

University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Environmental & Water Resources Engineering
Masters Projects

Civil and Environmental Engineering

5-2014

Integrating Emerging River Forecast Center Streamflow Products Into the Salt Lake City Parley's Drinking Water System

Rebecca Guihan

Follow this and additional works at: https://scholarworks.umass.edu/cee_ewre

 Part of the [Environmental Engineering Commons](#)

Guihan, Rebecca, "Integrating Emerging River Forecast Center Streamflow Products Into the Salt Lake City Parley's Drinking Water System" (2014). *Environmental & Water Resources Engineering Masters Projects*. 65.
<https://doi.org/10.7275/0DNE-KT29>

This Article is brought to you for free and open access by the Civil and Environmental Engineering at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Environmental & Water Resources Engineering Masters Projects by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**INTEGRATING EMERGING RIVER FORECAST CENTER STREAMFLOW
PRODUCTS INTO THE SALT LAKE CITY PARLEY'S DRINKING WATER SYSTEM**

A Master's Project Report Presented By:

Rebecca F. Guihan

Submitted to the Department of Civil and Environmental Engineering of the University of
Massachusetts Amherst in partial fulfillment of the degree of

Master of Science in Civil Engineering

April 2014

**INTEGRATING EMERGING RIVER FORECAST CENTER STREAMFLOW
PRODUCTS INTO THE SALT LAKE CITY PARLEY'S DRINKING WATER SYSTEM**

A Master's Project Presented

by

REBECCA F. GUIHAN


Approved as to style and content by:



Dr. Austin Polebitski, Chairperson



Dr. Richard Palmer, Member



Dr. Sanjay Arwade
Civil and Environmental Engineering Department

Abstract

Seasonal forecasts of weather provide valuable information for reservoir operations, particularly during unusual events, such as floods or droughts. A challenge confronting reservoir operators today is whether to incorporate new climate products into their operations to help manage such extremes or to use historic data to guide them. This research evaluates the accuracy and value of the Hydrologic Ensemble Forecast System (HEFS) generated from the Climate Forecast System version 2 (CFSv2) using the operations of the Salt Lake City Parley's System and compares it to the accuracy and value of using an Ensemble Streamflow Predictions (ESP) approach. Streamflow reforecasts are generated and used to evaluate the predictive skill of the HEFS in reservoir management. Using the HEFS may offer more insight when responding to climate driven extremes than the ESP approach because the HEFS incorporates a fully coupled climate model into its forecasts rather than relying on the historic record.

Table of Contents

| | | |
|------------|--|-----------|
| 1.0 | PROBLEM STATEMENT..... | 1 |
| 2.0 | INTRODUCTION TO THE USE OF STREAMFLOW FORECASTS | 3 |
| 2.1 | ESP DATA | 8 |
| 2.2 | CFSv2 DATA AND THE USE OF REFORECASTS | 10 |
| 3.0 | CASE STUDY AND MODEL INPUTS | 14 |
| 3.1 | MODEL INPUTS..... | 16 |
| 4.0 | FORECAST ACCURACY | 18 |
| 4.1 | R-SQUARED CALCULATION..... | 18 |
| 4.2 | RANK PROBABILITY SKILL SCORE..... | 19 |
| 4.3 | PREDICTIVE TEST FOR FORECAST USE | 21 |
| 4.4 | GENERAL TRENDS IN VARIABILITY | 28 |
| 5.0 | DESCRIPTION OF MODELS | 31 |
| 5.1 | BASELINE SCENARIO..... | 31 |
| 5.2 | PERFECT SCENARIO..... | 33 |
| 5.3 | INTEGRATED SIMULATION OPTIMIZATION MODEL | 35 |
| 6.0 | RESULTS..... | 39 |
| 6.1 | METRICS..... | 39 |
| 6.1.1 | <i>Water Reliability</i> | 39 |
| 6.1.2 | <i>Bypass Release Metrics</i> | 40 |
| 6.1.3 | <i>Storage Metrics</i> | 40 |
| 6.2 | COMPARING OBSERVED METRICS TO THE BASELINE SIMULATION OUTPUT | 41 |
| 6.2.1 | <i>Water Reliability</i> | 41 |

| | | |
|-------|---|----|
| 6.2.2 | <i>Bypass Release Metrics</i> | 43 |
| 6.2.3 | <i>Storage Metrics</i> | 44 |
| 6.2.4 | <i>Conclusion from Baseline Simulation and Observed Record Comparison</i> | 47 |
| 6.3 | FORECAST PERFORMANCE..... | 47 |
| 6.3.1 | <i>Water Reliability</i> | 48 |
| 6.3.2 | <i>Bypass Release Metrics</i> | 50 |
| 6.3.3 | <i>Storage Metrics</i> | 51 |
| 7.0 | CONCLUSIONS AND FUTURE WORK | 55 |

List of Figures

| | |
|---|----|
| Figure 1: Schematic of SLC Parley's System | 15 |
| Figure 2: Part of CBRFC Combined Inflow Calibration | 17 |
| Figure 3: Annual April-July Inflow Volume – Little Dell SWE (1985-2010) | 24 |
| Figure 4: Annual April-July Inflow Volume – Little Dell ESP (1985-2010)..... | 26 |
| Figure 5: Regression of ESP and HEFS Forecast Means | 27 |
| Figure 6: Summed April-July ESP and HEFS Inflow (Little Dell - January Forecast)..... | 28 |
| Figure 7: Summed April-July ESP and HEFS Inflow (Little Dell - April Forecast)..... | 29 |
| Figure 8: Summed April-July ESP and HEFS Inflow (Little Dell - June Forecast)..... | 30 |
| Figure 9: Salt Lake City Combined Historic Storages by Week | 32 |
| Figure 10: Schematic of the Integrated Simulation Optimization Process | 35 |
| Figure 11: Optimization of January 1983 ESP Forecast..... | 36 |
| Figure 12: DSS ESP Operational Output for 1983 | 38 |
| Figure 13: Water Supply Reliability - Baseline v. Observed..... | 42 |
| Figure 14: Volume of Water Delivered - Baseline v. Observed..... | 43 |
| Figure 15: Bypass Releases - Baseline v. Observed | 44 |
| Figure 16: Days Combined Storage Exceeds 20,000 af - Baseline v. Observed | 45 |
| Figure 17: Total Summed Volume of Storage over 20,000 af - Baseline v. Observed | 46 |
| Figure 18: Maximum Storage Volumes - Baseline v. Observed | 46 |
| Figure 19: Aug. 1 Combined Storage Volume (af) - Baseline v. Observed | 47 |
| Figure 20: Water Supply Reliability - Forecast Performance..... | 49 |
| Figure 21: Volume of Water Delivered - Forecast Performance | 49 |
| Figure 22: Bypass Releases - Forecast Performance | 50 |

Figure 23: Days Combined Storage Exceeds 20,000 af - Model Performance 53

Figure 24: Total Summed Volume of Storage over 20,000 af - Model Performance 53

Figure 25: Maximum Storage Volumes - Model Performance 54

Figure 26: Aug. 1 Combined Storage Volume - Model Performance 54

List of Tables

| | |
|--|----|
| Table 1: R-squared Values of April – July Inflow Volumes for Raw Forecast Data | 18 |
| Table 2: Rank Probability Skill Score for Raw Forecast Data | 21 |

1.0 Problem Statement

Despite their evolving skill, state-of-the-art climate forecasts are not being used to inform reservoir operations and decision making. This research quantifies the benefits of using climate forecasts to create a framework to link global-scale climate research and forecast production to operational watershed-scale streamflow prediction to improve water management and decision making.

This thesis compares and contrasts the new HEFS forecasts made from the CFSv2 model to the older ESP products which rely on climatology, or the historic record, using a case study of Salt Lake City Parley's Drinking Water System. By viewing seasonal forecast data through the lens of water management decision support tools, this study increases the understanding of how climate systems impact water resources management. Reforecast data, or historic forecasts, have been created so the performance of new types of forecasts can be evaluated over a larger sample of data. Analyzing reforecast data in the context of decision making and consequent changes to system performance provides a quantitative assessment of the utility of seasonal forecasting that is meaningful to water resources managers.

Successfully incorporating climate products in the broader water community will create more informed users and improved decision making concerning the stewardship of both infrastructure and natural resources. This will also create a more climate literate water community because managers can use the best available climate science to inform their operations and policy decisions. The framework for generating climate-informed streamflow products by the River Forecast Centers is intended to be easily implemented, setting a pathway for producing national reliable and timely climate-informed streamflow forecasts. The

framework for determining benefits of seasonal forecasts and conditioning operations on forecast skill is flexible and readily transferable to water managers across the nation.

2.0 Introduction to the Use of Streamflow Forecasts

This chapter introduces how streamflow forecasts have been used in reservoir decision making. Although several studies have proven forecasts to be useful, there are also many challenges associated with incorporating forecasts as a tool in real decision making settings. Two specific streamflow forecast products will be explained in more depth because they are the products that will be examined in this study: the Ensemble Streamflow System (ESP) and the Hydrologic Ensemble Forecast System (HEFS).

Forecasts can be an important tool in the management of water resource systems. When water resource systems are managed properly, there is a thoughtful trade-off between the services provided by infrastructure in the built environment and natural ecosystem services. For example, short-term forecasts can be used to minimize flood risk or to maximize the revenues from a hydropower facility. Medium to long range forecasts can be used to reduce the risks of extreme events like flood or drought. Often a reservoir is operated for multiple purposes. Managers can use forecasts to find a balance between competing objectives. To provide adequate water supply, even in times of drought, managers try to maintain high reservoir levels. Conversely, they must keep reservoir levels low to reduce risk of flooding in an extreme precipitation event.

Ensemble streamflow predictions and energy price forecasts are examined by Alemu et al (2010) to help improve dam operations in a complicated and interconnected system. The Jackson Hydropower Project in Washington State is used as the case study to test the decision support system, which first simulates the general operating rules and then optimizes the multipurpose reservoir system using state variables and medium range forecasts as input. The goal of the

decision support system is to maximize revenue from hydropower production, while meeting the drinking water demand for nearby City of Everett and federal streamflow regulations.

Alemu et al. (2010) found that substantial gains were possible using forecast information to aid the water management process. Forecasts of two types were explored: streamflow forecasts and energy price forecasts. However, scenarios with perfect energy price forecasts yielded better results than those with perfect streamflow input and forecasted energy prices. For this case, accurate energy prices improved the timing of the releases which was more important than having perfect information of the inflows to the system. This research uses reforecast data to quantify potential operational improvement forecasts could provide for their system. The range of improvement may be limited with reforecast data because the system depends on a variety of factors in its operation.

Streamflow forecasts can improve the decisions made by water suppliers and watershed stakeholders. Like Alemu, Gong (2010) presents a simple framework to incorporate streamflow forecasts into an existing management system by predicting the future seasonal inflow forecast, future daily inflow forecast, and future reservoir level forecast. The reservoir release policy is then modified and the performance of the new system can be evaluated. This approach was tested on New York City's supply to offer a more flexible release policy for environmental flows and flood control, which could provide additional ecological performance without jeopardizing reliability of the NYC supply. Moradkani (2010) also showed forecasts generated with independent component analysis could improve skill by simply using basic climatology in the western United States.

Lee et al. (2009) improved flood control rule curves for dam operation in the Columbia River Basin by considering the effects of global warming. An optimization model created a new

set of operation rule curves for global warming scenarios. The study found that flood protection for all of the control points improved under climate change scenarios using these new flood control methods when compared to the current flood control curves. The reliability of reservoir refill did not change when the new rule curves were implemented. Under changing climate conditions, the Columbia River Basin could be successfully operated to provide effective flood control if optimized flood rule curves replace the current flood control rule curves.

After exploring the improvements possible by modifying flood control curves for a climate change scenario, Lee et al. (2011) improved their model by organizing El Niño-Southern Oscillation (ENSO) years into three retrospective classifications: warm, neutral, and cool. Penalty functions were calibrated specifically for each ENSO state which accounted for the varying temperature and precipitation anomalies ENSO brings. Many papers have shown that ENSO greatly affects precipitation, snow accumulation, and stream flow when analyzed at individual sites and across a larger spatial area (Cayan, 1999). Anomalies caused by ENSO are expected to be important predictive factors and a specific means to evaluating CFS skill through reforecast data.

Although forecasts are helpful when accurate (Alemu, 2010), unskilled forecasts can result in negative impacts. In 1977, water resources managers predicted a dry spring and summer, because of low snow pack levels stored from the winter (Glantz, 1982). The drought proved not nearly as bad as predicted. In order to conserve water, the Yakima Project Superintendent severely restricted water use based on early forecasts. It was later announced that the original water supply predictions grossly underestimated available water due to a modeling error and for the growing seasons farmers received about 70% of their normal water allocations. Many farmers had made accommodations to deal with an extreme water shortage, which proved

unnecessary. Glantz's paper demonstrates the importance of accurate forecasting and warns of the economic and social consequences if significant mistakes are made.

Given the potential impacts of inaccurate forecasts, there has been some reluctance in including forecasts into operational planning. When water resource managers are asked why they do not regularly use climate forecasts to help them improve hydrologic management, the answer is often because they do not find the forecasts reliable (Rayner, 2005). The forecast models must consider complicated and unique natural spatial terrain and the role of infrastructure. The complicated nature of the models that generate forecasts could cause some of the manager's lack of confidence and deter usage. Another reason forecasts are not being used are because they can be poorly presented and difficult to use. When applying reforecast data to an operations system, acknowledging these challenges will help streamline the process to be more user-friendly.

However, O'Connor (2005) found that the quality of the forecast did not impact the likelihood of the forecast being used. He determined that convincing managers of the utility of weather and climate forecasts would not increase usage. This directly conflicts with Rayner's (2005) argument that managers would be more likely to use forecasts if they understood how it would be more reliable. The water managers were most likely to use weather and climate forecasts when they were feeling at risk; in fact, the forecasts usage about doubled when risk perception increased. Managers may feel more inclined to use forecasts as the demand for water and therefore risk increases, because the margins for making operational management decisions will decrease (Day, 1985).

Water systems are often highly fragmented because they are designed to serve many purposes. There are often multiple stakeholders invested in conflicting objectives as to how a system should be operated. The goal of water suppliers is remain unnoticed because mistakes,

such as poor water quality or having to place restrictions on water use, attract the negative attention of politicians and customers. System cost is not valued as highly as customer service and reliability. “Cost is third in the hierarchy; to be minimized within the constraints of reliability of quality of supply”, so improving an already functioning system to save money is seen as a risk (Rayner, 2005). Gong (2010) also claims the conservative nature of most water managers inhibits fundamental changes in management practices.

Streamflow forecast products were proven to generate benefit in a system that must be optimized to serve multiple purposes (Alemu 2010) and improve flood control operational regimes along the Columbia Basin (Lee 2009). Lee et al. (2009) and Cayan (1999) have shown that climate signals such as the ENSO cycle can have an influence on streamflow and thereby are useful to the forecasting process. Raynor (2005) found that forecasts are not being used as widely as possible because there are political challenges that discourage diverting from the status quo of operation practices in addition to the challenge in understanding how to use the forecasts. This thesis will demonstrate how two forecast products can be implemented into a small reservoir system, in the hopes of that it will serve as an example for similar systems and encourage the use of forecast products in general.

There are two types of forecast products used in this study: the Ensemble Streamflow System (ESP) and the Hydrologic Ensemble Forecast System (HEFS). The HEFS forecasts include the Climate Forecast System version 2 (CFSv2) model, which is state of the art technology. All forecasts were produced by the Colorado Basin River Forecast Center (CBRFC) for two inflow points of interest. The forecasts were generated on a monthly time-step from January to July (ie. issued on January 1, February 1, March 1, etc.). This generated 6 forecasts

per year. The ESP reforecasts were generated from 1981-2010, and the HEFS reforecasts, which contain the CFSv2 model, were generated from 1985-2010.

2.1 ESP Data

The Ensemble Streamflow Prediction (ESP) is a forecast product produced by the National Weather Service River Forecast Centers (NWS RFC) that uses a historic time series of temperature and precipitation to support water managers in understanding the range of streamflow that they could likely encounter in the next year. This product has been available in some form for decades, so there has been extensive research on its skill and usefulness. Rather than using the recorded historic streamflow data, the ESP uses historic recorded temperature and precipitation data to force a hydrology model which is initialized to the current starting state. The starting state should ensure the hydrology model accurately reflects the system conditions at the start of the forecast.

The ESP uses historical meteorological data and creates an ensemble of futures by assuming that each year of historic data is a possible representation of the future. One unique streamflow trace is simulated for each historical year using the current watershed conditions as the initial conditions for each simulation. Several initial conditions must be defined before each possible future, or trace, is simulated. Parametric information needed for initial conditions include up-to-date information including snow pack, soil moisture, and current streamflow along with other parameters such as melt factors, recession constants, and routing coefficients (Day, 1985).

There are a number of challenges in defining the initial state variables, and their estimates are important because they can bias the ESP results. The spatially aggregated precipitation and temperature data used to force the models are estimates. The physical process models performing

simulations are arithmetic approximations of natural systems, and hydrology model calibrations often present challenges (Day, 1985).

Wood and Schaake (2008) describe the value of ESP as accounting for uncertainty in future climate by assuming that historic climate variability is a good estimate of current climate uncertainty. However they note that uncertainty of the initial hydrologic state is a significant component of the overall forecast uncertainty. A deterministic estimate of the forecast ensemble's initial hydrologic state leads to an overconfident forecast because it implies a certainty that does not actually exist. The variability of the ensemble tends to be narrower than what the total forecast uncertainties warrant. Wood and Schaake (2008) investigated four post-process calibration methods that generally improved the probabilistic forecast skill, but at the cost of forecast sharpness, a forecasts ability to predict extreme events. Another limitation of the calibration methods noted is that they discard the raw ensemble estimates of uncertainty. One major strength was that the approach ensured that the model ensemble forecasts could not perform worse than the climatological forecast.

