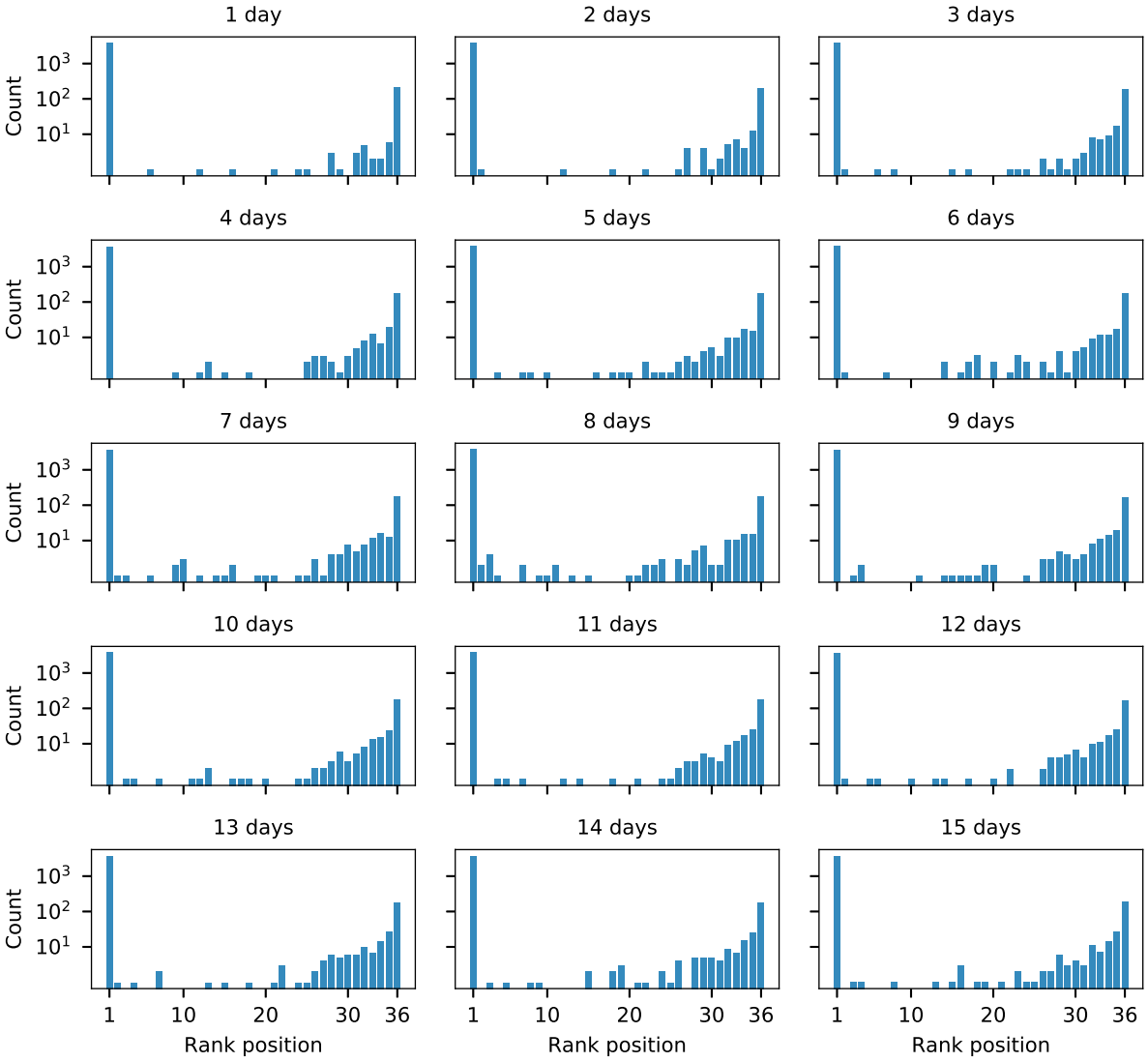

Figure A.26: Rank histogram for all lead times for the summer forecasts at HOPC1 (Ukiah), shown with a log scale.

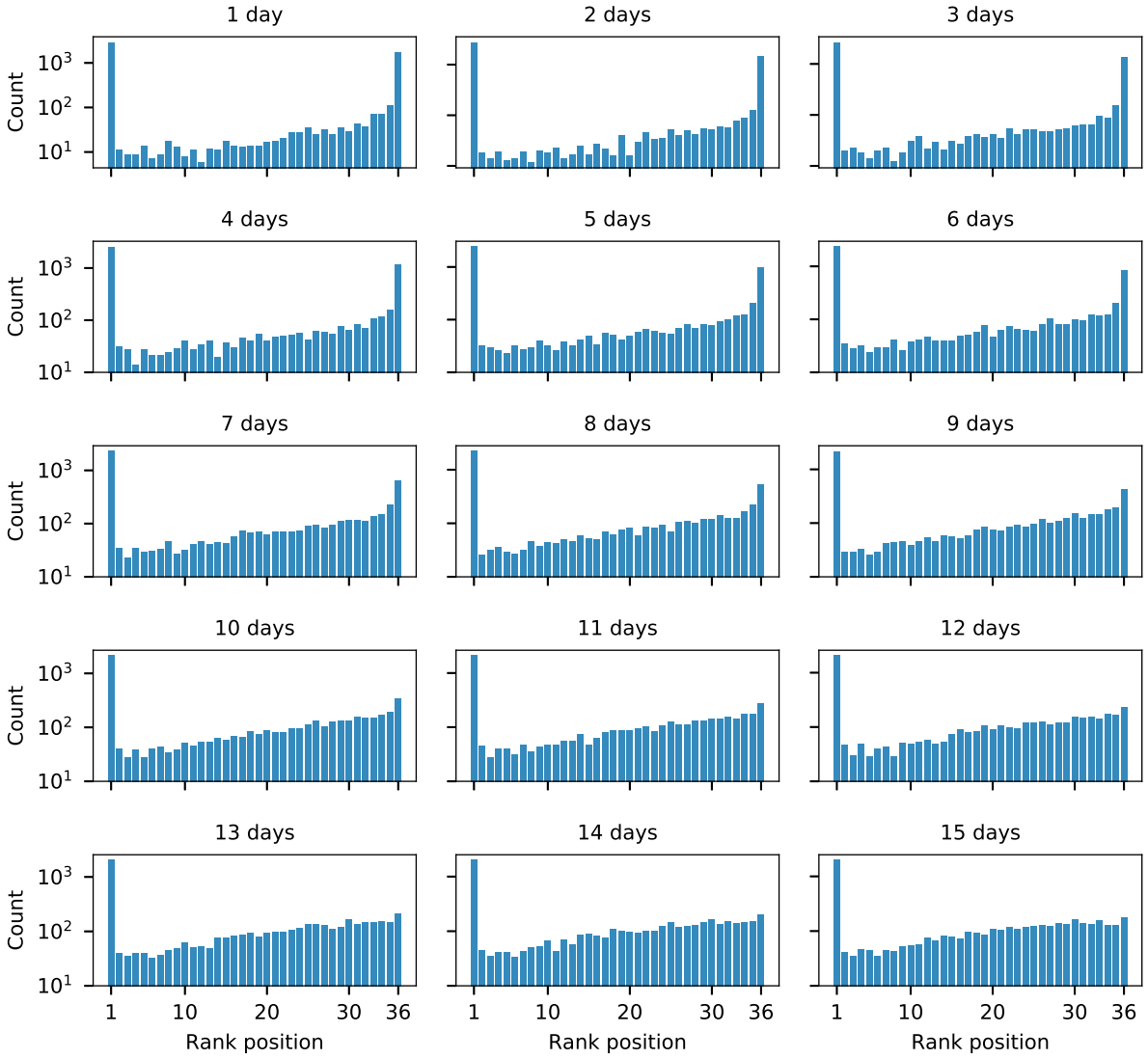

Figure A.27: Rank histogram for all lead times for the winter forecasts at CDLC1 (Cloverdale), shown with a log scale.