Wood and Lettenmaier (2008) tested how forecast accuracy depended on initial conditions, specifically examining initial soil moisture conditions. By comparing traditional ESP to what they called a "reverse-ESP" they tested whether forecast depended more on the starting soil moisture conditions or the meteorological inputs given to the hydrologic system. Traditional ESP uses a hydrologic model with initial conditions that are assumed to be perfect that is forced by a forecast ensemble that is sampled from observed historic meteorological sequences. The "reverse-ESP" combines an ensemble of sampled potential initial conditions forced with a perfect meteorological sequence. For some cases tested during specific seasons, they found initial condition knowledge can be more important than climate information. In studying the use

of ESP along with climate based forecasts, Yuan (2013) also found that for a drought plan in two basins being examined, the soil moisture initial condition may have more impact on the predictive skill for a forecast initiated during the dry season.

The mean or median of the ensemble forecast is commonly used as the only prediction, and when reported in this way, one would assume the mean or median is the best prediction (Najafi et al 2012). This assumes that the weather sequence based on each historical year of the record has an equal change of occurring in the current forecast year (Gobena and Gan 2010). Many studies have investigated parametric and non-parametric weighting methods to best use ensemble forecasts: Equal Weights, Index Different Weights, K-Nearest Neighbors, Fuzzy C-means clustering, etc. (Gobena and Gan 2010, Najafi et al 2012, Werner et al 2004). All of these methods have been analyzed on the ESP dataset, but could be applied to other ensemble based forecasts such as the CFSv2. These weighting methods have been shown to improve the utility of the forecasts, but the best method may vary from system to system. A water manager would have to perform a study to see what correction and weighting scheme gave his system the most improvement, whereas using the mean or median value is a far more simplistic approach. Werner et al. (2013) demonstrated that managers tend to rely on the median of a probabilistic forecast if forecasts are used at all, and that more work is needed to understand how to most effectively leverage forecasts to improve water the management of built water systems.

2.2 CFSv2 Data and the Use of Reforecasts

The Climate Forecast System version 2 (CFSv2) is the newest climate forecast product made by the NWS RFCs. This framework uses a fully-coupled atmospheric and oceanic climate model designed to provide high quality forecasts with longer-lead times. Although the CFSv2 has only been available operationally for a short time, there is a growing body of literature that

discusses the skill and quality of previous iterations of seasonal forecasts in comparison to the CFSv2 (Yuan et al 2013 and Najafi et al 2012).

Barston (1994) analyzed the performance of five ENSO prediction systems that were either operational or potentially operational. At the time, ocean and atmosphere models were not fully integrated and only provided a coarse representation the physics of atmospheric and oceanic motion and interaction. Barston pointed out that similarity of skill scores between forecasts did not imply that the forecasts themselves were similar. Also highlighted was that there was no bias-free method of estimating the forecast skill of the hindcasts and that there is no substitute for real-time forecasting, which will be discussed later. Barston found that the models tended to predict the strongest ENSO episodes best, and did not perform as well on weaker ENSO fluctuations and neutral periods.

Hamill (2004) applied retrospective forecast, or reforecast, data to improve medium range forecasts. This research attempted to improve six to ten day and week 2 probabilistic forecasts by using retrospective reforecast data. A model output statistics (MOS) approach improved raw numerical forecasts by implicitly removing model bias. One major challenge of this analysis was that the MOS approach requires large sample of forecasts and verification data has to be available for the regression analysis.

Hamill (2004) found that forecasts generated using reforecast calibration were significantly more skillful for both temperature and precipitation than ones produced only with raw ensembles or bias corrected ensembles. In particular, medium to extended range probabilistic forecasts can be improved by using MOS techniques, because these algorithms can filter the predictable from the unpredictable. This background information from earlier research frames the type of improvements possible when using reforecast data.

Hagedorn (2008) explains the calibration of two-meter temperature forecasts using reforecast data. Non-homogeneous Gaussian regression was selected as the method to validate and calibrate the model. Calibration was particularly successful in increasing the skill of the model, as demonstrated in other studies (Hamill, 2004). For each forecast lead time, a linear regression analysis was performed for each model. The result was a technique to predict the lowest root-mean-square error for each model's raw forecast. After calibrating and running the model, it was determined that using reforecast data produced more skillful calibrated forecasts than other datasets at longer leads. A forecast calibrated with multiple models was more skillful than an individually calibrated forecast. Hagedorn (2008) thoroughly demonstrates many different applications for reforecast data.

Hamill (2008) expanded on Hagedorn's (2008) research by investigating if precipitation reforecast data are useful. This study employs validation and calibration models similar to those in Hagedorn's study on reforecast temperature data. Hamill (2008) noted that the precipitation forecast calibration discussed in this article is very different in character from temperature forecast calibration. They found that both the ECMWF and GFS datasets improved in reliability and skill after being calibrated with weekly reforecast data. Both of these papers indicate that large improvements in forecast skill and reliability are possible when using reforecast data. Reforecasts have been found to be useful for improving skill, but there is a need to explore other approaches for optimal reforecast configuration.

In 2006, Hamill studied ten different analog techniques using reforecast data, which provided an underlying theoretical basis for using reforecast data and explaining practical approximations that need to be made to apply it effectively. They compared each analog method to explore if any one technique could be deemed more useful.

Yuan et. al. (2013) compared the skill of the ESP, CFSv1, and CFSv2 datasets against climatology for the conterminous United States. Six month seasonal hindcasts of monthly precipitation were used to force a VIC model producing 6 month, 20 member ensemble streamflow for the first of the month from 1982-2008. A Bayesian post-processing method was used to correct errors and biases for the forecasts. They found that all three forecast products generated more skillful forecasts out to two months when compared to climatological forecasts. Although these results are encouraging for general skill of the forecast products, this thesis will explore the skill when examined on much finer spatial and temporal resolution.

3.0 Case Study and Model Inputs

This research applies the NWS's ESP and HEFS forecast products to the Salt Lake City Parley's Drinking Water System located in the Colorado River Basin about 10 miles east of Salt Lake City. Parley's Watershed is one of four drainages that are included in Salt Lake City's "Protected Watershed" Canyons. "Protected Watersheds" are defined as the source water drainages along the Wasatch Front mountain range.

The Parley's Watershed contains two reservoirs, the Mountain Dell facility and the Little Dell facility (Figure 1). The Parley's system has two primary management objectives: provide Salt Lake City with reliable, safe, and cost effective public drinking water and provide flood control protection downstream of the system. Little Dell Reservoir has a maximum storage of 20,500 acre-ft, with 17,500 acre-ft of active storage. Immediately downstream of the Little Dell Reservoir is Mountain Dell Reservoir. Mountain Dell Reservoir has a maximum storage of 3,200 acre-ft, 2,200 acre-ft of which is active storage. The reservoirs are situated in Parley's Canyon, with the surrounding mountains reaching maximum elevations of 7,120 feet. Snowfall and its subsequent melting significantly influences the timing and volume of runoff in any given year.

Forecasts are most valuable for the Parley's System during the March through June period, when snowmelt and rainstorms can produce significant amounts of runoff over short periods of time. During this critical period, releases are often made for flood protection, in addition to the usual deliveries made for drinking water supply. Forecasts can provide value when trying to balance these two competing objectives on a monthly operating period as the active storages are small and sensitive to monthly operations.

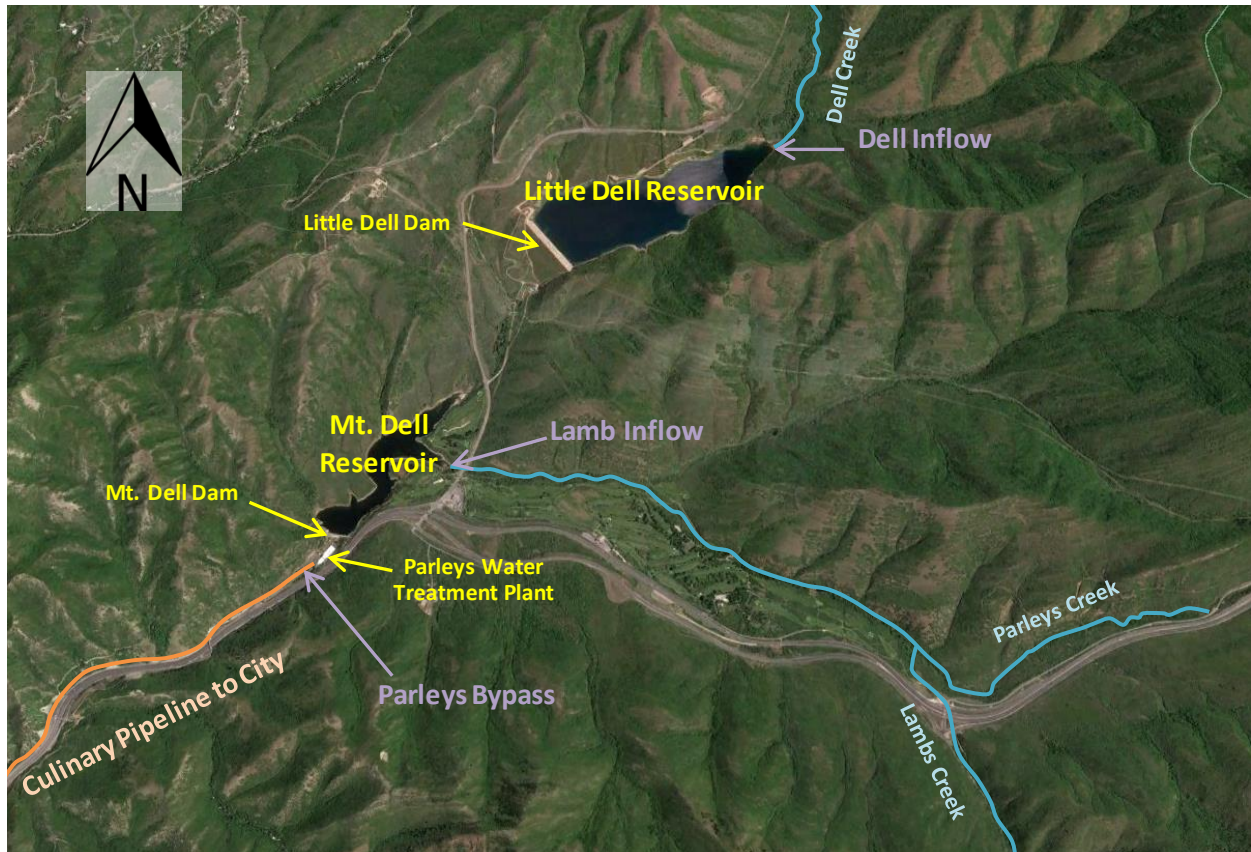


Figure 1: Schematic of SLC Parley's System

The decision support system developed for this study includes a simulation model that incorporates system constraints and operating policies and an optimization model which uses perfect foresight to determine best-case operating policies within given constraints. To determine the value of the reforecast products, performance metrics meaningful to managers are identified and quantified. Without such metrics and awareness of seasonal operational nuances, it is difficult to identify forecast improvements in meaningful ways.

Balancing the two primary management objectives (system reliability and flood control protection) requires maintaining reservoir pool levels as high as possible while providing sufficient storage for catching floods. The analysis of these metrics focuses on the critical time periods where forecasts can be used to effectively manage floods events. These system performance metrics are compared for the reforecast, climatology, and observed scenarios to

evaluate the potential benefits of using CFSv2 seasonal forecasts in systems decision making. Achieving these objectives is not only significant when analyzing results, but is a crucial part of the collaboration between researchers and water managers when building decision support systems.

3.1 Model Inputs

For this research, it is assumed that the calibrated output of the CBRFC hydrology model is the observed record. For the Parley's System, there are two inflow points of interest: Lamb's Creek and Dell Inflow. The recorded gage inflows for the inflow points are from 1994 to 2011. For decision making, it is advantageous to have a longer period of record to test over. By using the CBRFC flows, the full record of ESP and HEFS forecasts can be tested and directly compared to a baseline scenario. Figure 2 provides an example of the combined gage inflows compared with the CBRFC calibrated flows from 2008 to 2010. This segment is representative of the quality of the calibration for the overlapping records.

The ESP and HEFS forecasts were produced by the CBRFC for two inflow points of interest. The forecasts were generated on a monthly time-step from January to June (ie. issued on January 1, February 1, March 1, etc.). This generated 6 forecasts per year. Each forecast contains a 30 trace ensemble, or 30 potential futures. The ESP forecasts were generated from 1981-2010 and the HEFS forecasts, which contain the CFSv2 model, were generated from 1985-2010.

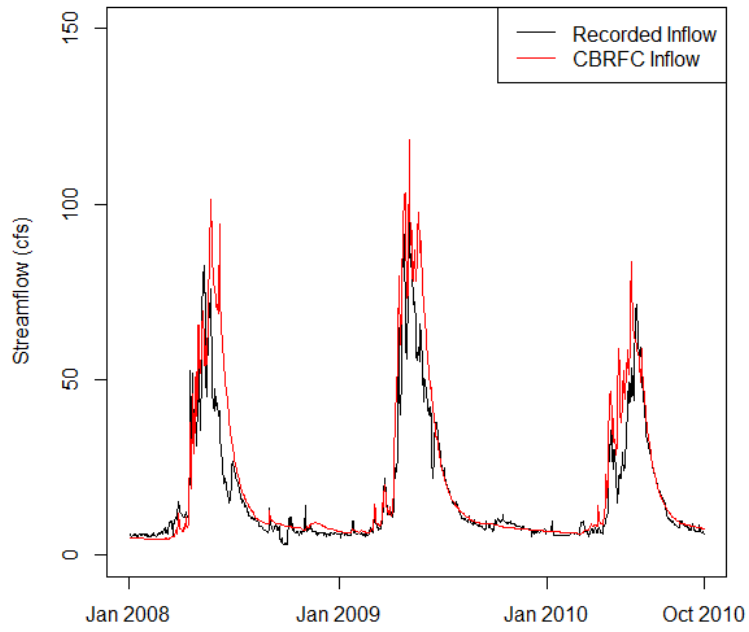


Figure 2: Part of CBRFC Combined Inflow Calibration

4.0 Forecast Accuracy

The ESP and HEFS streamflow forecasts were analyzed for their overall skill prior to incorporating them in the decision support system. This type of analysis provides water managers a better understanding of the forecast data they are using and the level of accuracy they can reasonably expect.

April through July is a critical time for reservoir management because it is when the majority of runoff enters the system, so there is heightened flood risk. This also coincides with making some of the largest water supply deliveries. Accurate forecasts for this period provide the opportunity to manage water for its most important purposes with the maximum information possible. The accuracy of the forecasts is calculated by analyzing how accurately the total volume from April to July is predicted.

4.1 R-squared Calculation

For each month, the R-squared values of the linear regression were calculated between the mean of the ensemble April through July inflow volumes against the observed volume (see Appendix A). For forecasts starting in May, the observed May through July volumes were used, and for June forecasts, observed June through July volumes were used.

Table 1 contains the R-squared values of the ESP and HEFS products for the Lamb and Dell inflows. The R-squared values improve as the forecast period of interest is approached (i.e. the April forecast predicts the April through July inflow volume better than the January forecast). This indicates that using the mean value of the HEFS ensemble may be slightly more informative than the mean of the ESP ensemble.

Table 1: R-squared Values of April – July Inflow Volumes for Raw Forecast Data

| | Dell ESP | Dell HEFS | Lamb ESP | Lamb HEFS |
|-----|-----------------|------------------|-----------------|------------------|
| Jan | .426 | .447 | .430 | .461 |
| Feb | .748 | .673 | .752 | .712 |
| Mar | .683 | .672 | .722 | .708 |

| | | | | |
|------|------|------|------|------|
| Apr | .810 | .836 | .832 | .858 |
| May* | .862 | .876 | .907 | .928 |
| Jun* | .942 | .947 | .971 | .965 |

**Note: May and June only include May–July volumes and June –July volumes respectively*

A drawback of this analysis is that the R-squared value is measuring the predictive power of the mean of the ensemble of the forecasts. This approach does not consider all of the useful information within the forecast (spread and bias among others). To provide a more complete understanding of forecast quality, other metrics are investigated.

4.2 Rank Probability Skill Score

One method to quantify the quality of forecasts is to determine if they provide more information than climatology. The Rank Probability Skill Score (RPSS) is a well-documented metric which represents the percent improvement of a probabilistic forecast with respect to a reference forecast (Wilks 1995, Werner et. al. 2004 and Gobena and Gan 2010). Using climatology as a reference forecast can provide us with an understanding if the forecasts are accurate and therefore informative to managers. Unlike the R-squared value which only uses the mean of the ensemble, the RPSS evaluates the entire distribution of an ensemble forecast.

In order to calculate the RPSS, the Rank Probability Score (RPS) must be calculated first. It is a measure of how well a forecast predicted the category an observation fell into. The RPS measures the sum of squared differences between the cumulative probability of a forecast (Y_m) and the cumulative probability of an observation (O_m) occurring in any number of forecast categories (J). In this example J , the number of forecast categories is equal to 3. The cumulative distribution function of the observed record is divided into terciles, and the corresponding streamflow limits create the bounds for the forecast categories representing low inflows, average inflows, and high inflows. In each year, the RPS is calculated by

$$RPS_{year} = \sum_{m=1}^J [Y_m - O_m]^2$$

RPS values can range from 0 to 1, with 0 being a perfect forecast.

For each year, the observation will fall into one of three categories. For example, when calculating the RPS for an average inflow year compared to climatology, the observed cumulative probability values would be $O_1 = 0$, $O_2 = 1$, and $O_3 = 1$ and the forecast cumulative probability values would be $Y_1 = 1/3$, $Y_2 = 2/3$, and $Y_3 = 1$ (because the forecast categories were created such that the probability of being in any category is equally likely according to a climatology forecast). When making the RPS calculation for the same year using an ensemble forecast where 10% of traces fell into the low flow category, 30% fell into the average flow category, and 60% of traces fell into the low flow category, the observed cumulative probability values would still be $O_1 = 0$, $O_2 = 1$, and $O_3 = 1$ and the forecast cumulative probability values would be $Y_1 = 0.1$, $Y_2 = 0.4$, and $Y_3 = 1$.

The RPS is calculated for each year using both a climatology forecast and either the ESP or HEFS forecasts compared with observations. The average RPS for both the climatology forecast and the forecast being compared is used to calculate the RPSS. RPSS is defined as

$$RPSS = 1 - \frac{\overline{RPS}_{forecast}}{\overline{RPS}_{climatology}}$$

Higher RPSS numbers correspond with more informative forecasts in reference to the climatology, with 1 being a perfect forecast. If the RPSS is negative, it would mean that the forecast was less accurate than climatology.

For the ESP and HEFS forecasted inflows, the Rank Probability Skill Score (RPSS) was calculated for the combined April through July inflow volumes using the observed record as a

reference forecast (Table 2). Three bins were chosen to represent terciles of the probable historic distribution of inflows. The years contributing to the RPSS were limited to where the ESP and HEFS data overlaps, 1985 – 2010.

From the results in Table 2, using a forecast is always more informative than only using climatology. There is useful and accurate information in both the ESP and HEFS forecasts when predicting April through July inflows, even in longer lead times such as January. The RPSS for ESP and HEFS are very similar, there is no clear trend indicating that one forecast product is more accurate than another. This suggests that the forecasts quality is driven largely by the starting state of the hydrology model, rather than a strong climate signal. A starting-state driven system also explains why lead time plays such a large role in forecast accuracy.

Table 2: Rank Probability Skill Score for Raw Forecast Data

| | ESP (26 years) | HEFS (26 years) |
|-----|----------------|-----------------|
| Jan | 0.269 | 0.279 |
| Feb | 0.497 | 0.464 |
| Mar | 0.508 | 0.501 |
| Apr | 0.614 | 0.617 |
| May | 0.629 | 0.665 |
| Jun | 0.906 | 0.906 |

4.3 Predictive Test for Forecast Use

The RPSS suggests that both the ESP and HEFS forecasts are more skillful than climatology in predicting the April through July inflow volume. A simple experiment was conducted to illustrate this point and compare the forecast’s ability to another metric, snow water equivalent (SWE) depth. For this experiment, a 26 year period common across all of the datasets was selected, 1985 to 2010.

With no forecast information available, a water manager may rely on climatology as a surrogate for a forecast. For the Salt Lake City Parley's System main inflow point, Dell Creek, the April through July inflow volume could range from approximately 1 to 10 thousand acre feet (kaf). A simple forecasting method would involve using SWE information. These values are readily available, and for this study, the NRCS website's SNOTEL values for Parley's Mountain were used. The SWE values on the date of each forecast (ie. January 1, April 1, and June 1) were recorded for each of the 26 years of interest. These SWE values were matched with the actual April through July inflow volume for the corresponding year and ranked from lowest to highest. These data were divided into three groups: the lowest SWE values, the middle SWE values, and highest SWE values. The observed April through July inflow volumes were plotted based on ranking of their corresponding SWE values.

Figure 3 indicates that snow water equivalent is a valid forecasting tool and is related to inflows to the system on January 1 and April 1. In both cases, the range of expected inflows decreases in relation to a SWE forecast being low, normal, or high. Between January and April, the accuracy of the snow water equivalent as a forecast increases as the lead time decreases. The April 1 snow water equivalent is a slightly better predictor of potential inflows than the January 1 snow water equivalent, as the range of inflows is smaller and more well defined when compared to climatology. However, the June 1 snow water equivalent does not provide as much useful information because most of the snow has melted by at that time. With many SWE values equal or very close to 0 inches, there is little useful information to be gained about the rest of the runoff season from this type of a SWE model. Relating historic snow information to the current state of a basin provides a water manager with simple forecast tools to plan ahead and potentially decrease vulnerability to anomalous climate conditions.

A similar experiment can be done with the ESP and HEFS products being investigated. Rather than plotting the April through July inflow volumes based on the lower, middle, and upper third of ranked of SWE values, they are plotted based on the lower, middle, and upper third of ranked means of the April through July inflow volumes forecasted by the ESP and HEFS products (Figure 4 and Appendix B respectively). The means of the ESP and HEFS forecasts are similar (Figure 5) resulting in identical rankings of the years. The distributions as represented by the boxplots are nearly identical, suggesting the information gained from using each forecast the same. This is not surprising as the R-squared values of the means were similar.

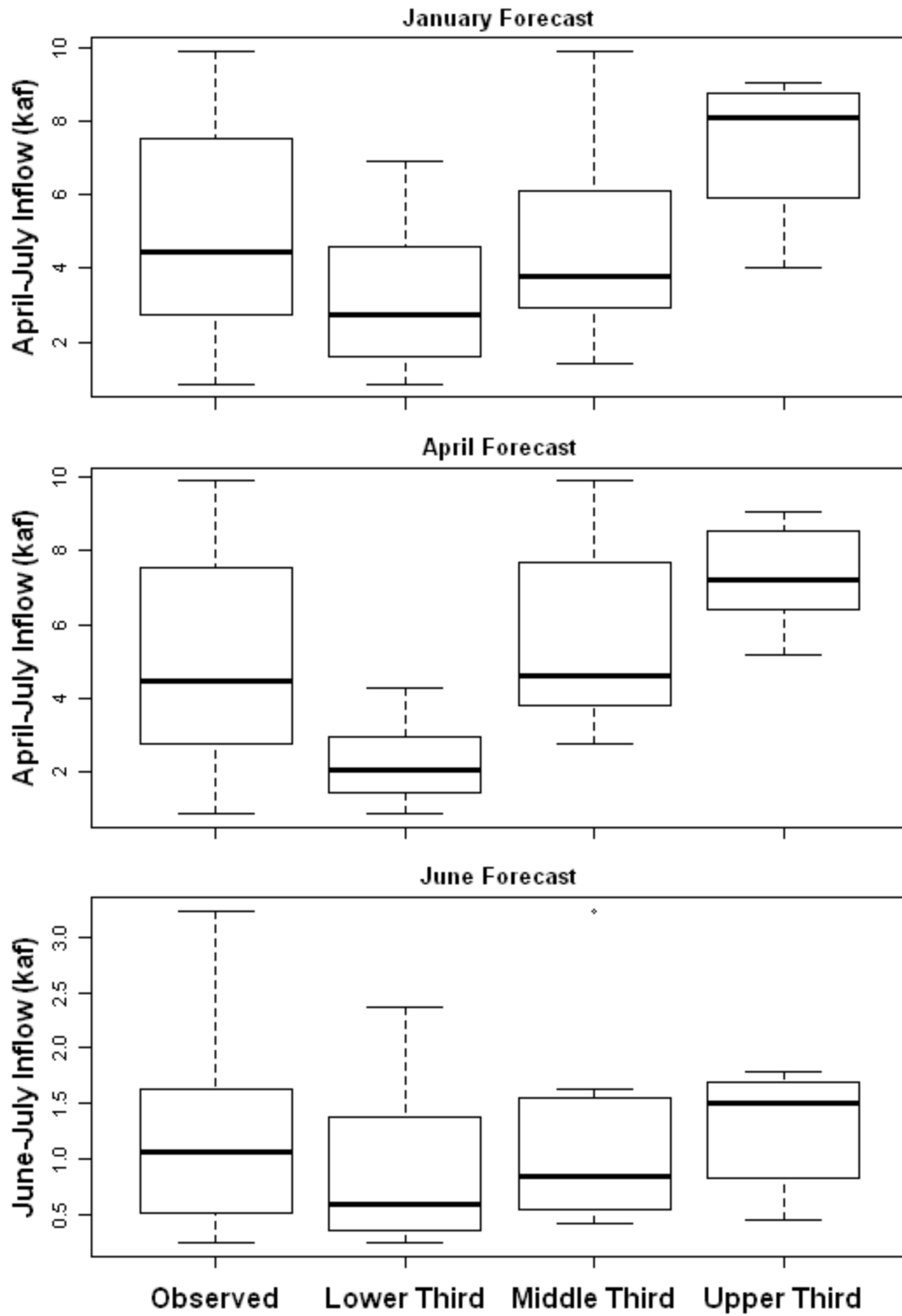


Figure 3: Annual April-July Inflow Volume – Little Dell SWE (1985-2010)

After calculating the R-squared values and RPSS, it is not surprising that both forecasts were informative; they provided a narrower and more accurate range of potential inflows based on the mean of the predicted forecast inflow being low, average, or high. For this inflow point, the ESP and HEFS forecasts provided a better defined, more accurate range of inflows (Figure 4), so they are more useful than using SWE alone. Also, the ESP and HEFS have a great deal of accuracy in May and June when SWE has little skill.

Based on the preliminary results for these two inflow points for Salt Lake City, using some a forecasting method, be it SWE, ESP or HEFS, is more informative than not using any type of forecasting information and relying on climatology. The researchers have worked closely with the CBRFC to clarify and debug issues with early versions of the hindcasts. Running the models in hindcast mode to produce the reforecasts datasets inherently presents computational challenges which do not need to be considered in real time. At this stage in the analysis, it is inconclusive whether the HEFS forecasts that include the CFSv2 model will provide significant benefit over using the ESP forecasts, but it is clear that forecasts tend to become more accurate as lead times decrease and that using the ESP or HEFS forecast will be more beneficial than not using any forecast information, or a simple method such as snow water equivalent.

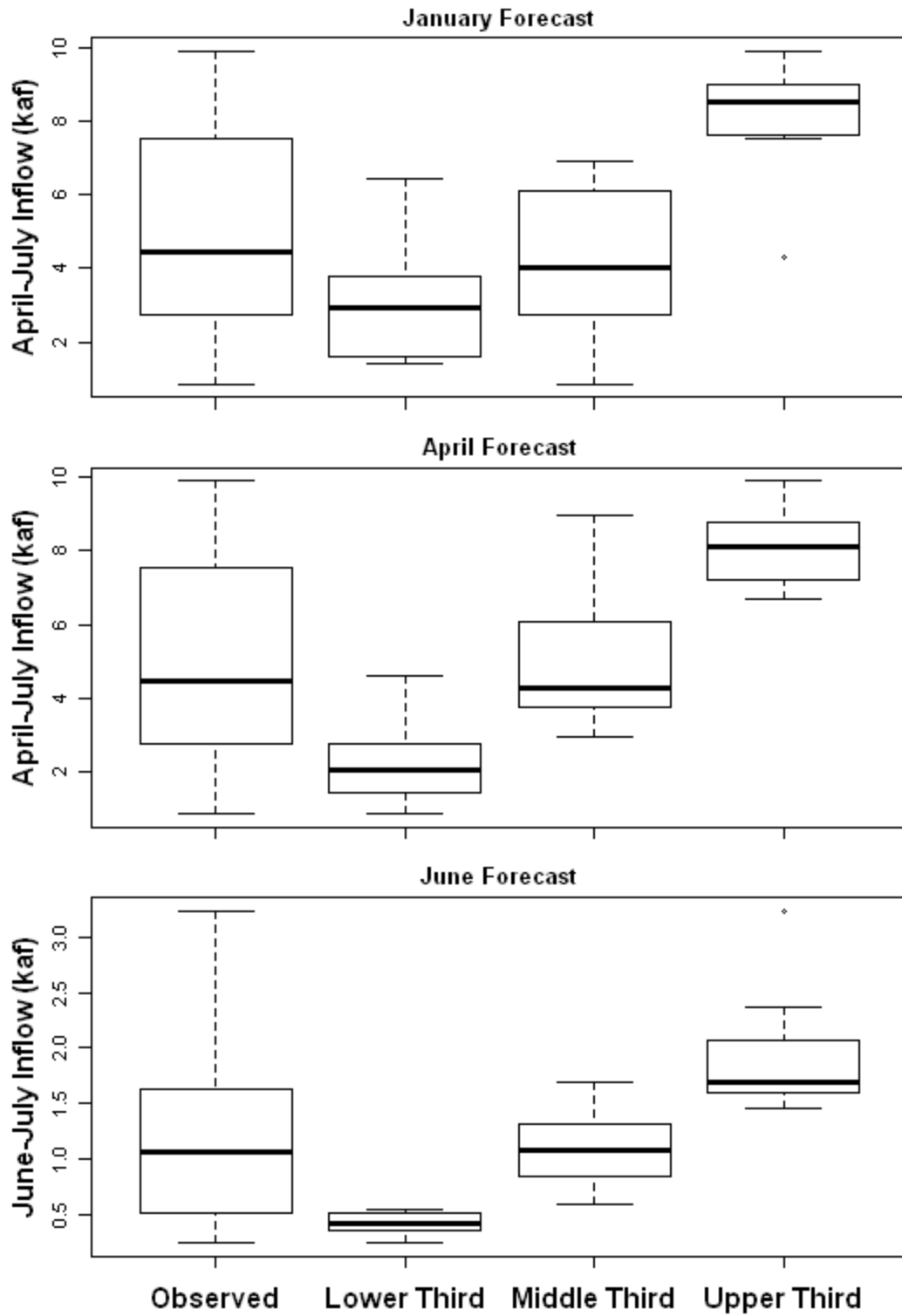


Figure 4: Annual April-July Inflow Volume – Little Dell ESP (1985-2010)

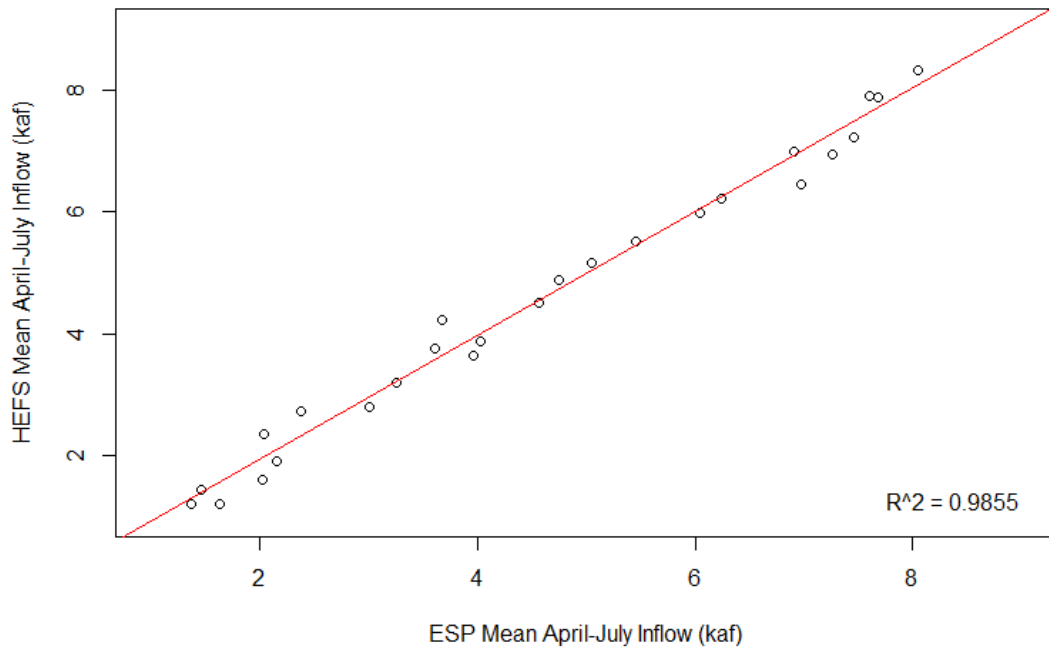


Figure 5: Regression of ESP and HEFS Forecast Means

4.4 General Trends in Variability

Figure 6 shows the boxplot of April through July inflow values of each trace in the January forecast ensembles compared with the observed volume (shown as a red “X”). The ESP and HEFS forecasts of the Little Dell inflow point are both shown in succession for each year. A similar graph of the April forecasts for Little Dell is shown in Figure 7. In all cases, the January forecast ensemble has a larger spread than what is predicted in the April forecasts. For the Dell inflow point shown, the April forecasts have about half of the variability as the January forecasts.

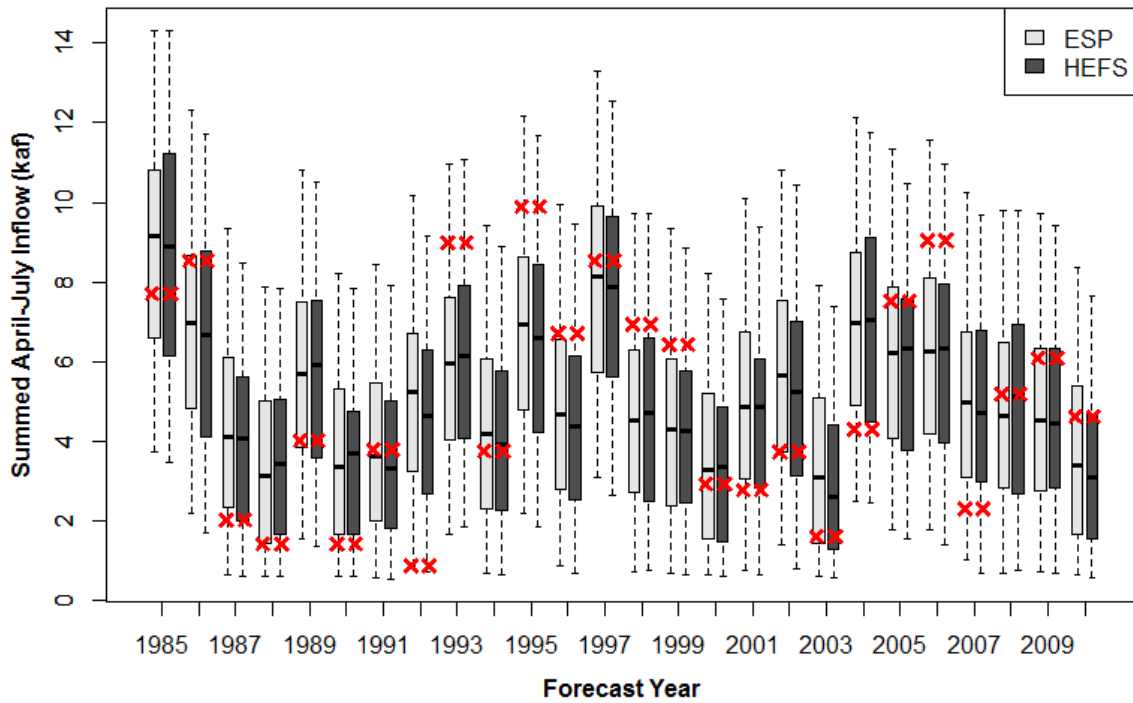


Figure 6: Summed April-July ESP and HEFS Inflow (Little Dell - January Forecast)