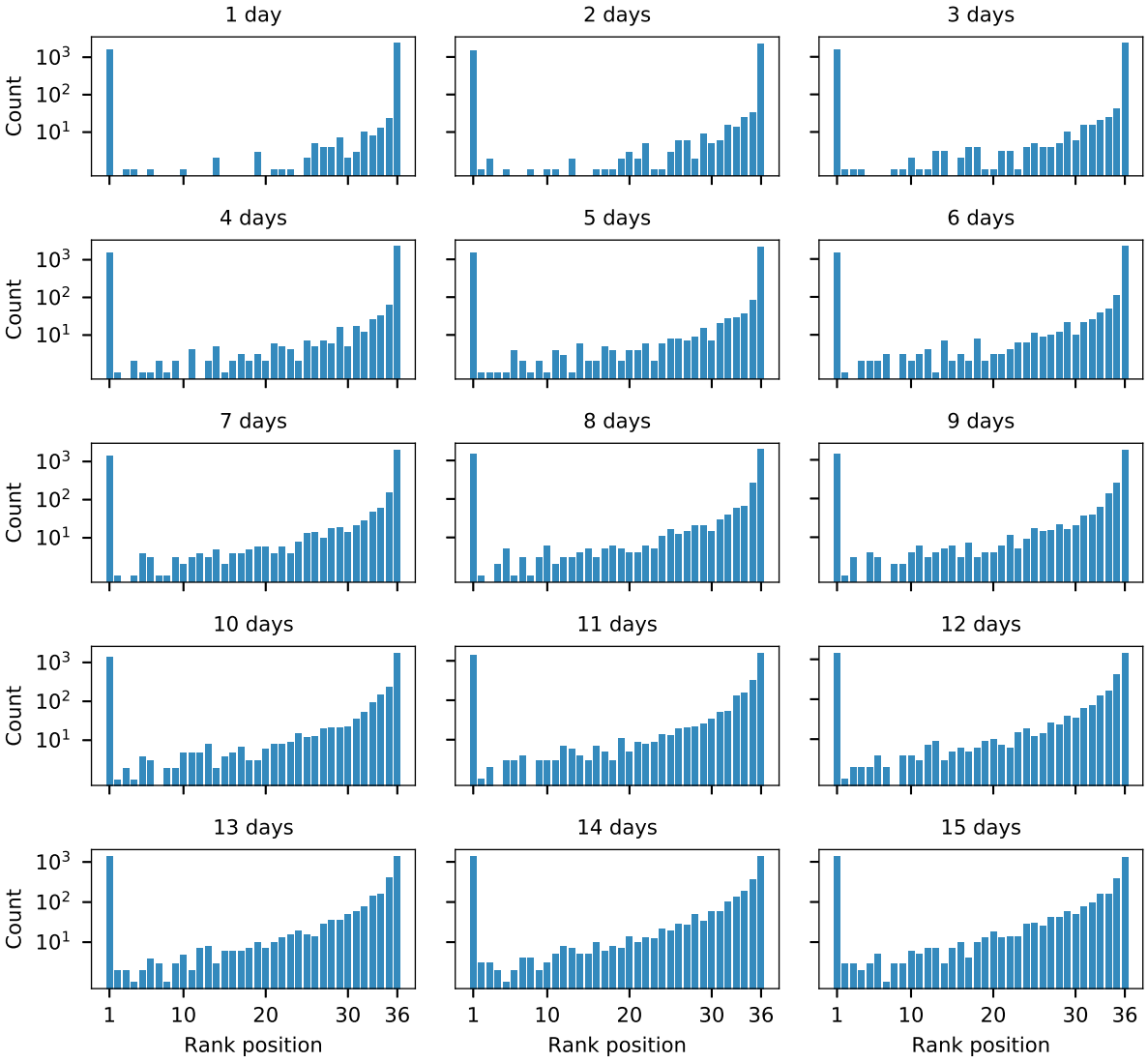

Figure A.28: Rank histogram for all lead times for the summer forecasts at CDLC1 (Cloverdale), shown with a log scale.

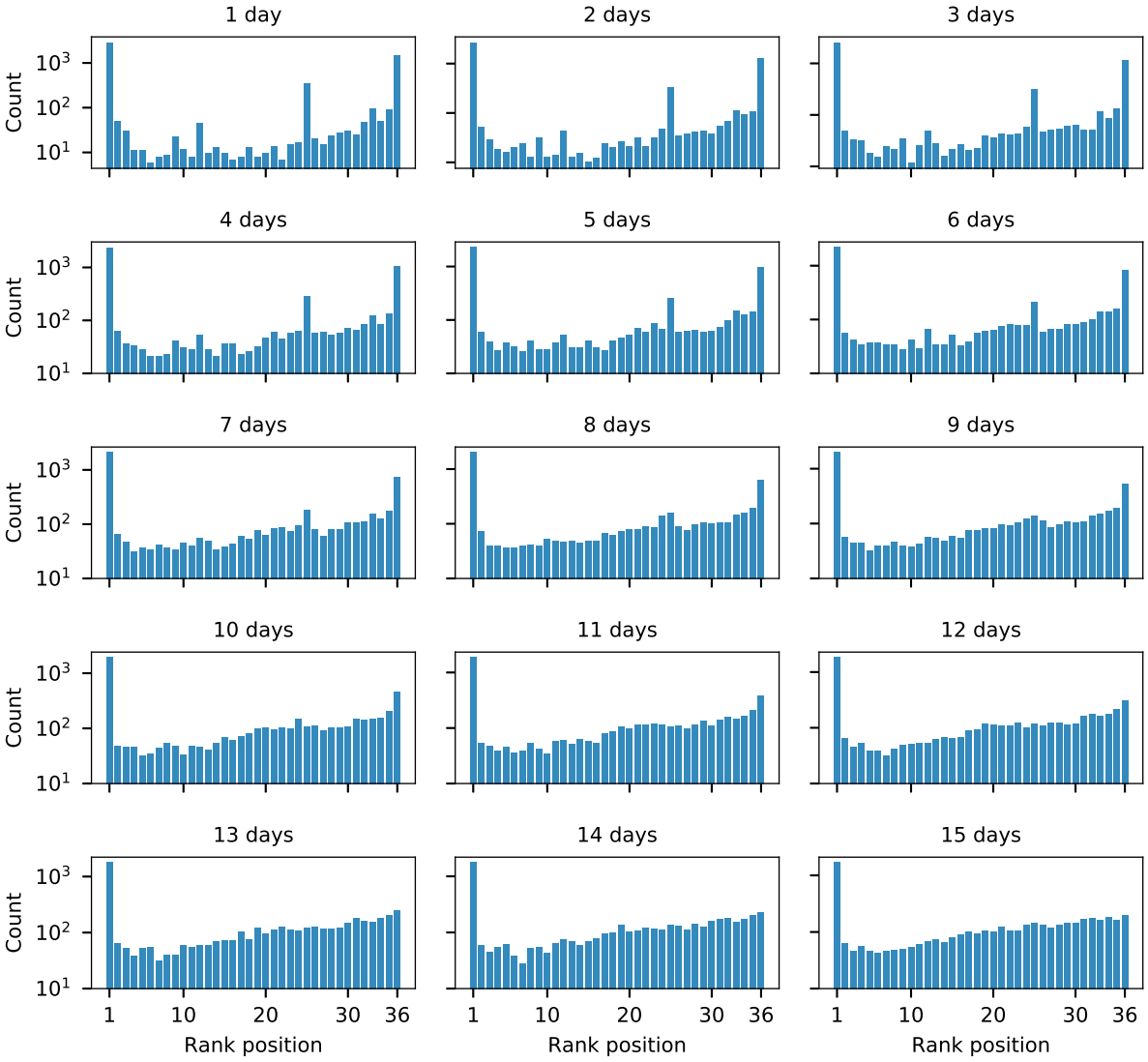

Figure A.29: Rank histogram for all lead times for the winter forecasts at HEAC1 (Big Sulphur Springs), shown with a log scale.

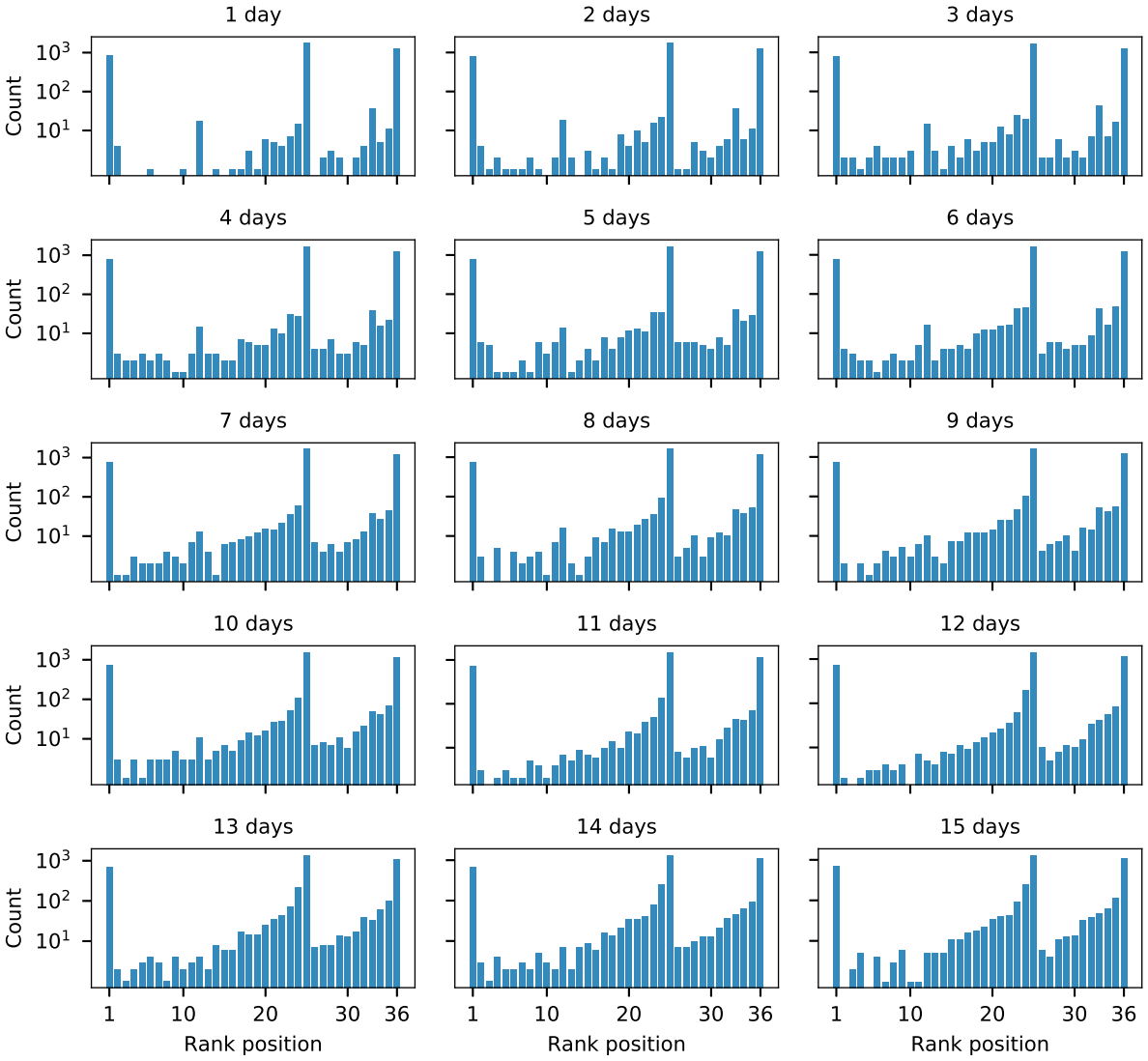

Figure A.30: Rank histogram for all lead times for the summer forecasts at HEAC1 (Cloverdale), shown with a log scale.

A.6 Distance From Ensemble

Figures A.31 - A.35 show the percentage of ensembles which do not span the observation grouped by season. The winter forecasts show a general trend of decreasing percentage of ensembles which do not span the observation. The summer forecasts show the reverse trend. In both seasons, the frequency of overprediction is higher than underprediction except for the summer forecasts at HEAC1. A tolerance is used when determining if the observation falls within the span of the ensemble. A tolerance of 10 af is used for the winter forecasts and a tolerance of 1 af is used for the summer forecasts.

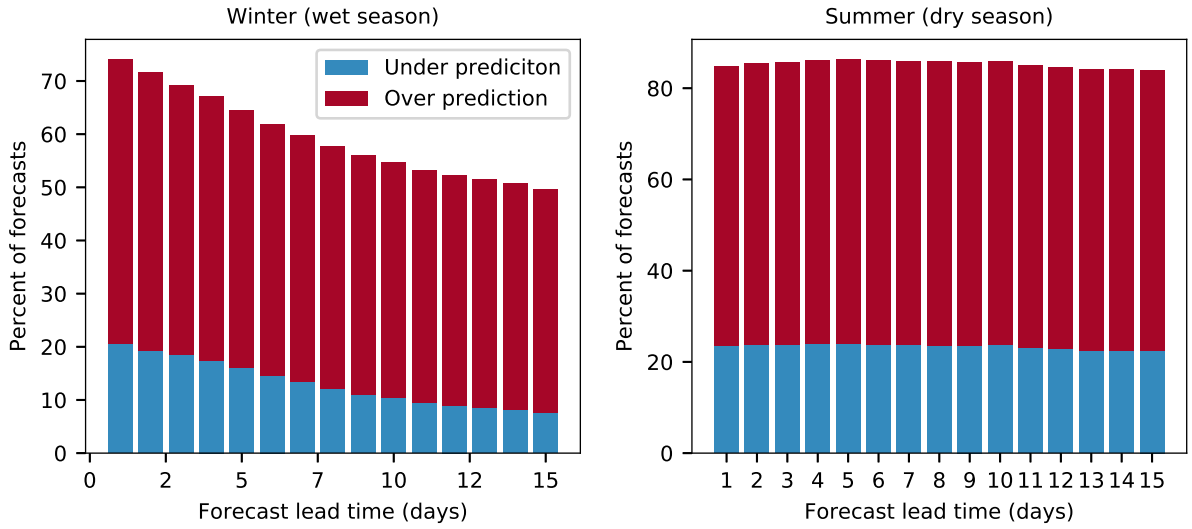

Figure A.31: Percentage of the ensembles which do not contain the observation for the forecasts at LM_Inflow (Calpella)

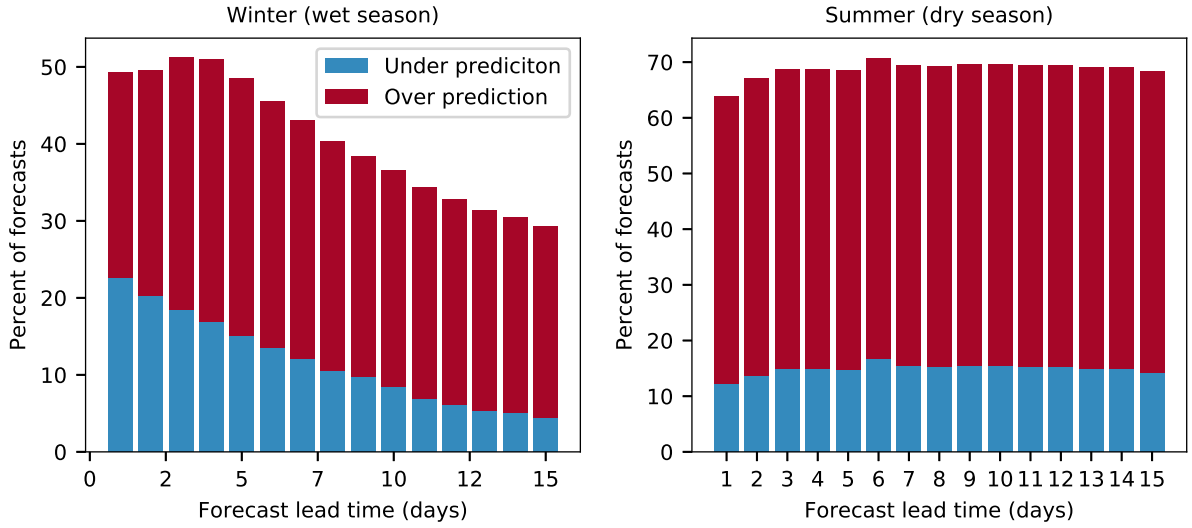

Figure A.32: Percentage of the ensembles which do not contain the observation for the forecasts at UKAC1 (Calpella)

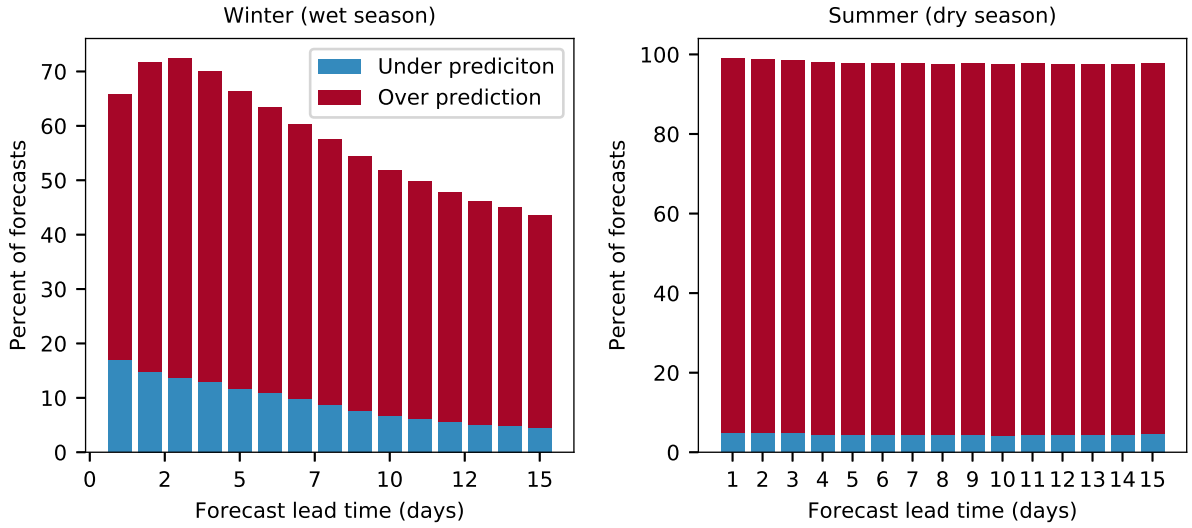

Figure A.33: Percentage of the ensembles which do not contain the observation for the forecasts at HOPC1 (Ukiah)

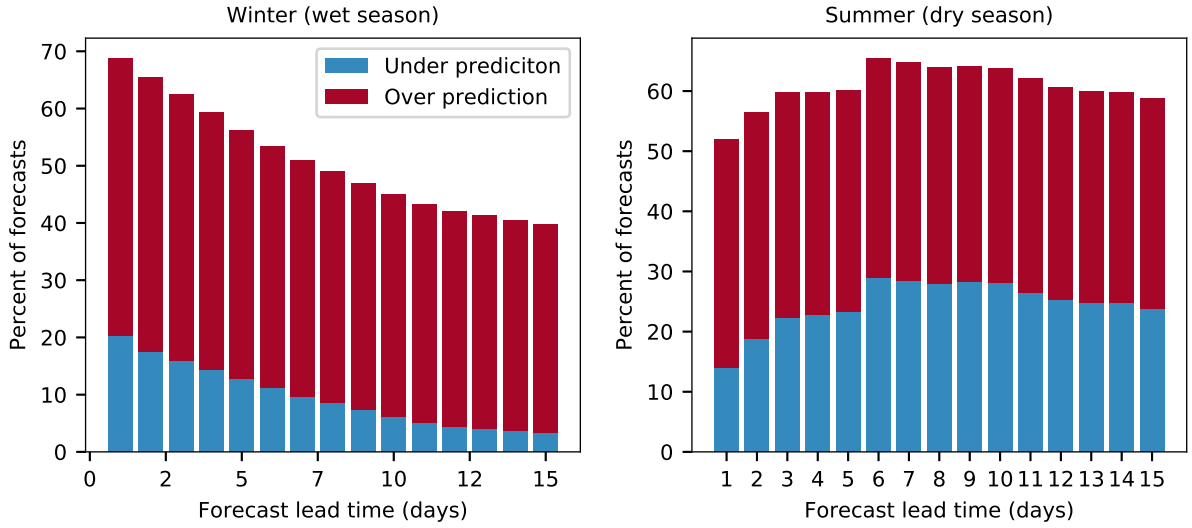

Figure A.34: Percentage of the ensembles which do not contain the observation for the forecasts at CDLC1 (Cloverdale)