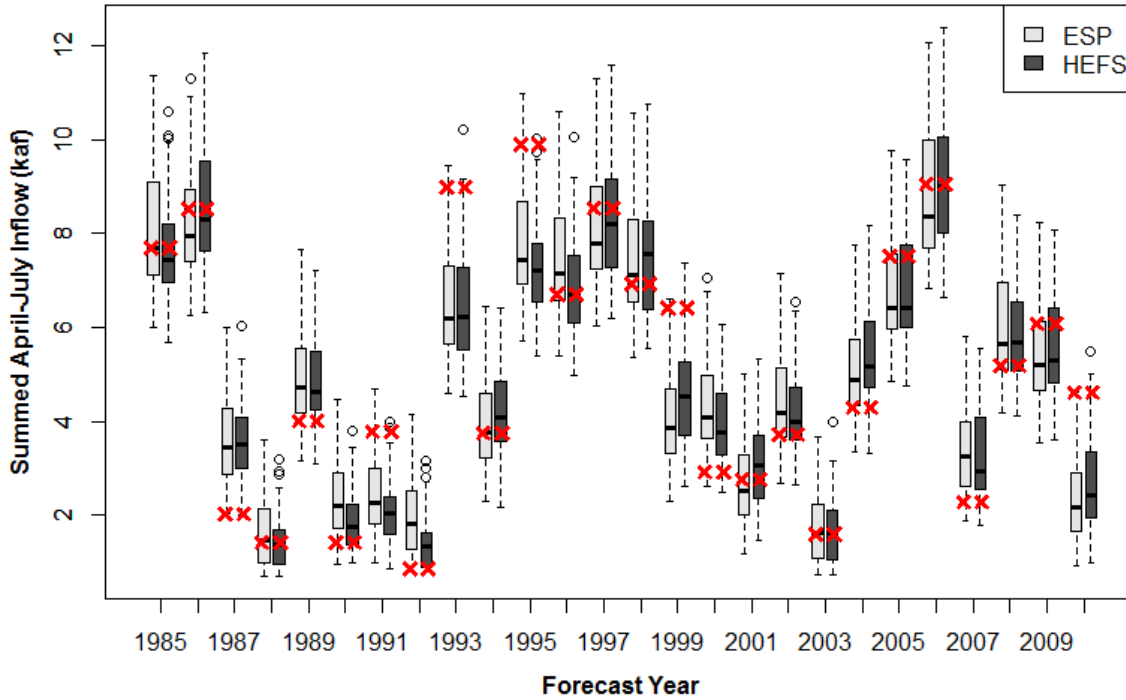


Figure 7: Summed April-July ESP and HEFS Inflow (Little Dell - April Forecast)

Similarly, the forecast issued on July 1 has the most precision, partially because half of the runoff season is over, so only the June and July inflow volumes are being predicted (Figure 8). It is critical that these late-season forecasts are accurate, because drinking water deliveries are still being made, and they set the stage for understanding where the system will be going into winter.

The less variable a forecast is, the more easily it can be incorporated into a decision framework. The relationship between these plots and the RPSS shown in Table 2 suggests that the forecast becomes more accurate and informative as the variability of the ensemble decreases. Conversely, there is still some accurate information in the forecasts with longer lead times although those forecasts are increasingly variable.

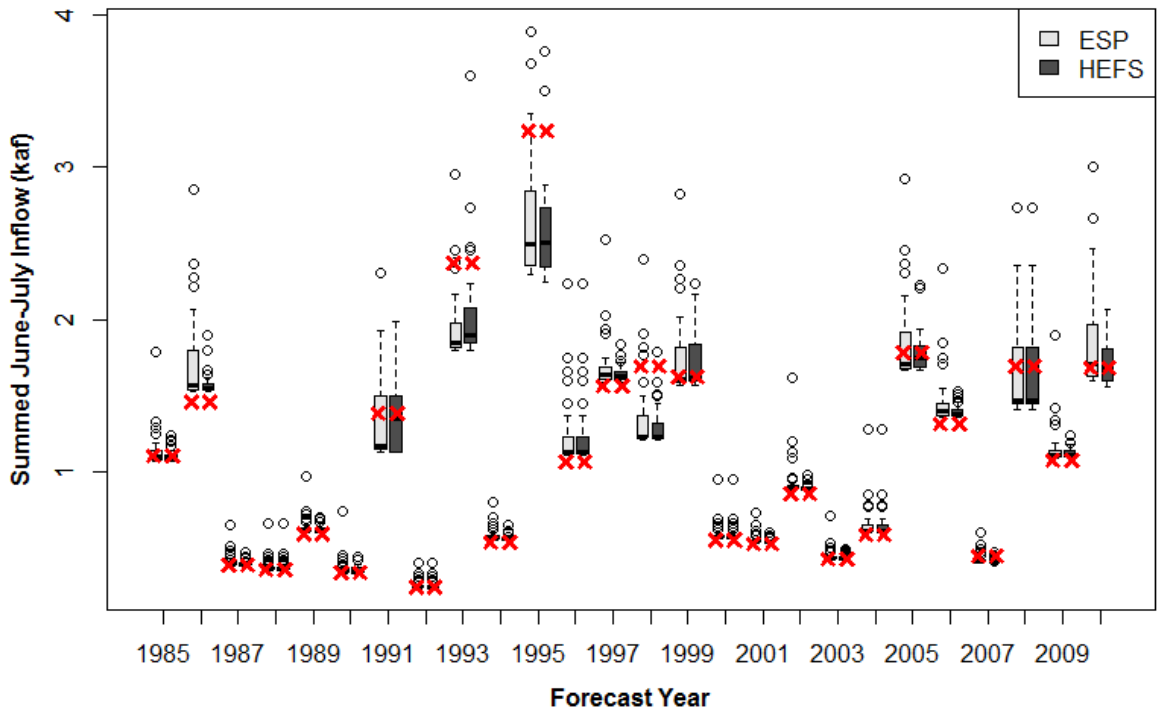


Figure 8: Summed April-July ESP and HEFS Inflow (Little Dell - June Forecast)

5.0 Description of Models

This chapter describes the three models used to analyze the benefit of incorporating ESP and HEFS forecasts into the Parley's System for decision making. A simulation model can create a status quo record of operations without the use of forecasts, or a "baseline scenario". An optimization model can create a "perfect information scenario" which describes ideal operations using a single, perfect forecast. Finally, a decision support system is built integrating the simulation and optimization approaches that can incorporate the ESP and HEFS forecasts.

5.1 Baseline Scenario

An idealized simulation model was built to create output data that could be considered a baseline scenario to compare the results of decision support tools. The Mountain Dell and Little Dell Reservoirs were combined into a single reservoir system. The two main inflows were similarly combined into a single inflow.

The simulation model was constructed in R and uses a variety of operational rules to determine releases. Two types of releases are made in this system. The first release is denoted as a "culinary release" and describes water that is drafted directly to a water treatment plant for municipal and industrial water supply. The second release is a "bypass release" that goes through Parley's Pipeline. These bypass releases are made to lower the pool elevation for flood control. Releases out of the Parley's pipeline are limited because large releases can cause flooding issues downstream.

Operational decisions are typically made on a weekly or daily time step depending on the time of year and flood risk. A nomograph developed by the US Army Corps of Engineers is used to determine ideal storage levels in the reservoirs during the year to reserve an adequate flood pocket. The nomograph is an integral part of how the dams are operated for flood control.

Depending on the time of year and the predicted amount of inflow from April through July, a certain volume of the reservoir is made available to catch potential floods.

The nomograph requires the use of a forecast, yet the baseline run could not explicitly use a forecast. A modified rule curve was created to address this. The 17 years of historic storage data were divided into 52 weeks in each year. The average combined storage in each week was calculated and plotted (Figure 9: Salt Lake City Combined Historic Storages by Week). The median of the historic data became the rule curve for the baseline simulation model. This 50% rule curve captures the seasonality of how the reservoir has been operated historically and inherently contains some forecast information and information from the nomograph.

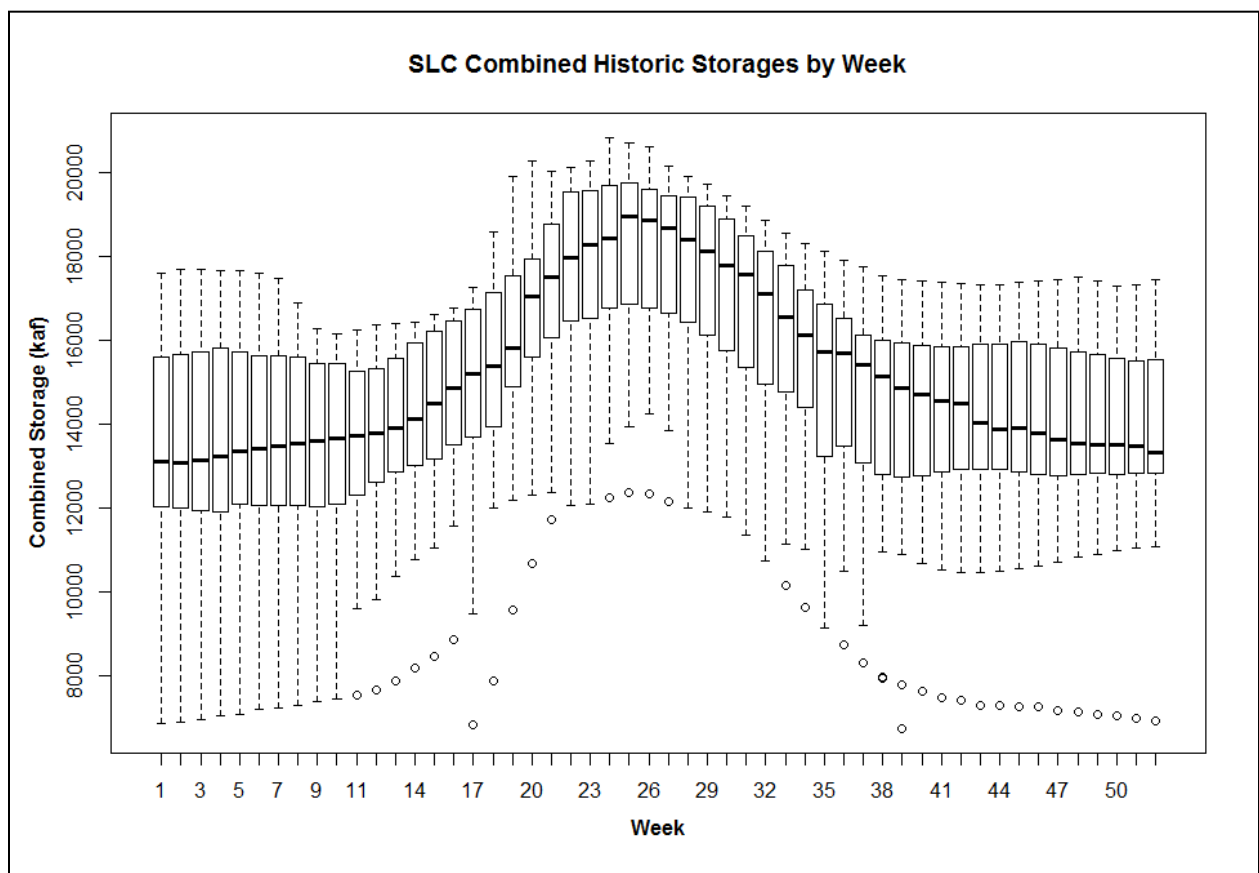


Figure 9: Salt Lake City Combined Historic Storages by Week

For each time step, the simulation model determines whether releases should be made for drinking water, flood protection, or both. For this analysis, the culinary water demands were assumed constant from year to year. In reality, the Parley's System is a supplementary source of drinking water rather than a primary source, so demands are not necessarily constant between years. A drinking water demand rule curve was developed by averaging the water delivered historically from 1994 to 2011 by Julian day. Years 2001, 2002, and 2003 were not included in this average, because water use restrictions were in place during those years due to a severe drought. This gave an approximate demand curve for the model to meet in every year of the simulation.

The simulation model uses a very basic set of rules to make releases. Inflows and the current state of the reservoir on any given day were considered, and then the releases were calculated. (For detailed information about simulation rules, please refer to Appendix C). For the baseline model the historic data was run through the simulation model. This model is very reactionary and is designed to keep the combined storage volumes to as close to a rule curve as possible. It does not have the benefit of forecast information. The decision support framework will try to improve upon the output from this baseline simulation by using the ESP and HEFS forecasts.

5.2 Perfect Scenario

To determine how the system would be operated if perfect forecasts were available, an optimization model was developed that incorporated the observed inflows for a one year horizon. This model can calculate the best operations possible because it is getting one year of perfect foresight, a level of forecast skill not available in reality. This exists to provide a contrast of the

baseline model, which does not use forecasts and determines releases based on one day of information.

The objective of the optimization model was to minimize penalties. Penalties were accrued when the combined storage was above or below a target range, set at 8,000 and 20,000 af. Bypass penalties would be collected when more than 200 cfs were released on any given day. Water supply penalties accumulated when the full drinking water demand given was not fully met. The objective function was:

$$\begin{aligned}
 \text{MIN } & \sum \text{Storage penalties over 20,000 af} \\
 & + \sum \text{Storage Penalties under 8000 af} \\
 & + \sum \text{Bypass Penalties over 200 cfs} \\
 & + \sum \text{Drinking Water Demand Penalties}
 \end{aligned}$$

The major constraints in the model were that the minimum combined storage was 3,000 af and the maximum capacity was 21,200 af. The bypass releases could not be less than 0 cfs and could not exceed 300 cfs.

Weights were multiplied to components of the objective function to make the optimization model results more consistent with historic operations. The heaviest weights were set on the storage penalties because staying within the target range was considered the priority. A smaller weight was placed on the bypass penalties because flooding downstream of the bypass is the second major priority. Supplying the full allocation of water was not given any additional weight. It is of the lowest priority because this system is a supplementary water source.

5.3 Integrated Simulation Optimization Model

The Decision Support System (DSS) that was found to be most effective in using the forecast information was a framework to guide operations that integrated both simulation and optimization techniques, similar to Alemu (2010). The process starts when the first forecast is available on January 1 (Figure 10). In January, the initial storage is defined as the current observed storage. Thirty traces of the forecast are individually optimized, and the result is an ensemble operational output for the combined storage, bypass release, and culinary release is given (Figure 11). The optimization model uses the same rules as previously described in Section 5.2. A decision maker could look at the operational ensemble and decide how he wants to operate his dam for the next month based on the 30 perfectly optimized futures. To simplify the model calculation, the average of both the bypass release policy and the water release policy for the thirty optimized traces is taken and used as the operational policy in the simulation for one month. The simple simulation mass balance, $S_{t+1} = S_t + Q_{in,t} - R_{bypass,t} - R_{demands,t}$, is calculated daily using the optimized release rules and the observed inflow for one month.

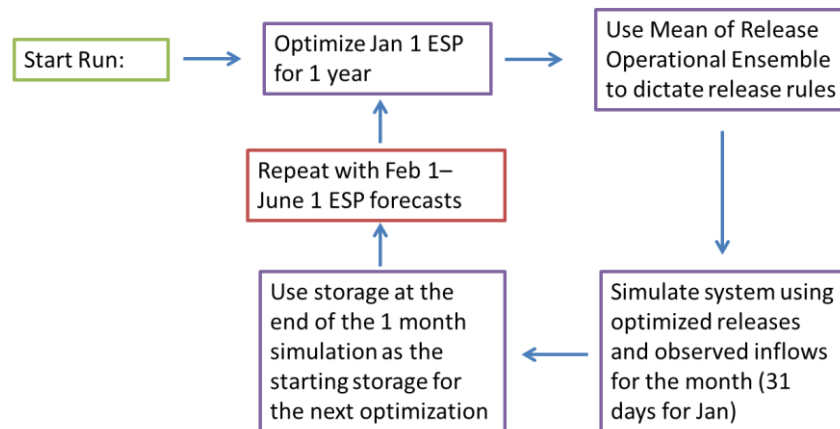


Figure 10: Schematic of the Integrated Simulation Optimization Process

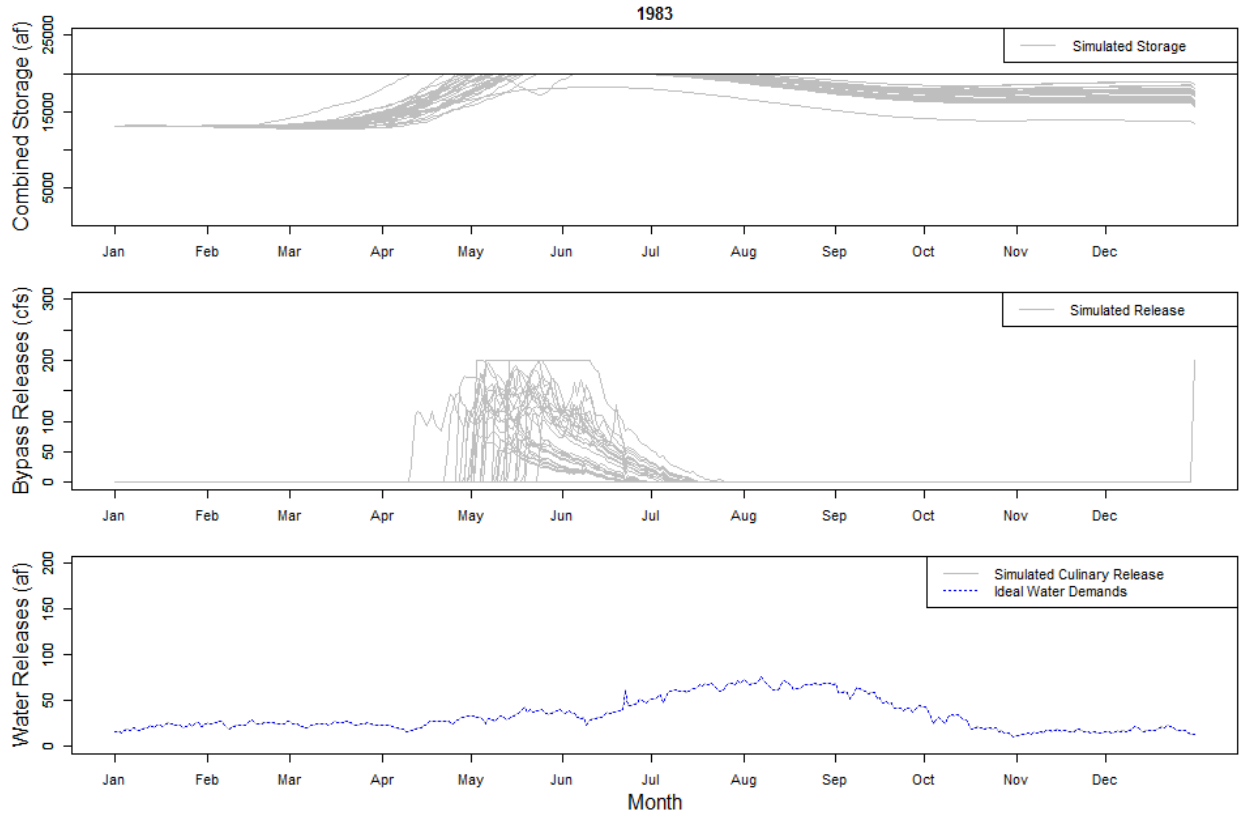


Figure 11: Optimization of January 1983 ESP Forecast

Once the model is simulated for the first month using the optimized operational policy and the observed inflows, the ending simulated storage is taken and used as the initial storage for the next round of optimization. The simulation shows what storage the operator would have ended with had he used the forecast. Using the simulation model as the next initial condition of storage, the second month forecast is optimized. This same process is repeated until all of the forecasts in a year are used, deriving an operational policy from the optimization output and simulating those optimized releases with the observed input. This process is repeated for each of the monthly forecasts, January through June. The June simulation is calculated for the remainder of the calendar year, through December 31, rather than just one month. Figure 12 shows the final operational output after completing this process for the year. Metrics are calculated on this

output to provide insights into whether these forecasts contain useful information for decision makers.