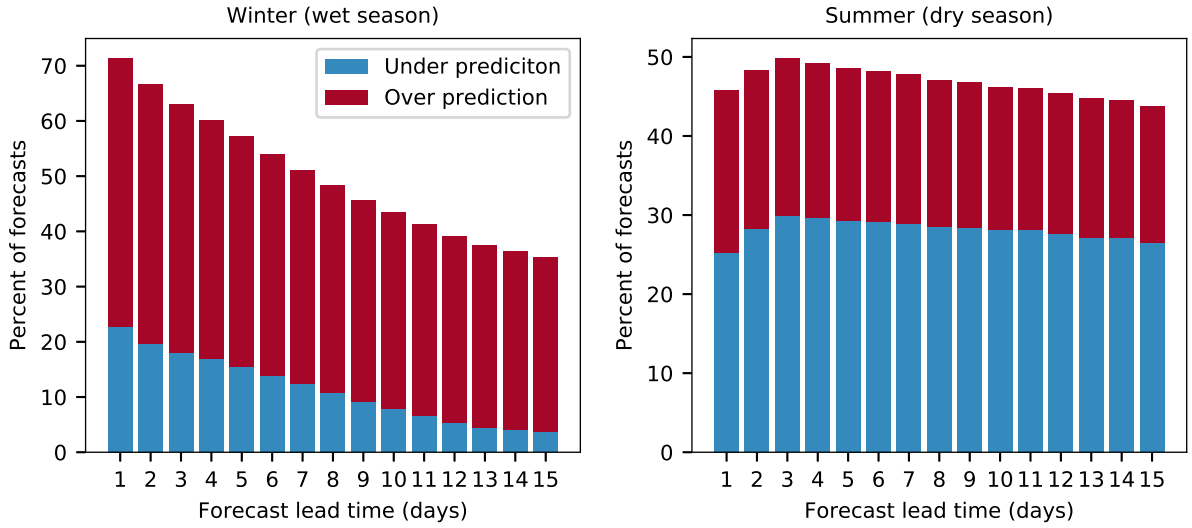

Figure A.35: Percentage of the ensembles which do not contain the observation for the forecasts at HEAC1 (Big Sulphur Springs)

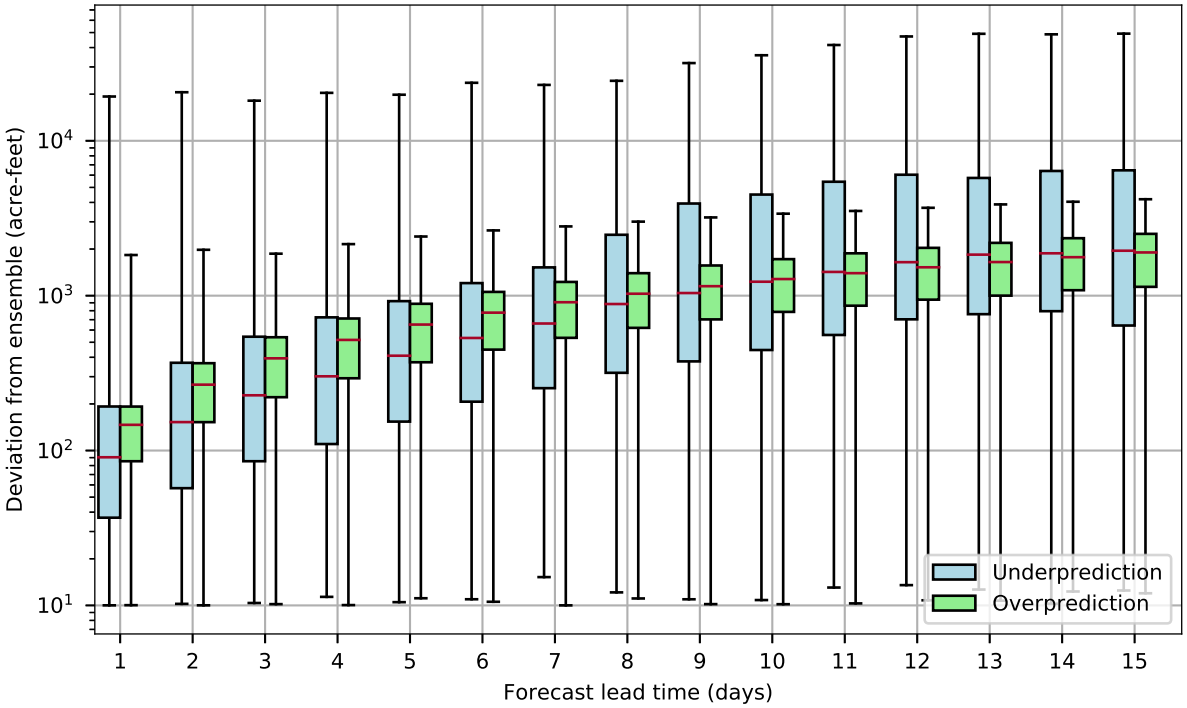

Figure A.36: Box plot diagram of the deviations outside of the ensemble for LM_Inflow (Calpella)

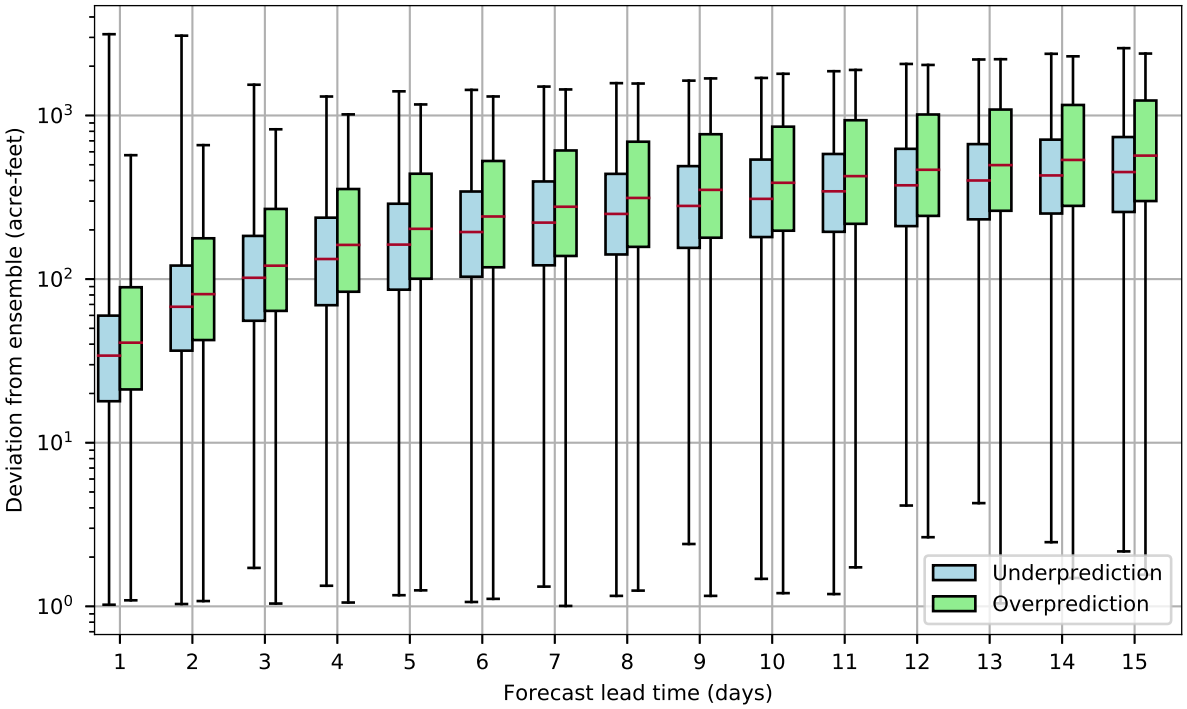

Figure A.37: Box plot diagram of the deviations outside of the ensemble for LM_Inflow (Calpella)

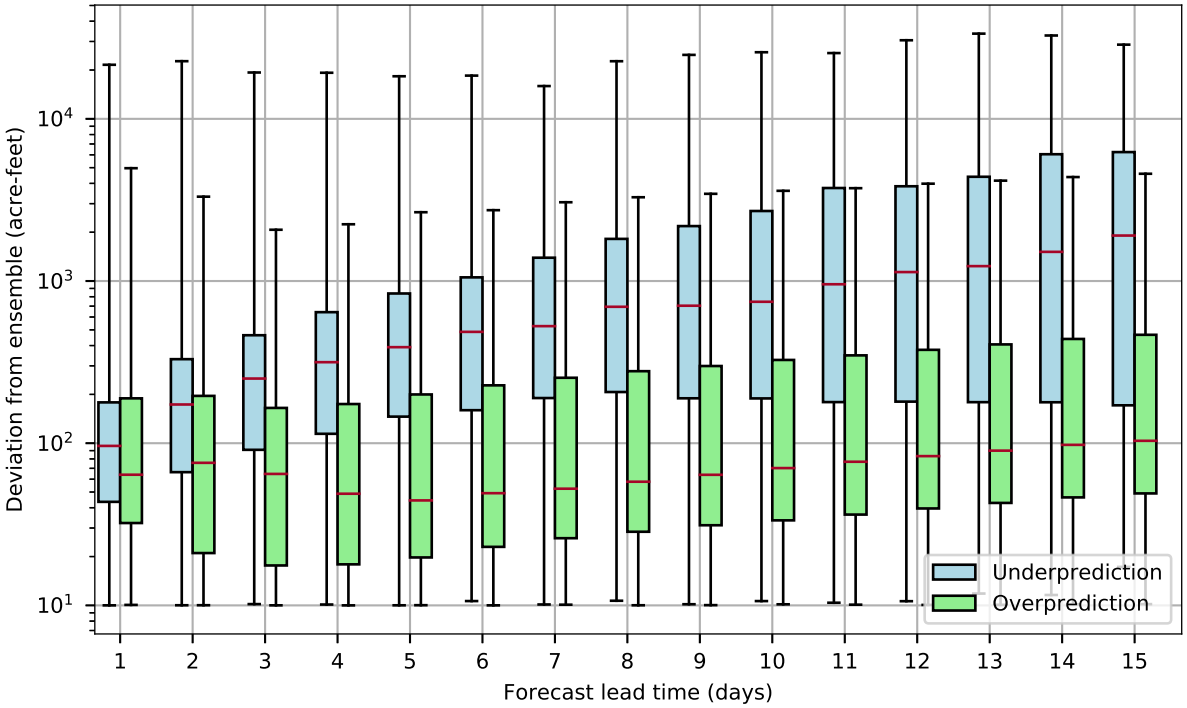

Figure A.38: Box plot diagram of the deviations outside of the ensemble for UKAC1 (West Fork)