In addition to a simple mass balance, there are some “common-sense” simulation operating rules that can override the operational policies set by the optimization model in potential flood stages. These “common-sense” simulation rules are a necessary corrective action because of the iterative structure of the integrated simulation and optimization DSS. A mass balance approach to the simulation portion of the DSS causes reservoir storage levels to exceed the capacity in certain high flow years. If the starting storage of an optimization run exceeds the reservoir capacity, the optimization model cannot perform any calculations because of the way the model is constrained. These overriding rules aim to keep the reservoir storage from exceeding capacity when actual releases are contributing to the mass balance simulation (Werner et. al. 2013). Specifically they state that with a storage level above 20,000 af, the culinary releases had to be meeting the full drinking water demand. The Parley’s Pipeline is forced to release 200 cfs which is about the maximum possible release before there are potential flooding issues. With a storage above the 21,200 af capacity, the simulation is forced to release however much water out of the Parley’s Pipeline that it needs to in order to not exceed capacity, mimicking spill and allowing the optimization model to run in the subsequent month.

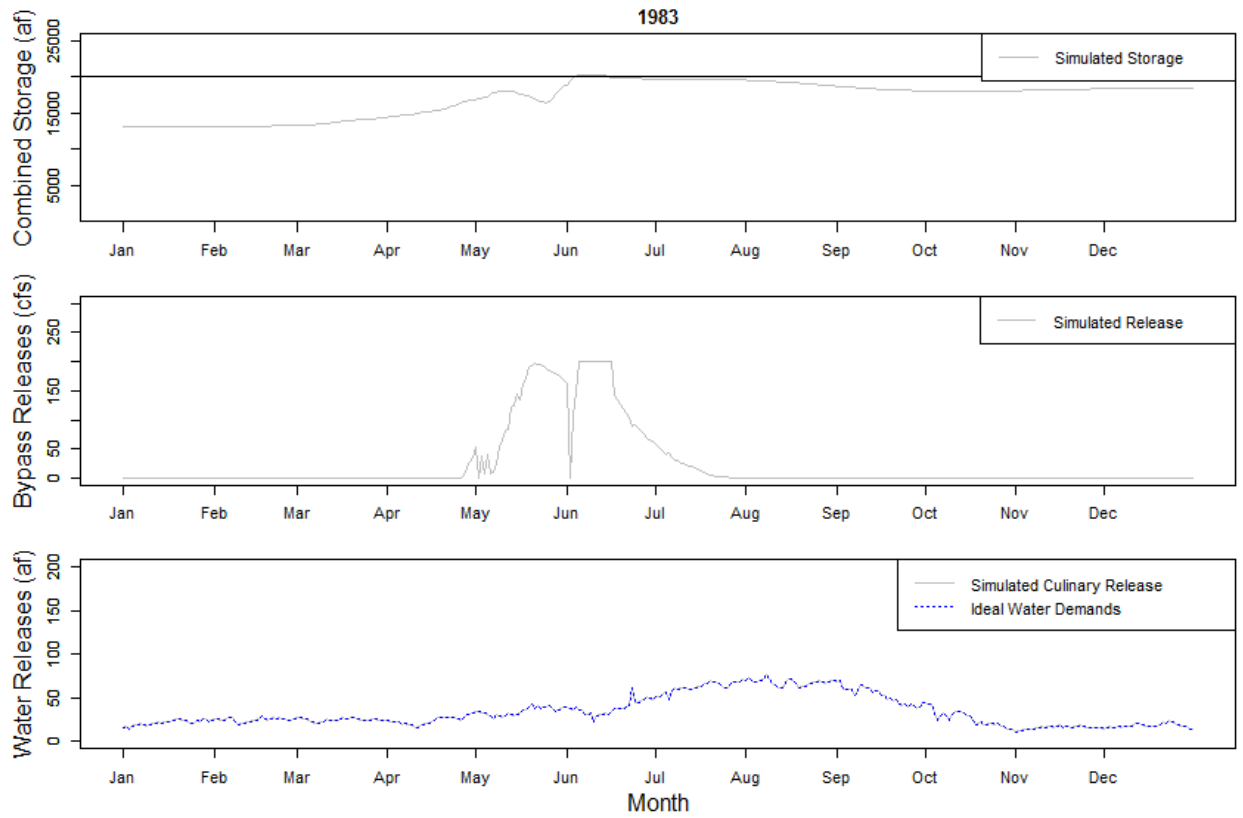


Figure 12: DSS ESP Operational Output for 1983

6.0 Results

This chapter reviews the results from the analysis of incorporating forecast information on the Parley's System. The metrics used to evaluate model performance are presented. The results from the baseline simulation model are compared to metrics calculated in from the observed record to show that the baseline is an adequate version of real system operation. Finally, the DSS is evaluated when both ESP and HEFS forecast data are used as input. These results are compared with the baseline metrics and the perfectly optimized scenario metrics.

6.1 Metrics

To determine the value of the using forecast products in managing the Parley's System, performance metrics meaningful to managers are identified and quantified. The two most important functions of the Salt Lake City Parley's System are reliably providing municipal water supply and maintaining flows below an operational target, which would cause downstream flooding. Balancing these objectives requires maintaining reservoir pool levels as high as possible while providing sufficient storage to reduce floods. These metrics of system performance are calculated for five scenarios: the observed record, output from the baseline simulation model, output from DSS using the ESP data, output from the DSS using the HEFS data, and output from optimization model using perfect forecasts. By comparing metrics in this experimental design, the potential benefits of using forecasts can be quantified in a systems decision making context.

6.1.1 Water Reliability

Reliability is calculated by summing the number of days the releases to the culinary pipeline are equal to the ideal demand and dividing by the number of days per year. This gives

the percent of days in a year that the full water demand was delivered. Ideally, this metric would be at 100% every year.

To capture the magnitude of shortfalls when they occur, the total volume of water delivered for the entire year was calculated. Knowing the volume delivered indicated the vulnerability of the system and can be a better indicator of the severity of a water shortage than reliability alone.

6.1.2 Bypass Release Metrics

The bypass release metrics track the number of days the bypass release is at or greater than 200 cfs, the threshold of the bypass releases managers are comfortable with due to potential flooding in downtown Salt Lake City. These metrics gauge how frequently each version of the model is nearing or exceeding this threshold.

6.1.3 Storage Metrics

In addition to the threat flooding poses when bypass releases exceed 200 cfs, managers do not want the reservoir storage to potentially overtop, which occurs when the capacity of the reservoir, 21,200 af, is exceeded. This concern is evaluated by calculating the number of days storage is in excess of 20,000 af, the threshold of maximum storage managers are comfortable with having at any time.

To capture the magnitude of exceedances of the maximum storage target, the storage volume over 20,000 af will be summed for each year. When examined in conjunction with the number of days the 20,000 af threshold is exceeded, this metric clarifies how much of a flooding concern there is from one year to another.

The maximum pool volume in May and June is also calculated for each year. The peak storage typically occurs in one of these two months. Having an idea of potential peak storage levels is useful to managers because they want to have the reservoir storage peak as full as possible without crossing into a threshold that could put them at risk for flooding. This metric should also be used in conjunction with the days exceeding the 20,000 af threshold to understand the severity of flood risk.

The volume stored in the reservoirs is also recorded on August 1, which is the end of the season where the greatest drinking water demands are. Higher storages on August 1 mean that the following year will start out the season with more water. A good management strategy will not only perform well in a given year, but leave the system in a state to be successful in following years.

6.2 Comparing Observed Metrics to the Baseline Simulation Output

The historic record of Salt Lake City's reservoir storages, gaged inflows, and operational releases and flows was available for sixteen years, from 1994 to 2009. In reviewing the records, several inconsistencies were found related to the mass balance which may be a result of recordkeeping errors. To address this challenge, an idealized simulation model was constructed to create output data that can be considered a baseline scenario with which results of the DSS could be compared. This section presents an overview of the baseline model's performance compared to metrics calculated on the observed record.

6.2.1 Water Reliability

Water reliability for the baseline simulation is generally higher than the observed record and is able to deliver a larger quantity of water (Figure 13 and Figure 14). The water demands were idealized from an average of all of the year's deliveries, so the simulation model only

releases up to a set amount, around 12.5 kaf, which was not the case in actual operation. In all of the years where the observed scenario delivers more water, the amount was outside of what the simulation model would ever deliver. For both reliability and the total volumes of water being delivered, the baseline simulation results generally follow the trends of the observed record. This shows that the simple simulation of the system is behaving in a realistic way.

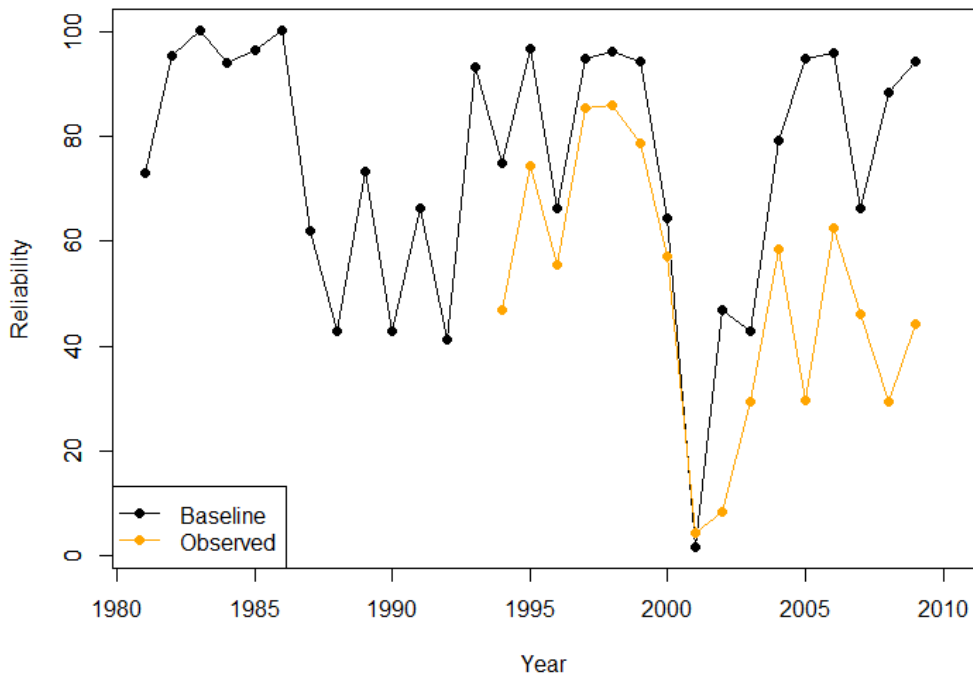


Figure 13: Water Supply Reliability - Baseline v. Observed

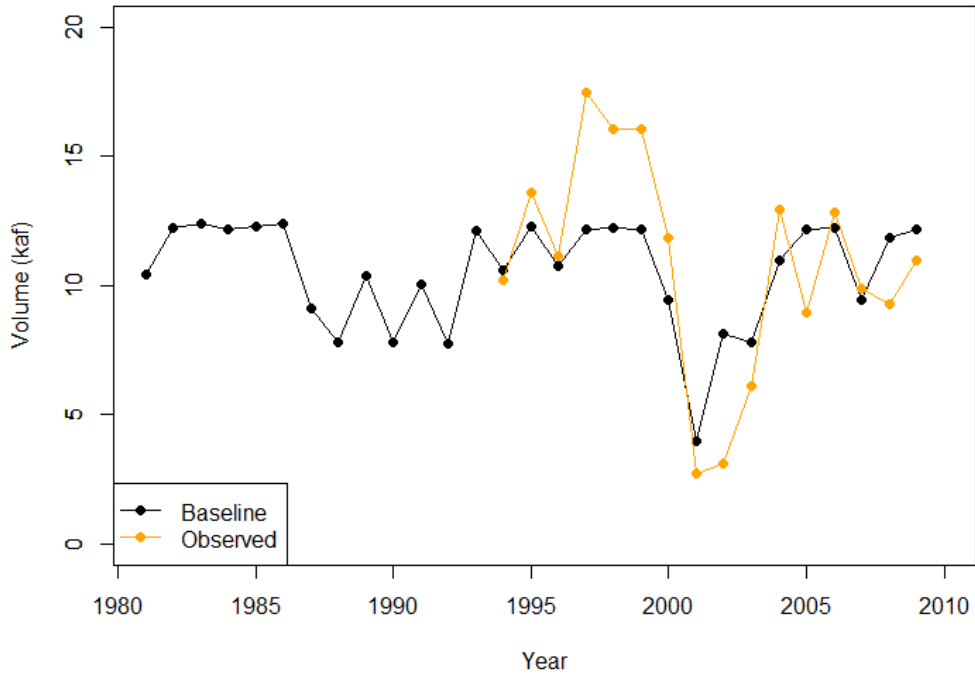


Figure 14: Volume of Water Delivered - Baseline v. Observed

6.2.2 Bypass Release Metrics

The trends of the bypass releases made are also very similar between the baseline simulation model and the observed record (Figure 15). Days where releases near or at the threshold of flooding were very similar. The baseline simulation was making realistic bypass releases relative to the flooding threshold.

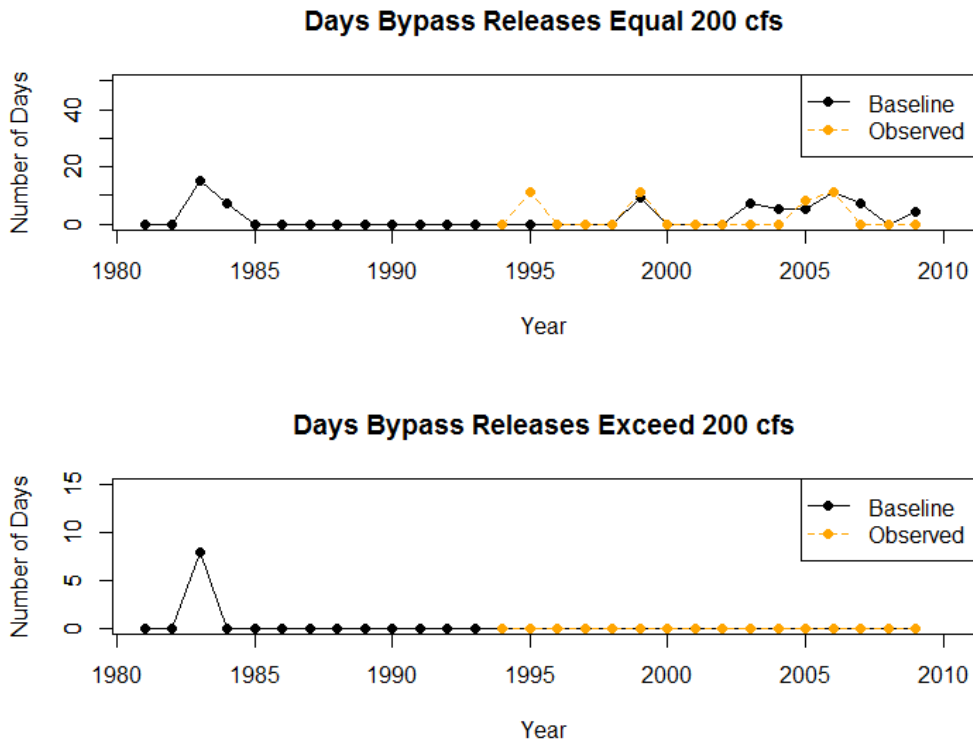


Figure 15: Bypass Releases - Baseline v. Observed

6.2.3 Storage Metrics

In addition to the threat flooding presents when bypass releases exceed their targets, managers want to preserve a pocket for catching large inflow to prevent overtopping. The baseline simulation minimizes the number of days the storage exceeds 20,000 af and the volume by which the threshold is exceeded when compared to the observed record (Figure 16 and Figure 17). For this metric, the baseline simulation does a better job than history in regulating high storage volumes.

By tracking the peak storages for May and June, managers can see if the reservoir is able to fill that season and understand if filling is causing an exceedance of the maximum target storage. The baseline simulation does a better job than history at having peaks that are high, but

not so high that flooding would be a major concern (Figure 18). In the low flow years in the early 2000s, the baseline model reproduces trends from the observed record well.

Figure 19 shows how the combined storages on August 1 produced by the simulation model compare to the observed record. The static rule curve in the baseline simulation determines that the ideal storage on August 1 should be about 17.5 kaf, so the simulation results generally do not exceed that threshold, when the observed values do at times. The simulation is able to reproduce a trend similar to the observed while showing improvement (higher storage) in some of the lower flow years, 2000 and 2003.

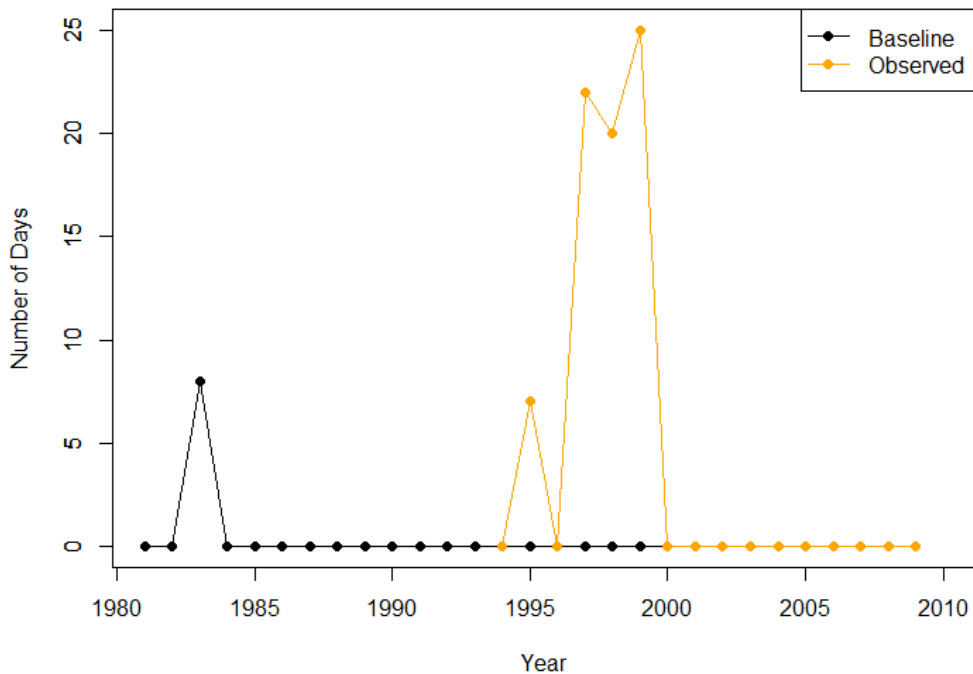


Figure 16: Days Combined Storage Exceeds 20,000 af - Baseline v. Observed

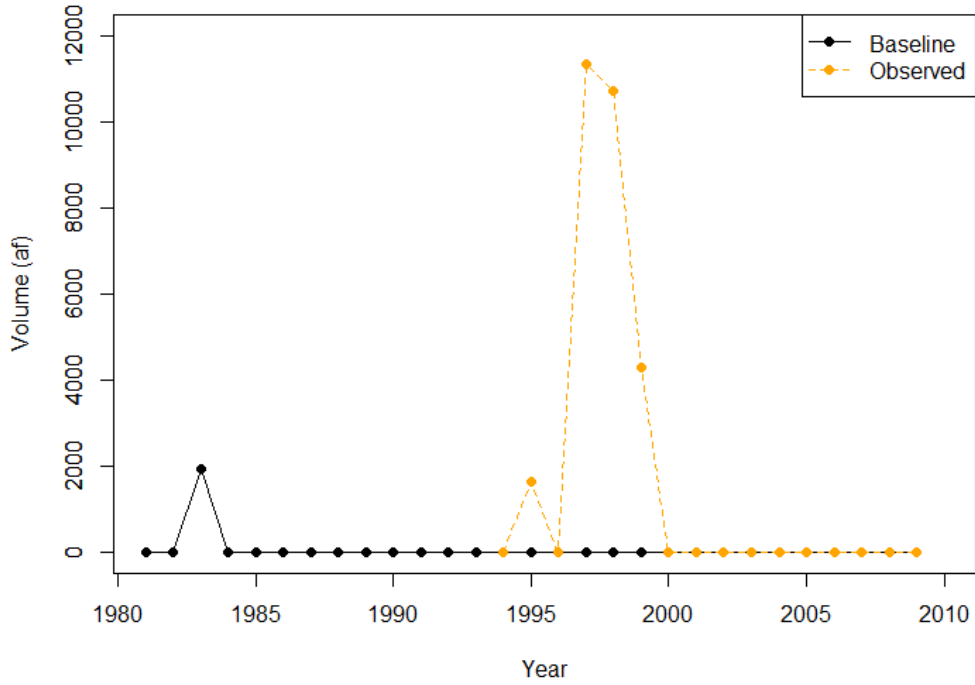


Figure 17: Total Summed Volume of Storage over 20,000 af - Baseline v. Observed

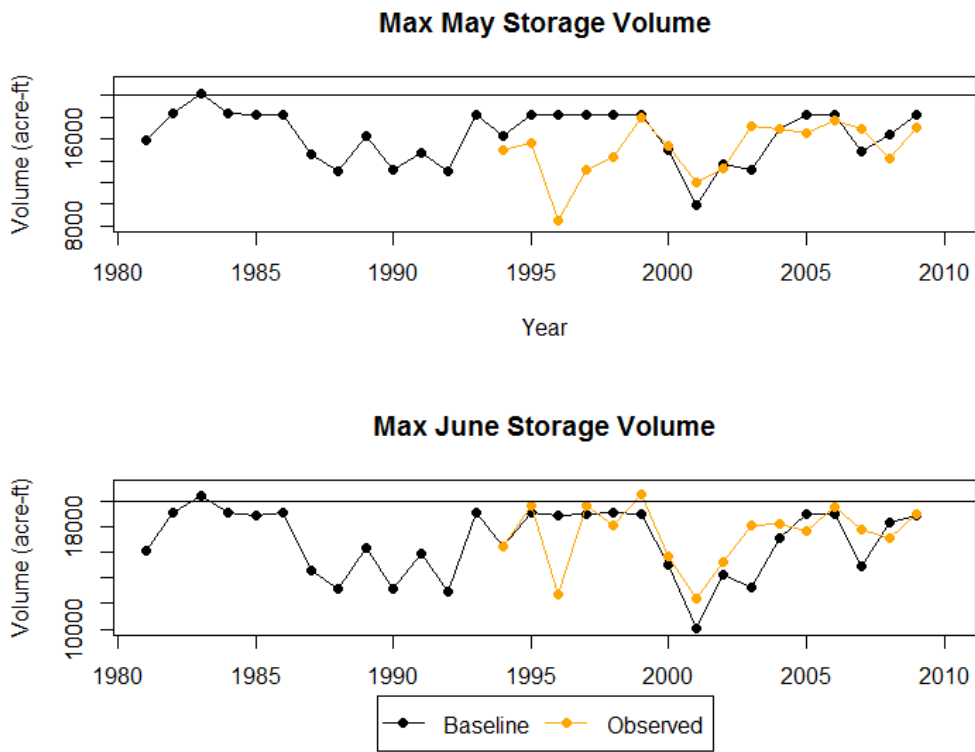


Figure 18: Maximum Storage Volumes - Baseline v. Observed

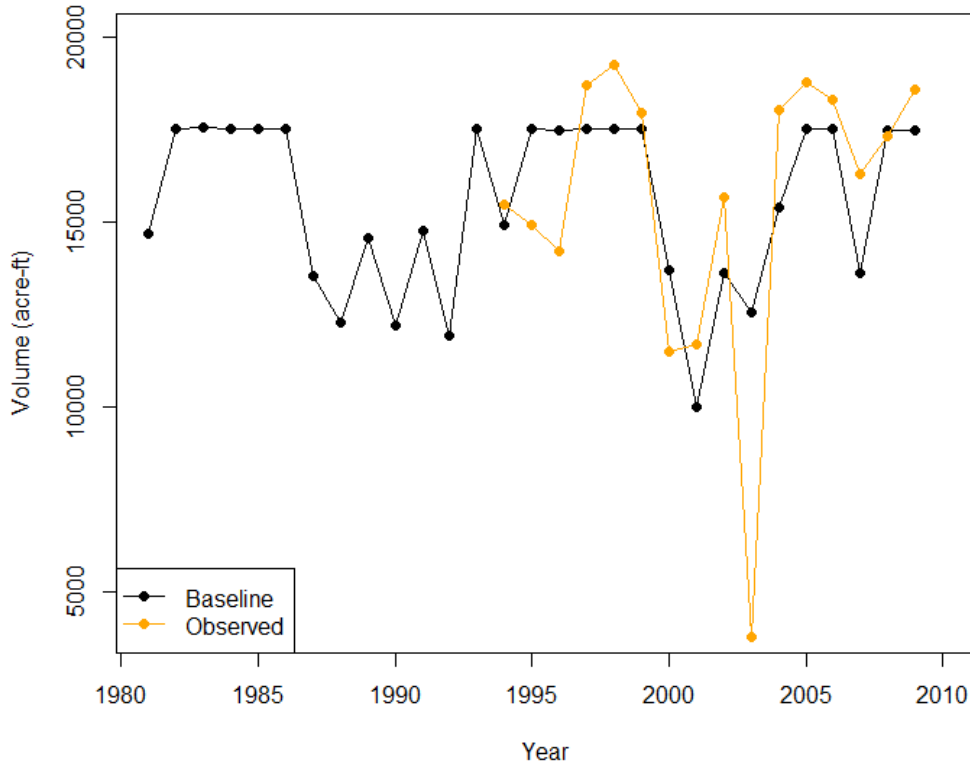


Figure 19: Aug. 1 Combined Storage Volume (af) - Baseline v. Observed

6.2.4 Conclusion from Baseline Simulation and Observed Record Comparison

From the comparison between the baseline simulation and the observed record, it seems like the simple simulation of the system is behaving in a realistic way, especially considering that it is an idealized model with many assumptions. The baseline simulation results are a good substitute for the observed record when showing the value forecasts can add to the Parley’s System.

6.3 Forecast Performance

A primary research objective is to determine the extent to which the ESP and HEFS forecasts provide an opportunity to improve system operations. To answer this question, ESP and HEFS forecast are used to inform decisions in the integrated simulation optimization DSS. These results are compared to the baseline simulation output and the perfectly optimized scenario. The

optimization of perfect information should perform the best of all of the scenarios. The value of applying forecasts into operations management can be quantified by how well the DSS, which incorporates the forecast information, can improve upon the baseline simulation.

6.3.1 Water Reliability

The DSS is able to improve water reliability, at least slightly, from the baseline simulation every year that was run (Figure 20). As expected, the perfect information scenario consistently schedules releases to have the highest reliability, but there are years (ie. 2004 to 2009) where using a forecast can improve reliability to the perfectly optimized amount. The DSS performance is not appreciably better or worse when using the ESP forecast or the HEFS forecast, but using the decision tool with forecasts can improve operations over the static rules of the baseline simulation.

Figure 21 presents the volume of water that can be delivered in each scenario, with the baseline delivering the least amount of water, and the perfect scenario delivering the most. This figure illustrates the complicated relationship between reliability and the amount of water delivered. For example, in years 1988 and 1990, the improvement for HEFS and ESP from the baseline reliability does not look very substantial, yet when examining the total volume of water delivered in those years, there were increases of about 1 kaf (12%). Similarly, the value of reliability in 1994 suggests that ESP forecasts were substantially outperforming the HEFS forecasts. However, the total volumes delivered with the two forecasts are quite similar. When looking at reliability and water supply in conjunction, it is clear that forecasts improved operations. The ESP and HEFS forecasts produce similar operational output schemes with neither being superior to the other.

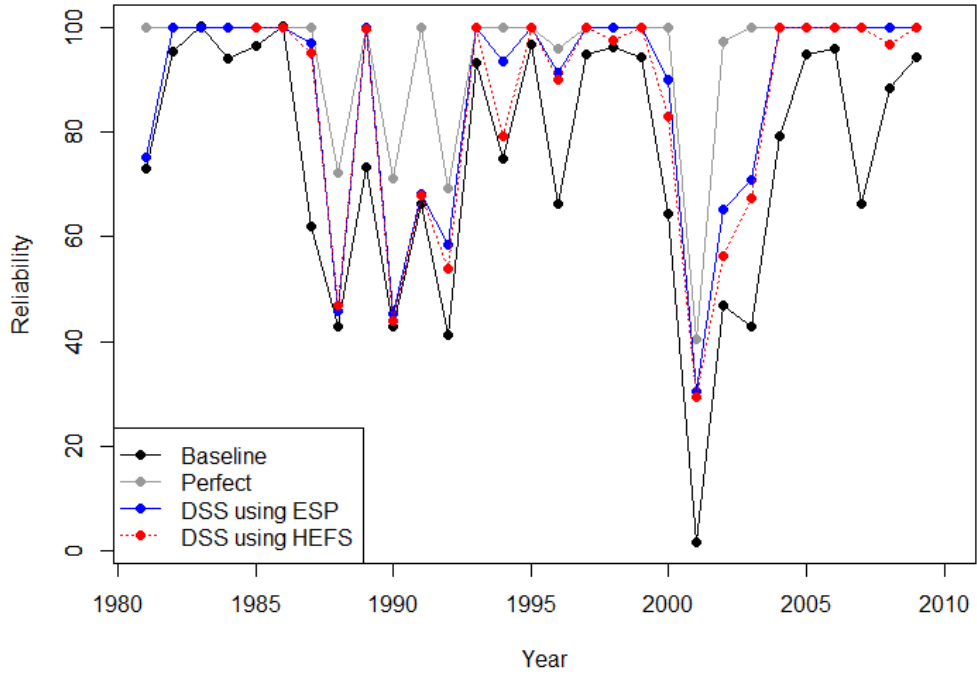


Figure 20: Water Supply Reliability - Forecast Performance

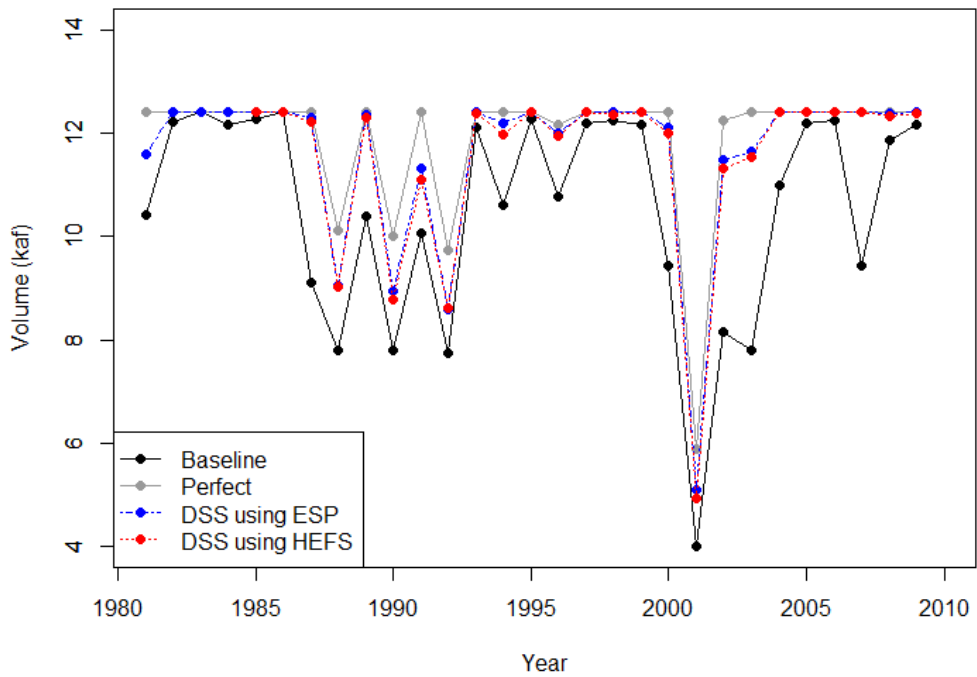


Figure 21: Volume of Water Delivered - Forecast Performance

6.3.2 Bypass Release Metrics

Results of the models indicated that the DSS never dictates a release schedule that causes the 200 cfs threshold to be exceeded (Figure 22). In each scenario, there are days where a full 200 cfs release is made, but minimizing exceedances is a higher priority. The only exceedance that occurred was in 1983 when the baseline rule curve scenario caused potentially damaging releases. Using the DSS method with the ESP forecast avoided having to make those types of high releases. (The HEFS dataset was not available for 1983.) Again, for the years where both datasets were available, the ESP and HEFS forecasts produce such similar operation output schemes, there seems to be no benefit in choosing either the ESP or HEFS forecasts product over the other.

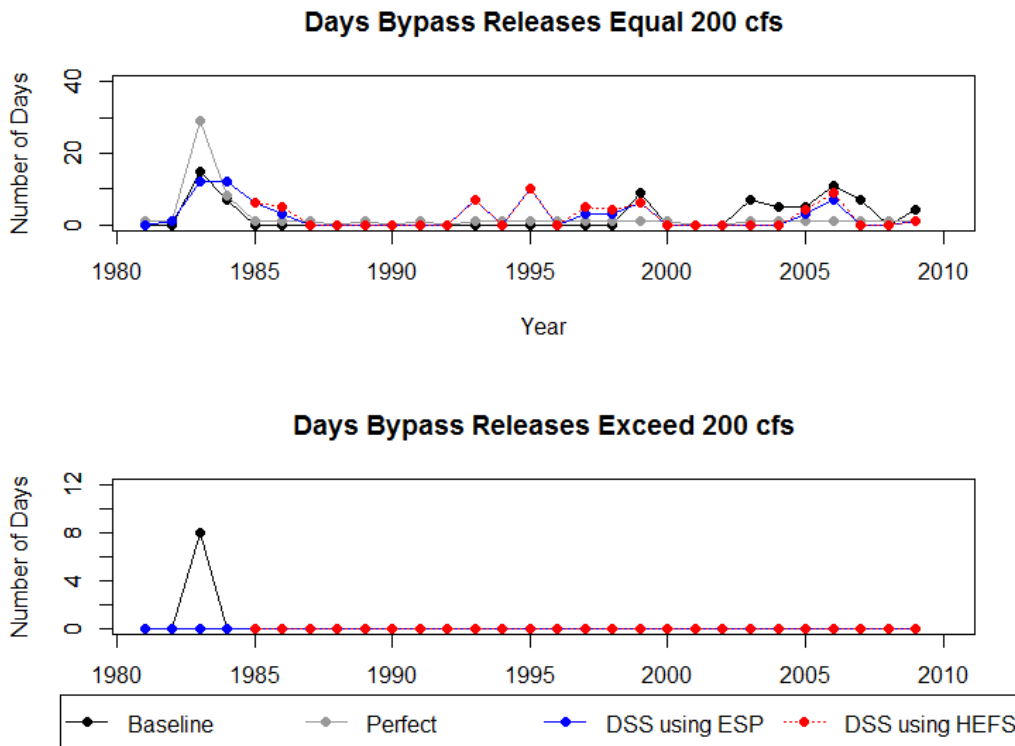


Figure 22: Bypass Releases - Forecast Performance

6.3.3 *Storage Metrics*

As discussed in Section 6.2.3, the baseline simulation does a good job of minimizing the number of days the storage exceeds 20,000 af and the volume by which the threshold is exceeded when compared with the observed record. Figure 23 shows that with a year of perfect information, the system can be optimized to never exceed the 20,000 af threshold. The decision support system's iterative methodology is inferior to the baseline and optimal scenarios with respect to this storage metric.