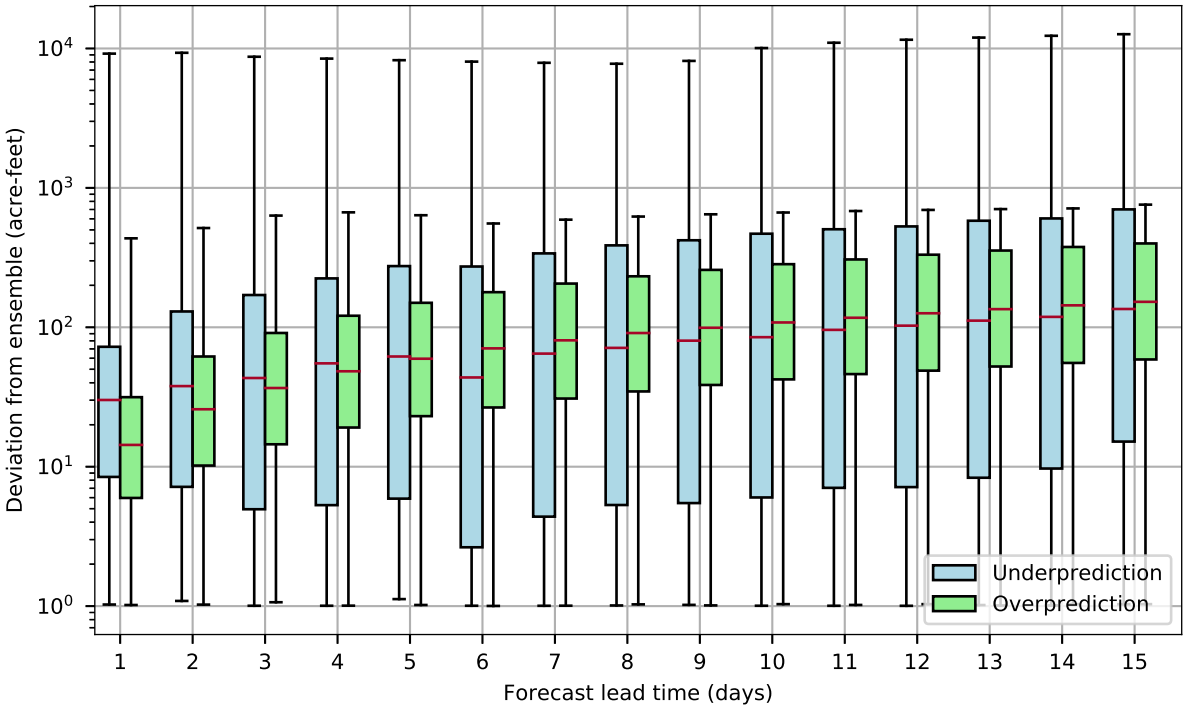

Figure A.39: Box plot diagram of the deviations outside of the ensemble for UKAC1 (West Fork)

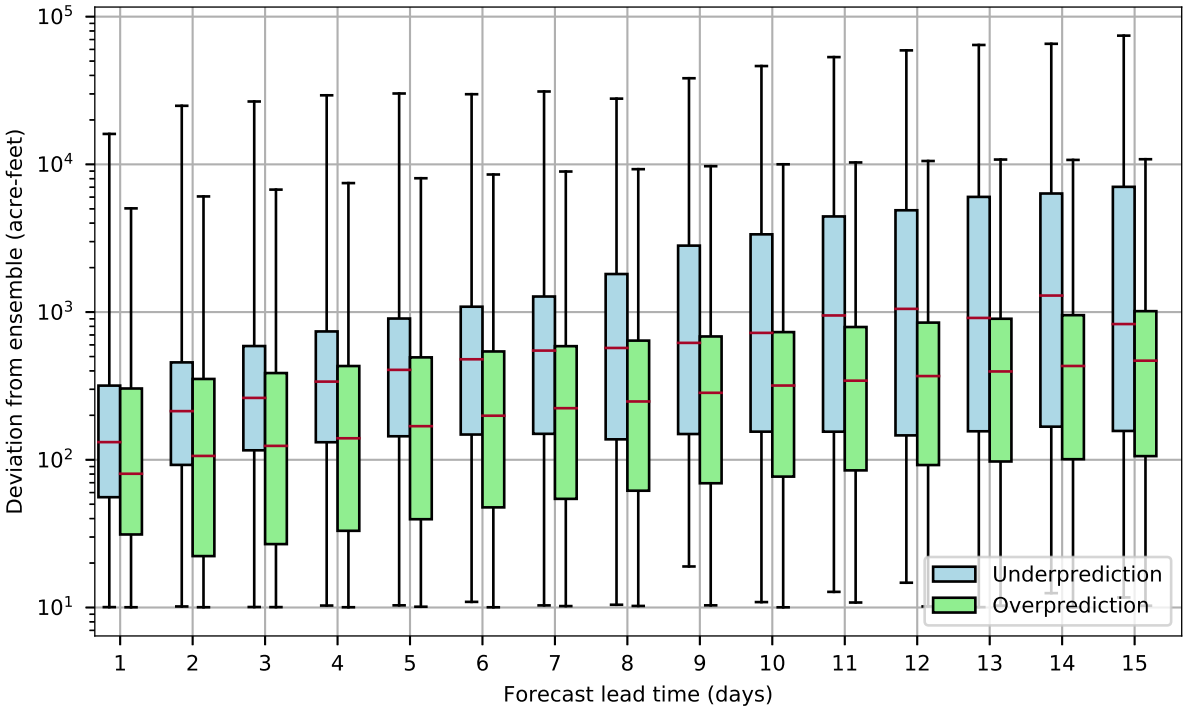

Figure A.40: Box plot diagram of the deviations outside of the ensemble for HOPC1 (West Fork)

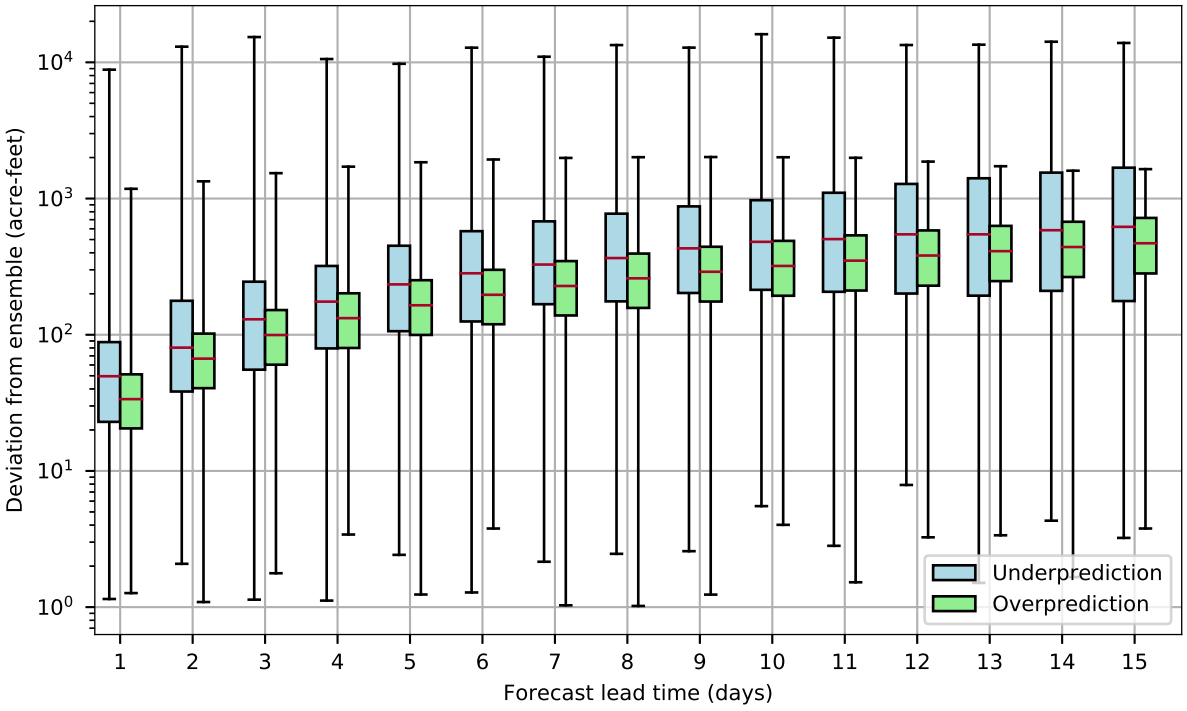

Figure A.41: Box plot diagram of the deviations outside of the ensemble for HOPC1 (West Fork)