Although each of the ESP and HEFS ensemble forecasts can be perfectly optimized to never exceed the 20,000 af threshold, the simulation portion of the model sometimes calculates storages that exceed 20,000 af if both the timing and magnitude of the observed inflows are not forecasted perfectly. The "common-sense" simulation rules over-ride the optimization dictated release regime they day after storages exceed 20,000 af (as described in Section 5.3) as a way to "penalize" the forecast for not being accurate enough. The result is a higher number of days where the threshold is exceeded compared to the baseline and optimal runs, although in most cases the volume of the overages is fairly low. With more frequent forecast updates, the DSS may be able to perform as well as the baseline and perfect optimization scenarios.

During the highest inflow year on record, 1983, using the DSS with an ESP forecast would have caused storage to exceed 20,000 af for 12 days (Figure 24). The total summed volume of overages in those 12 days was about 3,000 af, or around 250 af/day when averaged and the peak storage was 20,338 af. It is somewhat risky to use over 20% of the 1,200 af flood pocket, but not dire when considering that there is room to increase bypasses amounts. Other than this most extreme case of 1983, there were low cumulative overages when the DSS model caused threshold exceedances.

Figure 25 shows the peak storages for May and June. The figure is scaled so the top of the y-axis is set to the reservoir capacity, 21,200 af. The black line at 20,000 af shows the flood pocket that would ideally be preserved at all times. Even in 1983, the peak overages are not a large percentage of the flood pocket. The perfect optimization and DSS have more flexibility to peak higher than the baseline simulation's rule curve, which allows the reliability to improve compared to the observed record, without imminent flood risk.

Figure 26 shows storage volume values on August 1. Generally the perfect optimization and DSS with ESP and HEFS have higher year end storages which leave managers in a better position for the following year. In low flow years, the DSS with ESP and HEFS actually produce some of the highest values. The perfect optimization will drain the reservoir down to increase reliability, because there is no end target storage on the optimization. The iterative nature of the DSS allows for adjustments that decrease reliability slightly, but leave the storage in a better ending state.

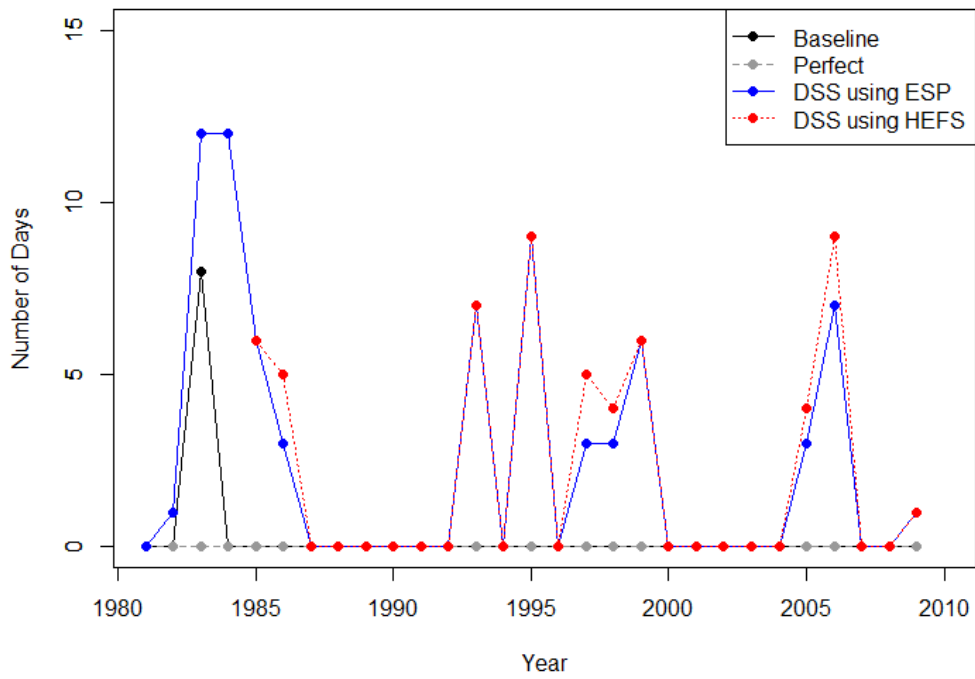


Figure 23: Days Combined Storage Exceeds 20,000 af - Model Performance

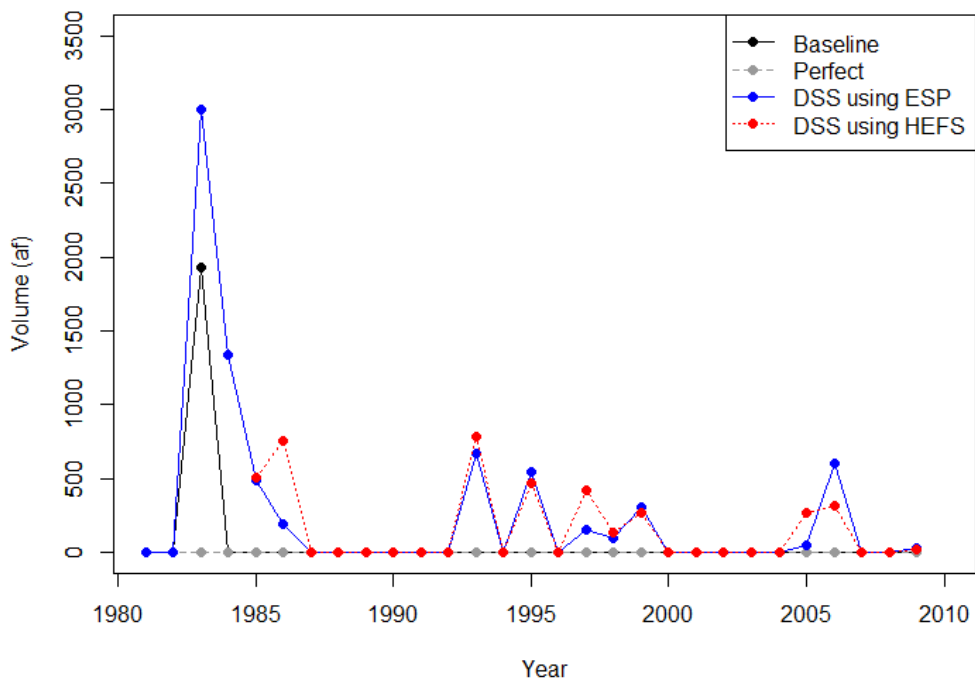


Figure 24: Total Summed Volume of Storage over 20,000 af - Model Performance

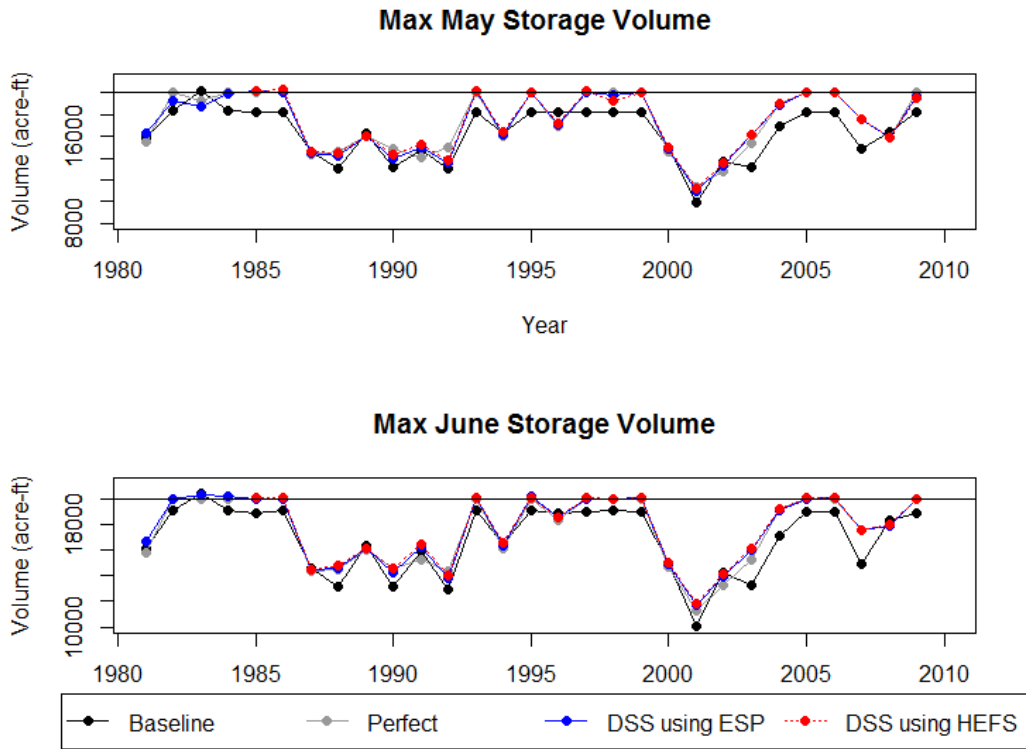


Figure 25: Maximum Storage Volumes - Model Performance

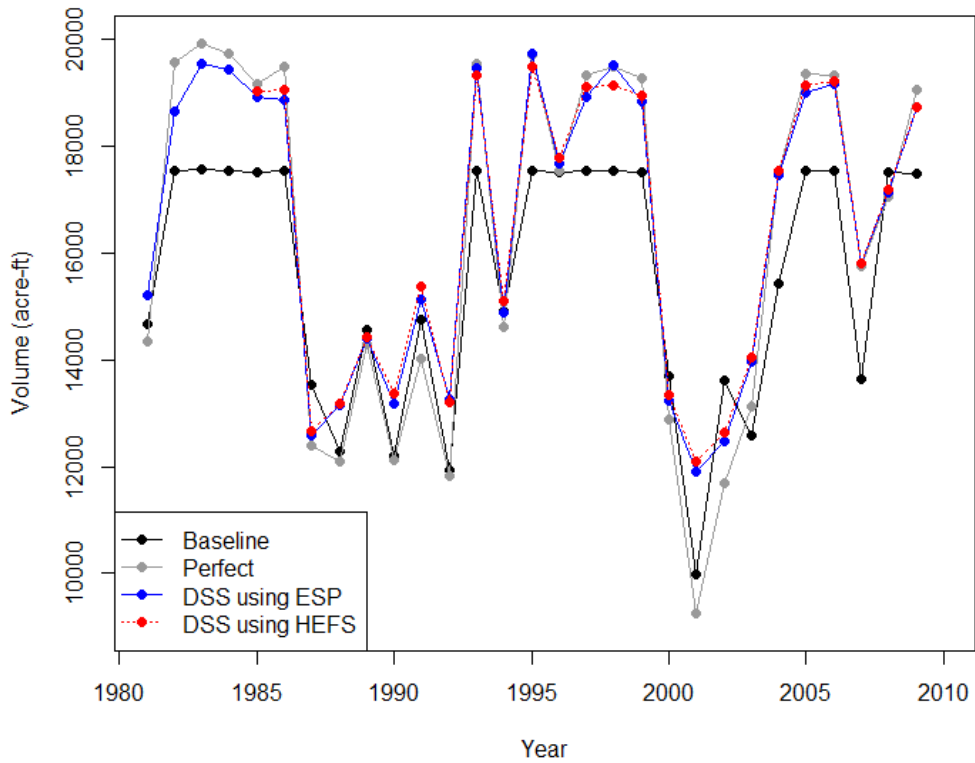


Figure 26: Aug. 1 Combined Storage Volume - Model Performance

7.0 Conclusions and Future Work

This research evaluates the accuracy and value of the HEFS and ESP streamflow products in a decision making framework using the operations of the Salt Lake City Parley's System as a case study. Metrics such as R-squared of the ensemble mean and RPSS were calculated to assess the accuracy of the streamflow forecasts outside of a decision making context. Three models were built to evaluate the forecasts in operations. A simulation model generated a baseline scenario, an optimization model created a perfect forecast scenario, and a DSS integrated the simulation and optimization approaches so the ESP and HEFS forecasts could be analyzed.

The DSS that incorporates ESP and HEFS forecasts shows improvement in many metrics related to drinking water reliability and flooding related to high bypass releases. The DSS did not improve the metric related to storages exceeding 20,000 af. This is not surprising because large flooding events are typically not caused by longer term climate signals, but shorter term convective processes. Forecasts issued on a monthly time frame will not be able to predict the timing and magnitude of flood events with enough accuracy to plan for them; weekly or daily forecasts are more appropriate to use for flood control.

However, this framework leaves enough flexibility to make bypass releases to moderate very high storage levels. In a real decision making scenario, release rules could be modified more frequently than on a monthly basis. Even in the highest flow years, there are not many days where even the full 200 cfs bypass release is being made. There is potential for a manager to evaluate and increase releases from what the operational output of the optimization model suggests to preemptively avoid high storage volumes when the forecasts under-predict the

magnitude of high inflows. For the Parley's System, it seems like being more aggressive in filling the reservoir would likely not cause major flooding issues.

Although the DSS performs well and provides beneficial information for the Parley's System, there is no appreciable difference between the two forecast products examined. The ESP and HEFS forecasts performed very similarly; even when isolating high or low inflow years, there is no distinct pattern for any metric where one product outperformed the other. Given the Parley's System location in Utah, it is not surprising that there was not a significant difference from the climate based HEFS from the climatology based ESP. If the HEFS product was tested in another location with stronger known climate signals, such as ENSO, the HEFS might have shown a larger improvement over ESP. It is encouraging that the HEFS is not performing worse than the ESP, so a manager could use it with the understanding that the quality of the forecast would not be worse than the ESP product.

The research results include the following:

- The R-squared values showed the mean of the ensemble is not a good predictor of observed April – July runoff until the April forecast.
- The RPSS calculations show that the forecasts are more informative than climatology for all months.
- The quality of the forecasts improves as lead time decreases.
- The DSS framework successfully incorporates ESP and HEFS forecasts and improves many metrics related to drinking water reliability and high bypass releases.
- The DSS framework did not improve the metric related to storages exceeding 20,000 af because of the iterative model structure. With more forecasts updated more frequently or different “common-sense” simulation rules this metric may be improved for the DSS.

Implementing forecasts into decision frameworks for reservoir operations is something that should be done based on each unique site. The 30 year reforecast dataset is a necessary tool to understand how the products perform in each individual system. Future work should include testing both ESP and HEFS forecast products on other sites in different regions of the U.S. with different climates. Other sites in other parts of the country could likely find that the climate informed HEFS is more useful than ESP. If this is the case, the focus of research in implementing forecast products should shift to the HEFS data since climate stationarity is no longer a valid assumption.