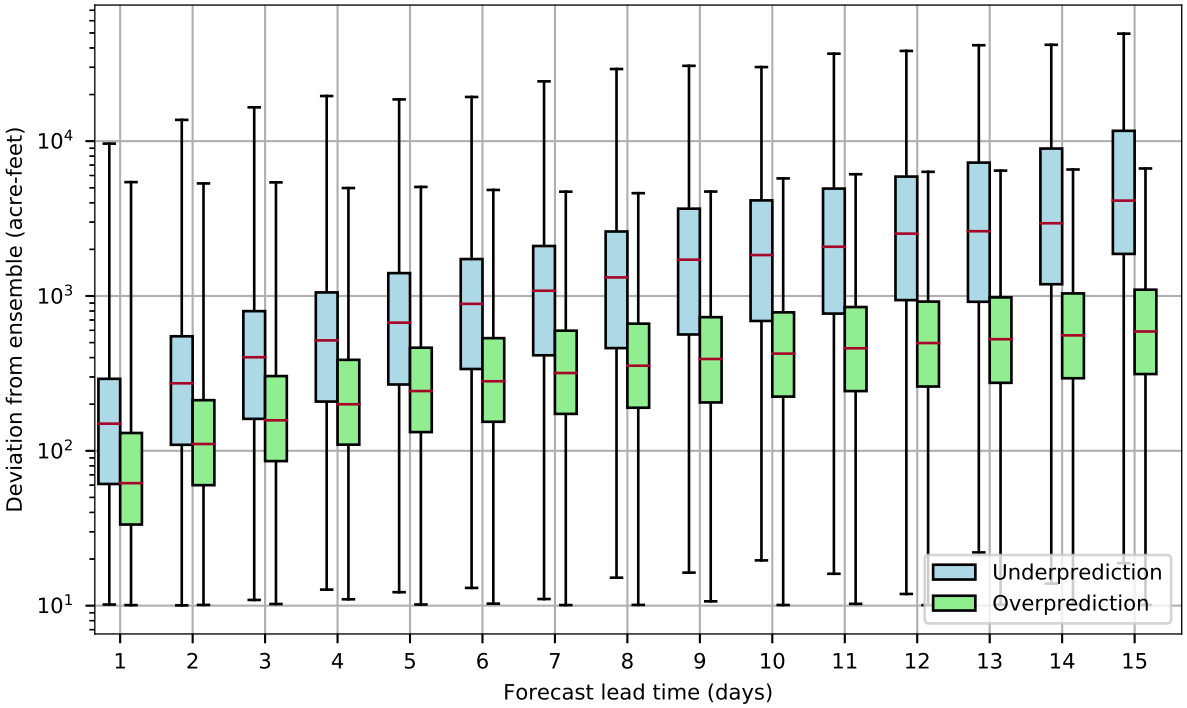

Figure A.42: Box plot diagram of the deviations outside of the ensemble for CDLC1 (West Fork)

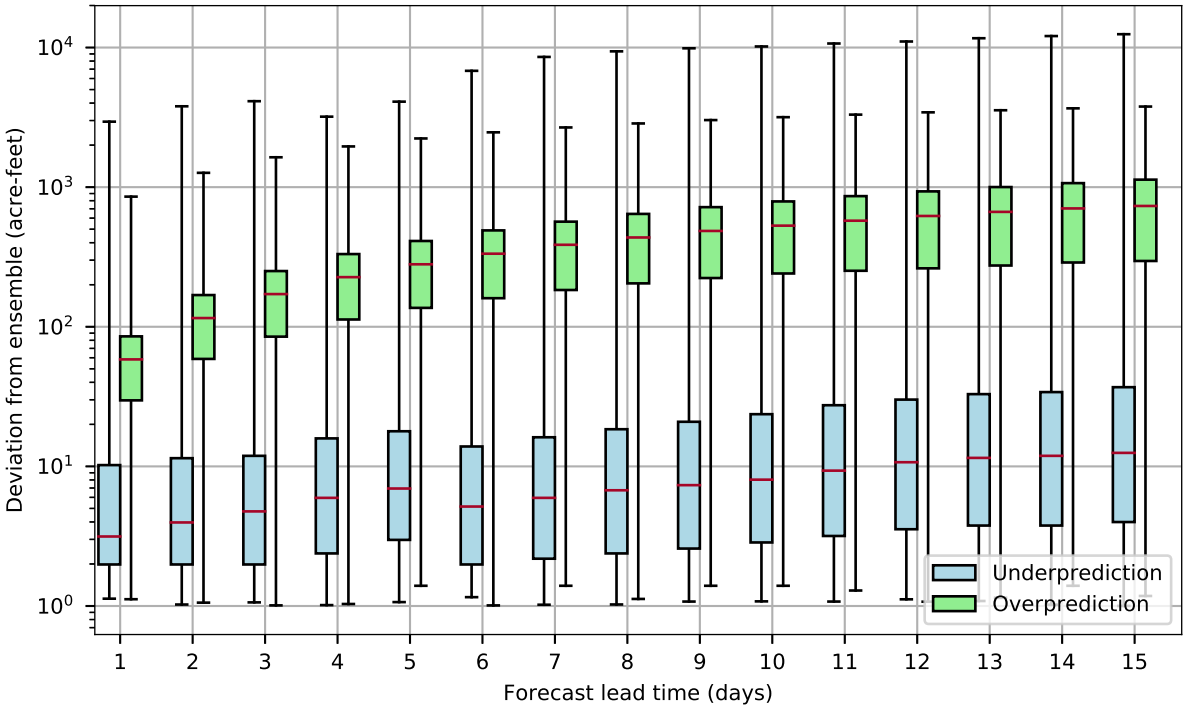

Figure A.43: Box plot diagram of the deviations outside of the ensemble for CDLC1 (West Fork)

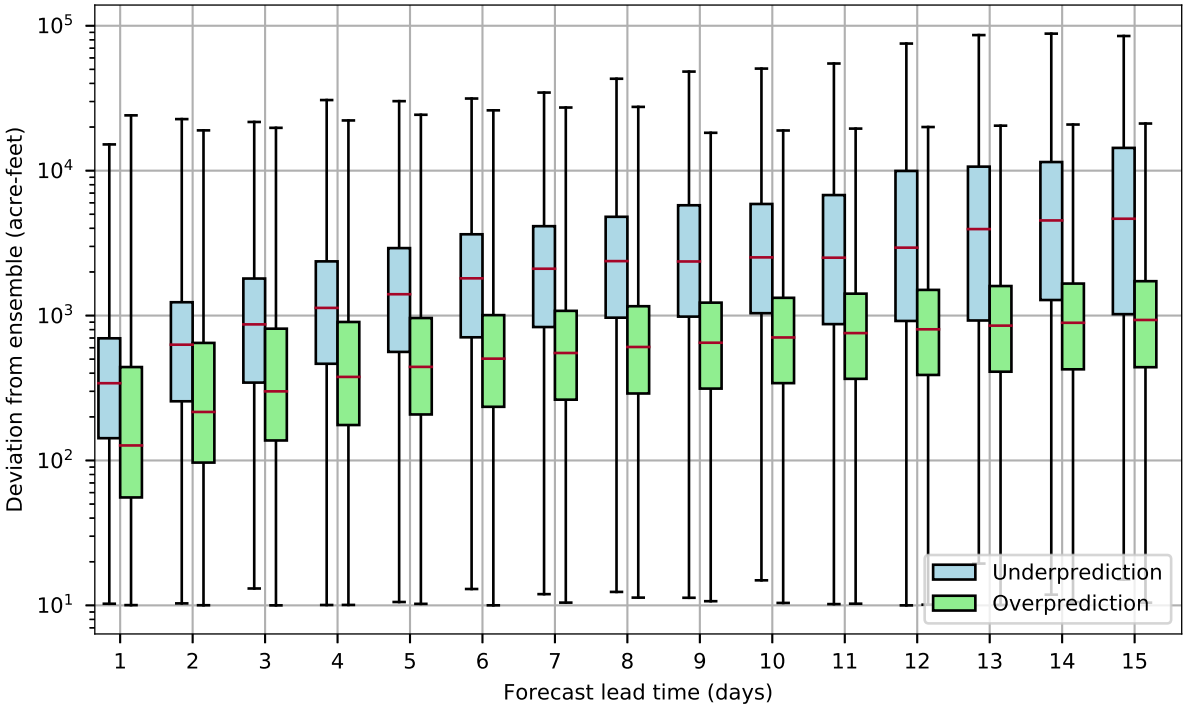

Figure A.44: Box plot diagram of the deviations outside of the ensemble for HEAC1 (West Fork)

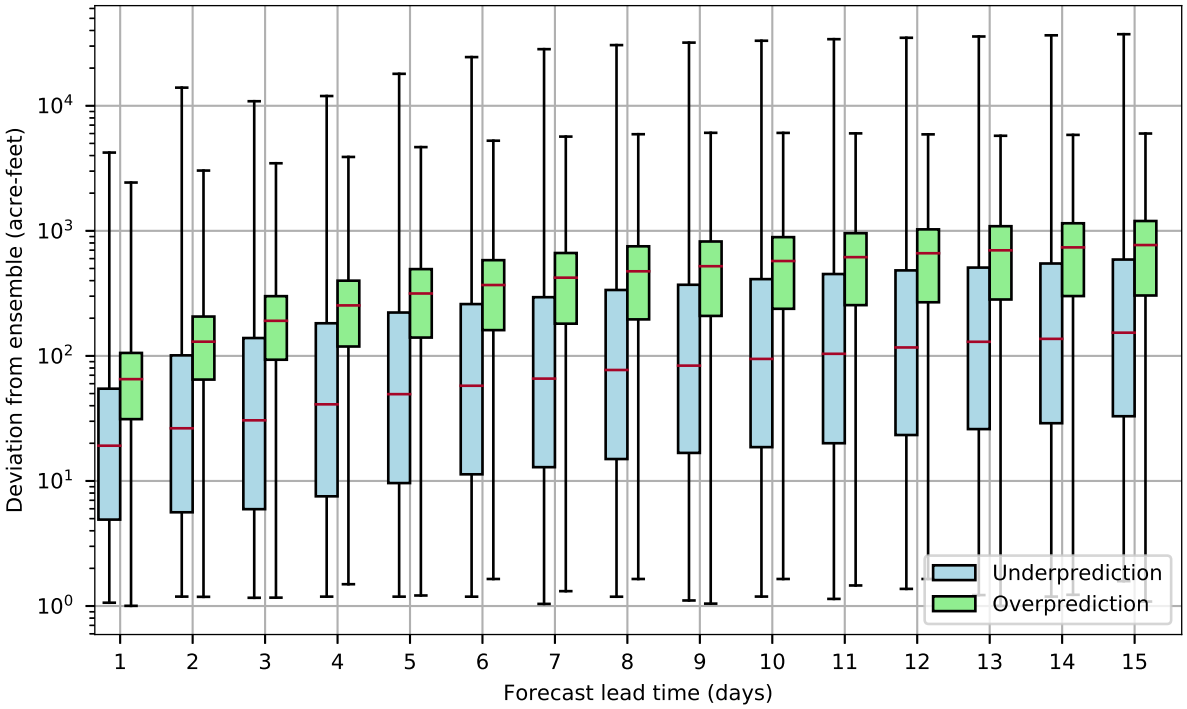

Figure A.45: Box plot diagram of the deviations outside of the ensemble for HEAC1 (West Fork)

Appendix B

Additional Performance Measures for the Simulations with Forecasts

B.1 10-day Forecast Horizon

Table B.1: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 10 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	41.36	24
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93113	5.19	37.12	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98105	3.42	40.48	28
ME	0.9989	2.00	1482.15	1.000	0.00	0.00	0.98854	2.82	32.24	29
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	1.00000	0.00	0.00	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92925	5.20	37.78	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98201	3.71	39.40	27
ME	0.9989	2.00	1501.35	1.000	0.00	0.00	0.98755	3.22	34.45	28
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	38.86	24
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92707	5.20	38.34	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98214	4.16	41.81	27
ME	0.9989	2.00	1372.67	1.000	0.00	0.00	0.98646	3.20	37.72	29
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	35.62	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92410	5.14	37.28	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98082	4.15	42.41	26
ME	0.9988	1.80	1315.65	1.000	0.00	0.00	0.98673	3.12	32.77	29
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.2: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 10 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9989	1.00	377.85	1.000	0.00	0.00	0.99993	1.00	34.12	26
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92251	5.31	37.44	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98181	3.77	39.91	28
ME	0.9982	1.40	1070.02	1.000	0.00	0.00	0.98904	2.79	35.20	29
Environmental contour 800										
PF	0.9987	1.00	356.18	1.000	0.00	0.00	0.99848	1.77	20.65	24
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91554	5.87	40.78	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97088	3.69	37.78	27
ME	0.9983	1.44	1489.72	1.000	0.00	0.00	0.98732	4.13	37.62	29
Environmental contour 1000										
PF	0.9988	1.00	382.68	1.000	0.00	0.00	0.99779	3.19	22.49	24
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90481	5.96	42.98	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95361	3.22	31.43	27
ME	0.9980	1.36	1061.24	1.000	0.00	0.00	0.98735	4.30	38.77	28
Environmental contour 1100										
PF	0.9987	1.00	365.25	1.000	0.00	0.00	0.99736	3.81	24.57	25
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89722	5.52	42.35	27
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94199	3.48	29.86	27
ME	0.9983	1.30	1461.48	1.000	0.00	0.00	0.98600	3.66	41.29	29
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

Table B.3: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 10 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	36.60	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93707	4.38	34.82	32
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98633	2.98	39.21	30
ME	0.9974	1.54	3416.10	1.000	0.00	0.00	0.99571	1.67	41.71	31
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	37.91	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93651	4.71	35.67	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98336	3.00	38.25	31
ME	0.9968	1.60	3346.26	1.000	0.00	0.00	0.99422	2.16	46.57	31
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	1.00000	0.00	0.00	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93578	4.91	34.74	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98458	3.22	44.26	31
ME	0.9968	1.60	3699.35	1.000	0.00	0.00	0.99419	2.15	46.75	31
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99997	1.00	32.76	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93159	4.50	35.91	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.98419	2.85	38.38	29
ME	0.9971	1.57	3478.48	1.000	0.00	0.00	0.99554	1.67	37.80	30
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.4: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 10 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9970	1.00	314.93	1.000	0.00	0.00	1.00000	0.00	0.00	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.93182	4.98	35.23	28
IP	0.9999	1.00	495.96	1.000	0.00	0.00	0.98359	3.82	41.42	31
ME	0.9952	1.29	3004.30	1.000	0.00	0.00	0.99211	1.99	44.22	31
Environmental contour 800										
PF	0.9971	1.00	360.05	1.000	0.00	0.00	1.00000	0.00	0.00	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92845	5.69	40.65	32
IP	0.9999	1.00	497.16	1.000	0.00	0.00	0.98006	3.28	35.41	29
ME	0.9954	1.30	2550.87	1.000	0.00	0.00	0.99181	1.78	43.45	31
Environmental contour 1000										
PF	0.9972	1.00	360.11	1.000	0.00	0.00	1.00000	0.00	0.00	26
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.92116	5.74	40.32	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96213	4.01	32.74	29
ME	0.9951	1.32	2614.88	1.000	0.00	0.00	0.99284	2.11	43.15	30
Environmental contour 1100										
PF	0.9972	1.00	364.45	1.000	0.00	0.00	0.99993	1.00	22.88	26
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91465	5.29	40.24	28
IP	0.9999	1.00	144.12	1.000	0.00	0.00	0.95774	3.22	29.55	29
ME	0.9954	1.30	3006.27	1.000	0.00	0.00	0.99194	2.18	39.42	31
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

B.2 5-day Forecast Horizon

Table B.5: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 5 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	43.92	28
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91089	4.68	39.23	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97114	4.01	43.64	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.97088	3.43	43.60	28
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99865	1.46	19.62	28
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90828	4.75	40.14	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97029	4.19	44.15	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96916	3.97	44.20	27
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99868	1.48	18.33	28
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90660	4.67	39.82	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96857	4.14	44.27	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96906	3.89	44.58	28
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99845	1.62	20.76	28
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90396	4.43	39.62	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96695	3.81	42.41	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96811	3.98	45.61	28
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.6: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 5 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9997	1.00	449.77	1.000	0.00	0.00	0.99990	1.00	32.88	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91026	4.45	38.72	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97154	3.94	43.25	29
ME	0.9999	1.00	498.75	1.000	0.00	0.00	0.97395	3.07	45.76	28
Environmental contour 800										
PF	0.9997	1.00	498.78	1.000	0.00	0.00	0.99739	4.16	27.72	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90125	5.07	44.28	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95404	3.56	38.52	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.97227	3.56	47.55	28
Environmental contour 1000										
PF	0.9997	1.00	499.47	1.000	0.00	0.00	0.99574	2.87	27.08	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88973	5.00	44.67	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.93516	3.87	34.95	29
ME	0.9997	1.00	349.02	1.000	0.00	0.00	0.96817	3.92	46.12	28
Environmental contour 1100										
PF	0.9997	1.00	498.73	1.000	0.00	0.00	0.99145	2.27	22.61	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88035	4.98	44.19	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.92485	4.02	34.62	28
ME	0.9997	1.00	202.84	1.000	0.00	0.00	0.96431	3.37	43.38	28
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

Table B.7: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 5 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	43.26	28
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91630	4.62	37.33	33
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97395	3.72	45.34	31
ME	0.9993	1.00	1010.98	1.000	0.00	0.00	0.97788	2.73	48.73	31
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	43.35	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91736	4.80	38.63	33
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97520	3.53	43.58	32
ME	0.9995	1.00	1182.64	1.000	0.00	0.00	0.97897	2.43	50.04	31
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	43.73	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91485	4.23	38.16	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97517	3.26	41.56	31
ME	0.9993	1.00	1306.25	1.000	0.00	0.00	0.97633	2.58	49.25	31
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99993	1.00	41.50	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91294	4.17	38.11	32
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97398	3.01	44.16	30
ME	0.9993	1.00	1387.44	1.000	0.00	0.00	0.97828	2.26	47.22	31
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.8: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 5 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9992	1.00	302.42	1.000	0.00	0.00	0.99990	1.00	32.25	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.91333	4.60	38.72	32
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.97329	3.38	43.89	34
ME	0.9991	1.00	1229.29	1.000	0.00	0.00	0.97854	2.51	47.25	34
Environmental contour 800										
PF	0.9991	1.00	283.49	1.000	0.00	0.00	0.99990	1.00	52.60	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90607	5.01	43.40	34
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95764	3.43	36.36	33
ME	0.9989	1.00	1000.64	1.000	0.00	0.00	0.97818	2.30	46.48	36
Environmental contour 1000										
PF	0.9992	1.00	387.15	1.000	0.00	0.00	0.99927	1.16	20.09	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89653	5.23	44.20	33
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94341	3.91	34.57	33
ME	0.9988	1.00	1007.49	1.000	0.00	0.00	0.97761	2.50	47.17	34
Environmental contour 1100										
PF	0.9989	1.00	367.22	1.000	0.00	0.00	0.99851	1.12	21.53	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88900	5.16	44.77	34
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.93565	3.83	33.93	32
ME	0.9988	1.00	1313.08	1.000	0.00	0.00	0.97705	2.46	44.73	34
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

B.3 3-day Forecast Horizon

Table B.9: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 3 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.16	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90115	4.47	39.21	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96606	4.21	43.19	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96939	3.89	43.64	28
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99865	1.46	18.45	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89702	4.37	39.25	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96451	4.37	43.35	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96811	4.16	43.70	28
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99858	1.43	20.19	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89448	4.34	39.15	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96183	3.85	42.10	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96784	4.22	43.66	28
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99855	1.52	20.49	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89223	4.22	39.01	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96144	3.68	41.34	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96797	4.27	43.75	28
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.10: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 3 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.20	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89999	4.53	39.49	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96556	4.42	44.08	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96827	3.72	43.61	28
Environmental contour 800										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99567	3.36	29.72	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89253	4.95	43.70	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94932	3.65	40.88	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96504	4.14	45.69	28
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.98560	2.26	22.10	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.87870	4.79	44.59	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.92750	3.70	35.77	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.95817	3.52	42.18	28
Environmental contour 1100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.97378	1.86	20.52	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.86807	4.61	44.47	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.91492	3.77	35.40	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.94701	2.74	37.32	28
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

Table B.11: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 3 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.20	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90191	4.19	39.41	36
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96431	3.73	44.62	35
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96527	3.73	43.30	30
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.00	31
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89729	4.10	39.52	36
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96365	3.26	43.05	34
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96533	3.82	43.66	30
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99980	1.00	22.17	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89220	4.10	39.52	38
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95995	3.13	41.70	36
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96490	3.89	43.69	31
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99967	1.00	18.30	31
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89190	4.00	39.81	33
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95965	3.00	41.88	36
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96471	3.99	43.81	31
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.12: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 3 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	0.9999	1.00	498.94	1.000	0.00	0.00	0.99987	1.00	27.38	31
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.90145	4.29	39.02	34
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96520	3.55	41.41	36
ME	0.9997	1.00	498.48	1.000	0.00	0.00	0.96457	3.64	43.26	31
Environmental contour 800										
PF	0.9999	1.00	498.76	1.000	0.00	0.00	0.99792	1.62	22.02	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89187	4.67	43.48	37
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94367	3.64	37.68	33
ME	0.9997	1.00	498.81	1.000	0.00	0.00	0.96345	4.08	44.32	32
Environmental contour 1000										
PF	0.9999	1.00	498.61	1.000	0.00	0.00	0.99260	2.17	19.65	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88065	4.60	44.64	35
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.92806	3.44	35.30	35
ME	0.9996	1.00	357.49	1.000	0.00	0.00	0.96071	4.05	43.91	33
Environmental contour 1100										
PF	0.9999	1.00	498.65	1.000	0.00	0.00	0.98432	1.82	18.37	29
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.87120	4.44	44.68	33
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.91360	3.44	34.57	34
ME	0.9997	1.00	316.22	1.000	0.00	0.00	0.95645	3.45	42.01	32
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

B.4 1-day Forecast Horizon

Table B.13: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 1 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.21	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89524	3.79	39.45	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96705	4.60	46.50	27
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96946	3.70	43.19	26
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99746	3.67	27.00	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89270	3.89	40.14	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96546	4.93	47.91	27
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96758	4.04	44.41	26
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99676	2.80	25.22	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88794	3.71	39.78	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96378	4.34	46.84	26
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96728	3.96	44.38	25
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99581	1.98	22.53	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88560	3.55	39.46	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96147	3.69	45.13	26
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96613	3.70	43.50	26
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.14: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 5,000 af/day with a forecast horizon of 1 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.22	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89461	3.79	39.49	26
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96616	4.46	45.87	27
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96890	3.67	42.86	26
Environmental contour 800										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.98382	2.54	23.73	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88695	4.01	43.44	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95242	3.61	43.59	27
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.95767	3.43	42.32	26
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.95721	2.00	22.54	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.87322	3.87	44.30	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.92674	2.72	36.61	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.93519	2.51	35.36	26
Environmental contour 1100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.93625	2.07	22.98	27
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.85994	3.62	44.11	28
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.90577	2.64	34.07	28
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.91508	2.39	32.47	26
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

Table B.15: Performance measures for the testing period 1990-2010 in the Full PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 1 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99990	1.00	33.22	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89402	3.80	39.52	30
IP	1.0000	0.00	0.00	1.000	2.00	16.09	0.96302	4.59	49.49	30
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96794	3.85	45.70	29
Environmental contour 8000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99746	3.67	27.09	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89032	3.86	40.14	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96451	4.82	47.83	30
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96490	3.65	44.65	29
Environmental contour 8200										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99660	2.58	24.61	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88342	3.53	39.49	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95084	3.05	40.81	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96272	3.27	43.43	28
Environmental contour 8250										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99541	1.83	21.60	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88035	3.43	39.42	29
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.95345	3.07	41.85	29
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96487	3.43	42.62	28
Guide curve										
	0.9989	4.00	5484.99	1.000	0.00	0.00	0.99997	1.00	147.77	48

Table B.16: Performance measures for the testing period 1990-2010 in the Zero PVP scenario with a maximum release of 12,694 af/day with a forecast horizon of 1 days.

	Flood Control			Water Supply			Environmental Flow			Hopland
	re1	res	vul	re1	res	vul	re1	res	vul	high flow
Environmental contour 100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.99987	1.00	27.45	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.89319	3.77	39.35	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.96461	4.43	45.82	30
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.96784	3.65	43.85	29
Environmental contour 800										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.98336	2.56	23.67	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.88537	4.00	43.48	30
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.94972	3.47	42.58	30
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.95658	3.35	42.00	28
Environmental contour 1000										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.95698	2.01	22.60	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.87196	3.84	44.54	32
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.92294	2.72	35.90	31
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.93341	2.48	35.08	28
Environmental contour 1100										
PF	1.0000	0.00	0.00	1.000	0.00	0.00	0.93582	2.06	22.98	30
HC	1.0000	0.00	0.00	1.000	0.00	0.00	0.85915	3.60	44.27	31
IP	1.0000	0.00	0.00	1.000	0.00	0.00	0.90263	2.77	33.92	32
ME	1.0000	0.00	0.00	1.000	0.00	0.00	0.91363	2.42	32.40	28
Guide curve										
	0.9995	4.00	5652.72	0.989	13.44	33.85	0.96067	19.52	132.28	43

Appendix C

Experiment Details

The neural networks for the critic, $Q(s, a)$, and actor, $\mu(s)$, share many similarities, and in general, the structure of the two networks derives from those used in (Lillicrap et al., 2015). Originally, both networks had only 2 hidden layers as in (Lillicrap et al., 2015), but this seemed to under-fit both value functions. Another layer was added to the μ function and 2 layers were added to the Q function merely because it has an additional state dimension. The μ network has 3 hidden layers with 400 nodes each. The Q network has 4 hidden layers, each with 400 nodes. The networks in (Lillicrap et al., 2015) began with a layer of 400 nodes followed by a layer of 300 nodes. Again, because this seemed to underfit, it was desired to have a larger network, and a layer of 400 nodes was near the limit that could be handled by the RAM in a desktop Linux workstation. These additional layers did not affect training time significantly. Similar to (Lillicrap et al., 2015), the actions are not included until the second layer in the Q network.

Both networks are fully-connected, feed-forward networks using leaky rectified linear unit (LReLU) (He, Zhang, Ren, & Sun, 2015) activation functions for all hidden layers. The output layer for each network is a single node with a linear activation function. A Xavier initializer (Glorot & Bengio, 2010) is used for all weights and biases except for the output node bias in the μ network. This bias was initialized to 0.5.

The Q network uses the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 10^{-3}. In the Q network this optimizer performs the gradient descent on the loss function and updates the parameters. The μ network has an Adam optimizer with a learning rate of 10^{-4}, though this is only used for applying the gradients provided by the Q network. Both of these learning rates were selected based on the work of (Lillicrap et al., 2015) and changing these values did not seem to improve the learning process.

A value of $n = 30$ was used for the n-step rewards. The maximum length of an episode was limited to 60 state transitions. The incremental updates used a value of $\tau = 0.01$. A batch size of 32 was used when training the ANNs. A value of $\epsilon = 0.1$ was used for the ϵ-greedy technique when learning given the ENV or COMBO rewards. The FC reward used a value of $\epsilon = 0.5$.

The particle swarm optimization used swarms of 50 particles. The inertia was given a value of $\omega = 0.8$, the cognitive and social parameters (c_1 and c_2 in Equation 3.39) each had a value of 0.9. The decision is a 15 dimensional vector of the release for each time step during the forecast horizon. This was normalized to a range of [0,1] corresponding to the maximum and minimum limits of the release. The maximum velocity that a particle could have was clipped in each dimension to [-0.05,0.05]. The modified-puls routing does not handle rapid changes in flow well as this can cause the output to oscillate or even become negative. To manage this, the decisions are also limited to a minimum of 50% of the previous time-step's value. This addition effectively implements a ramp-down rule that is more restrictive than the existing ramp-down rule set by the guide curve.