References

- Alemu, E. T., Palmer, R. N., Polebitski, A., Meaker, B. (2011). Decision Support System for Optimizing Reservoir Operations Using Ensemble Streamflow Predictions. *J. Water Resour. Plann. Manage.*, 137(1), 72–82.
- Barnston, A. G. et. al. (1994). Long-Lead Seasonal Forecasts—Where Do We Stand?. *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114. doi: [http://dx.doi.org/10.1175/1520-0477\(1994\)075<2097:LLSFDW>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2)
- Cayan, D. R., Redmond, K. T., Riddle, L. G. (1999) ENSO and Hydrologic Extremes in the Western United States*. *J. Climate*, **12**, 2881–2893. doi: [http://dx.doi.org/10.1175/1520-0442\(1999\)012<2881:EAHEIT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2)
- Day, G. N. (1985). Extended Streamflow Forecasting Using NWSRFS. *J. Water Resour. Plann. Manage.*, 111(2), 157–170.
- Glantz, M. H. (1982), Consequences and responsibilities in drought forecasting: The case of Yakima, 1977, *Water Resour. Res.*, 18(1), 3–13, doi:10.1029/WR018i001p00003.
- Gobena, A.K., Gan, T. Y. (2010) Incorporation of Seasonal Climate Forecasts in Ensemble Streamflow Prediction System. *J. Hydrology*, 385: 336-352. DOI: 10.1016/j.jhydrol.2010.03.002
- Gong, G., Wang, L., Condon, L., Shearman, A. Lall, U. (2010). A Simple Framework for Incorporating Seasonal Streamflow Forecasts into Existing Water Resource Management Practices. *JAWRA Journal of the American Water Resources Association*, 46: 574–585. doi: 10.1111/j.1752-1688.2010.00435.x
- Hagedorn, R., Hamill, T. M., Whitaker, J. S., (2008). Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures. *Mon. Wea. Rev.*, **136**, 2608–2619. doi: <http://dx.doi.org/10.1175/2007MWR2410.1>
- Hamill, T. M., Hagedorn, R., Whitaker, J. S. (2008) Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, **136**, 2 620–2632. doi: <http://dx.doi.org/10.1175/2007MWR2411.1>
- Hamill, T. M., Whitaker, J. S. (2006) Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application. *Mon. Wea. Rev.*, **134**, 3209–3229. doi: <http://dx.doi.org/10.1175/MWR3237.1>

- Hamill, T. M., Whitaker, J. S, Wei, X. (2004) Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447. doi: [http://dx.doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2)
- Lee, S.-Y., Hamlet, A. F., Fitzgerald, C. J. and Burges, S. J. (2009). Optimized Flood Control in the Columbia River Basin for a Global Warming Scenario. *J. Water Resour. Plann. Manage.*, 135(6), 440–450.
- Lee, S.-Y., Hamlet, A. F., Fitzgerald, C. J. and Burges, S. J. (2011), Methodology for Developing Flood Rule Curves Conditioned on El Niño-Southern Oscillation Classification. *JAWRA Journal of the American Water Resources Association*, 47: 81–92. doi: 10.1111/j.1752-1688.2010.00490.x
- Moradkhani, H. and Meier, M. (2010). Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis. *J. Hydrol. Eng.*, 15(10), 744–762.
- Najafi, M. R., Moradkhani, H., Piechota, T. C. (2012) Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *J. Hydrology*, 442: 105-116. DOI: 10.1016/j.jhydrol.2012.04.003
- O'Connor, R. E., Yarnal, B., Dow, K., Jocoy, C. L. and Carbone, G. J. (2005), Feeling at Risk Matters: Water Managers and the Decision to Use Forecasts. *Risk Analysis*, 25: 1265–1275. doi: 10.1111/j.1539-6924.2005.00675.x
- Rayner, S., Lach, D., Ingram, H. (2005) Weather forecasts are for wimps: Why water resource managers do not use climate forecasts. *Climatic Change*, 69: 197–227.
- Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. Peña Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker (2013) The NCEP Climate Forecast System Version 2. *J. Climate*. doi:10.1175/JCLI-D-12-00823.1
- Werner, K., Averyt, K., & Owen, G. (2013). River Forecast Application for Water Management: Oil and Water? *Weather, Climate, and Society*, 5, 244-253.
- Werner, K., Brandon, D., Clark, M., Gangopadhyay, S. (2004) Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts. *J. Hydrometeor*, **5**, 1076–1090. doi: <http://dx.doi.org/10.1175/JHM-381.1>

Wilks, D. (1995) *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

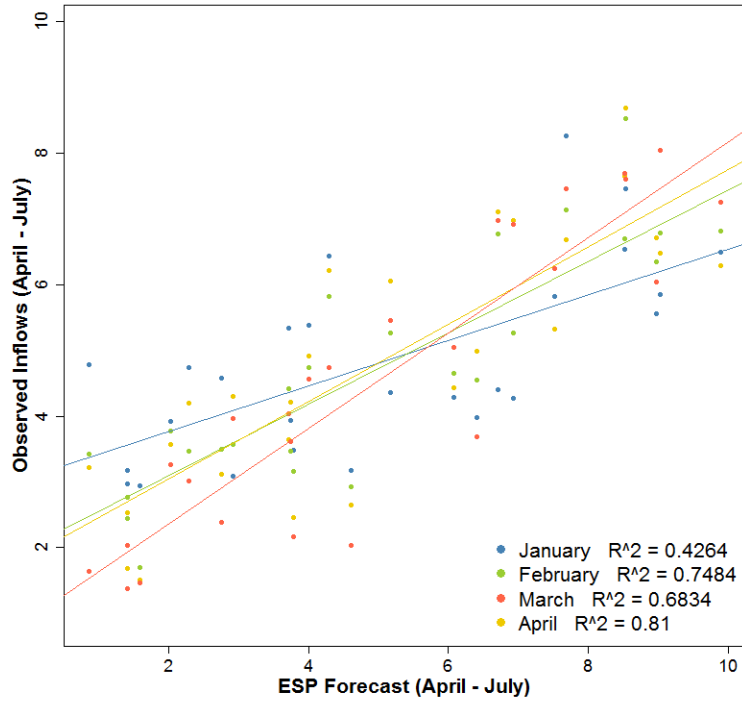
Wood, A. W., and D. P. Lettenmaier (2008). An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, 35, L14401, doi:10.1029/2008GL034648.

Wood, Andrew W., John C. Schaake (2008), Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. *J. Hydrometeor.*, **9**, 132–148.
doi: <http://dx.doi.org/10.1175/2007JHM862.1>

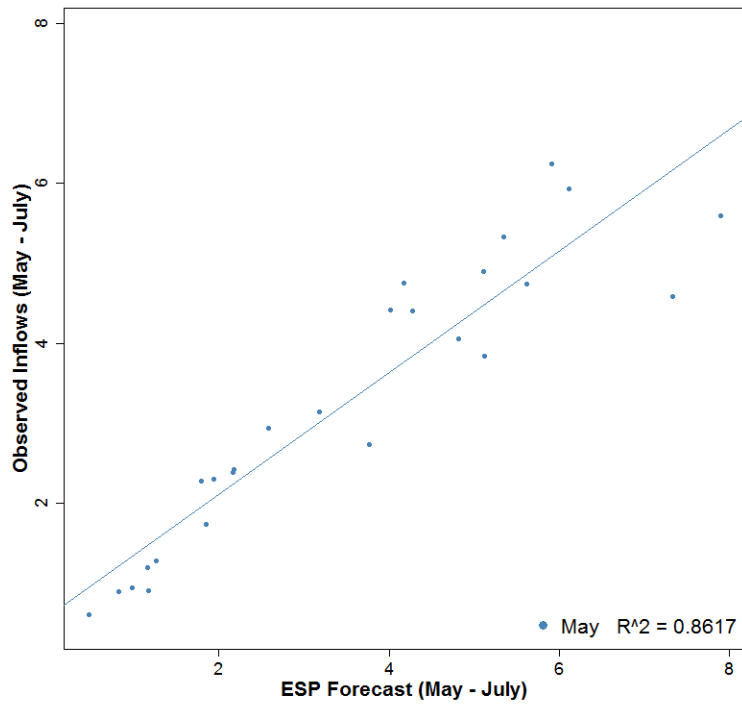
Yuan, Xing, Eric F. Wood, Joshua K. Roundy, Ming Pan. (2013) CFSv2-Based Seasonal Hydroclimatic Forecasts over the Conterminous United States. *J. Climate*, **26**, 4828–4847.

APPENDIX A

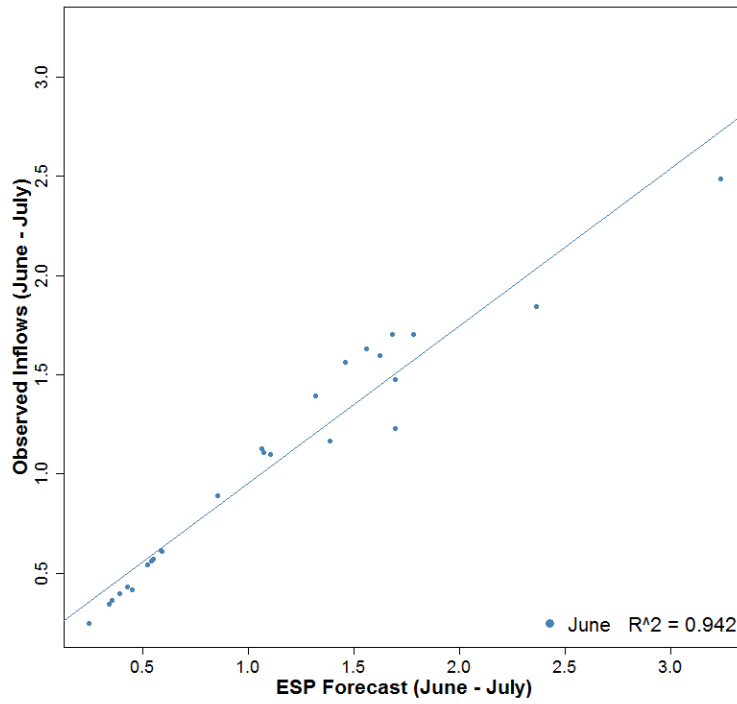
ESP Forecast vs. Observed Little Dell Inflow (April - July)



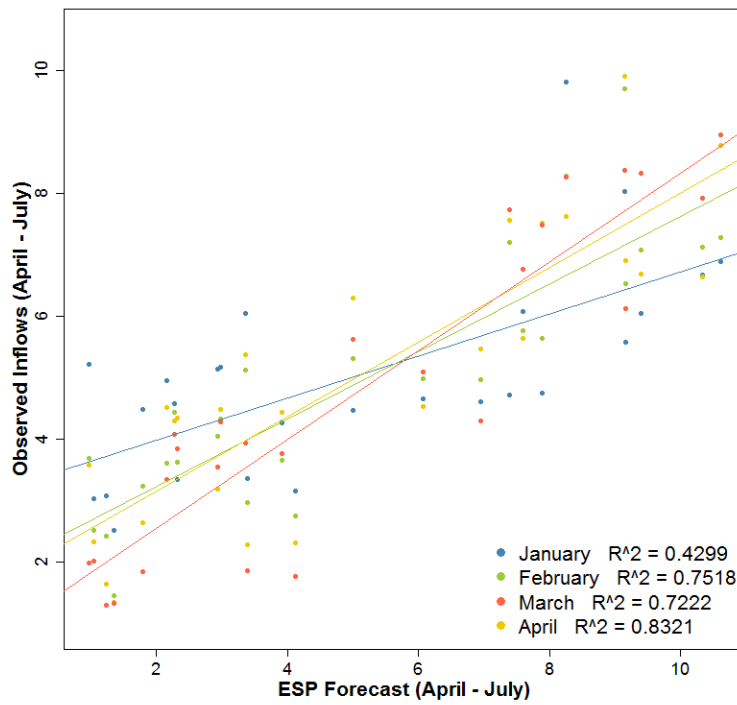
ESP Forecast vs. Observed Little Dell Inflow (May - July)



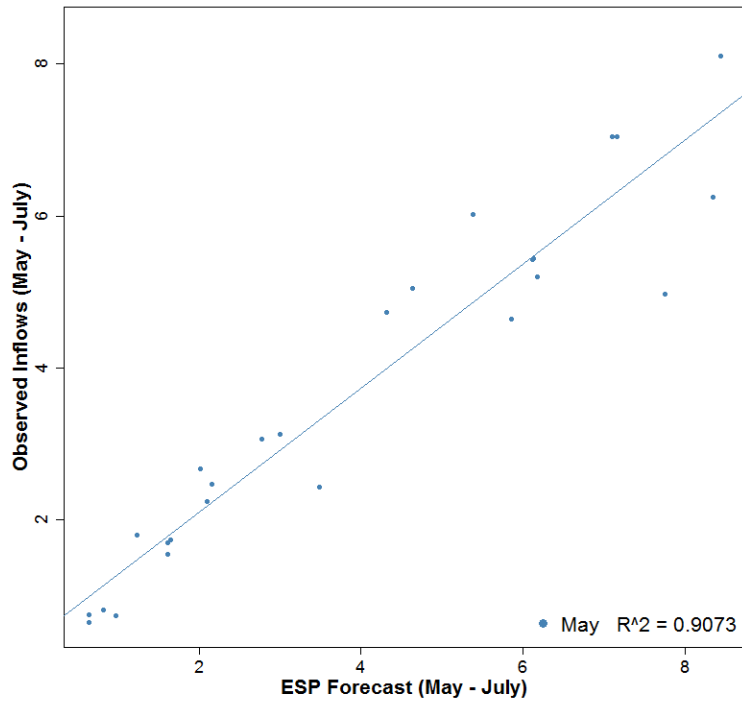
ESP Forecast vs. Observed Little Dell Inflow (June - July)



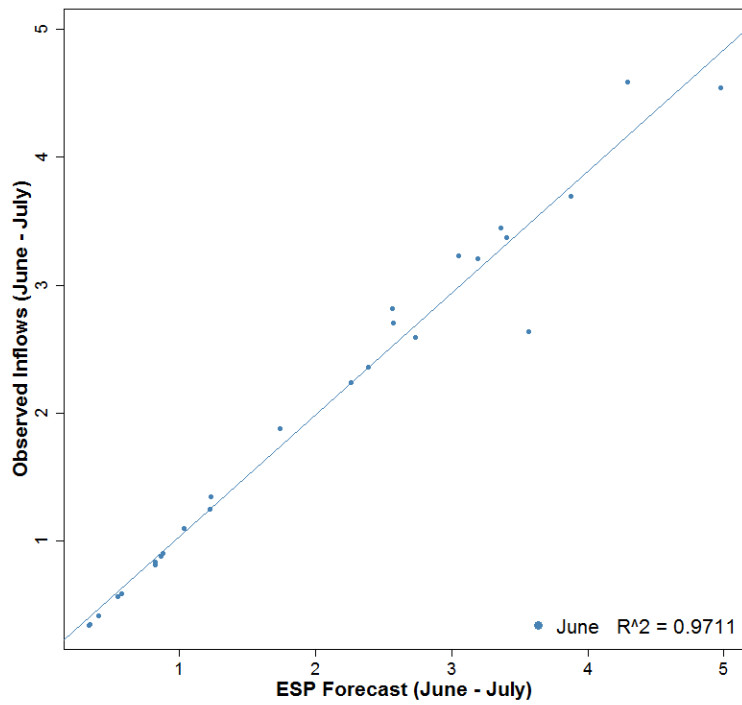
ESP Forecast vs. Observed Lamb Inflow (April - July)



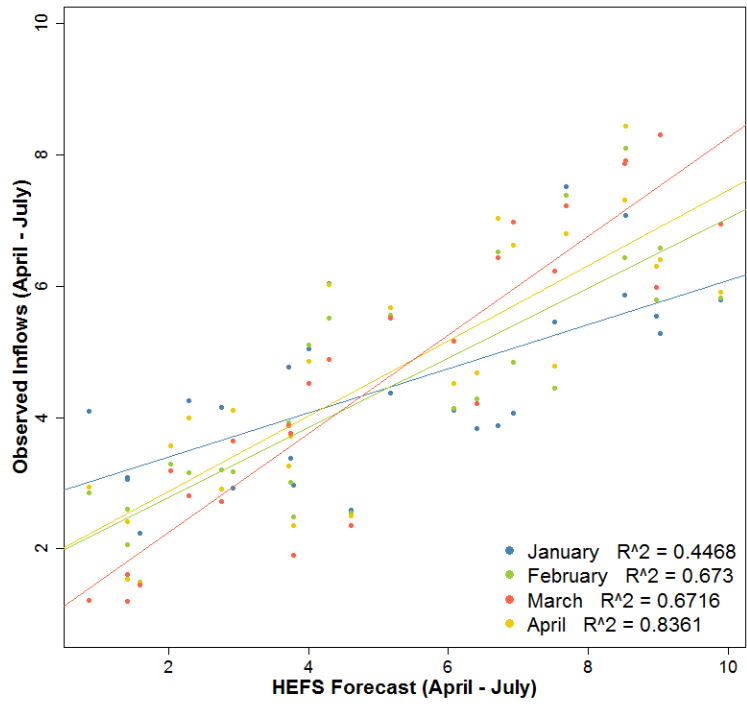
ESP Forecast vs. Observed Lamb Inflow (May - July)



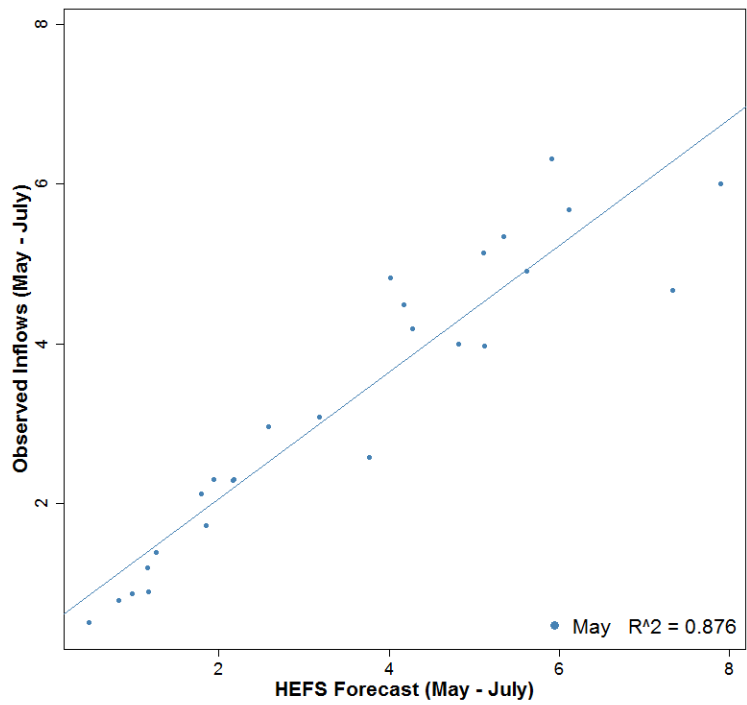
ESP Forecast vs. Observed Lamb Inflow (June - July)



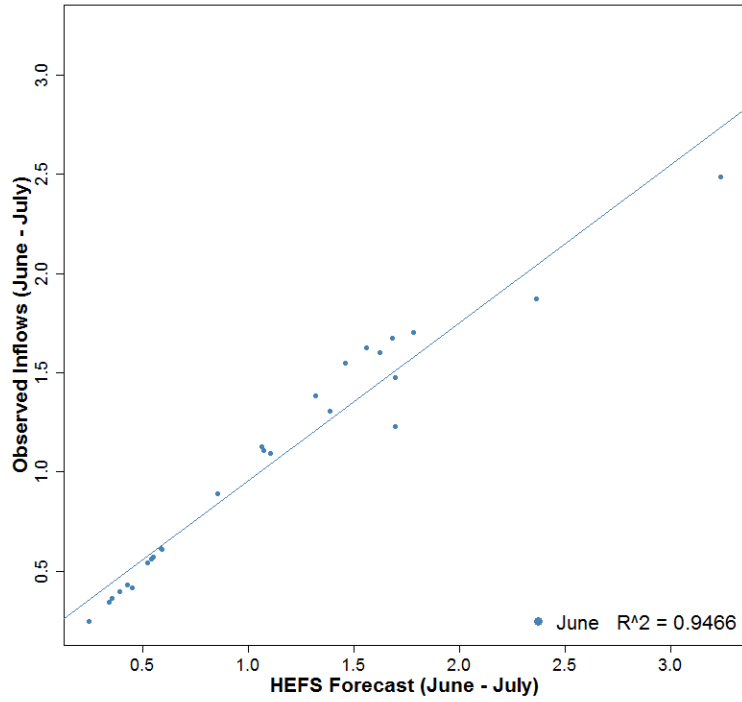
HEFS Forecast vs. Observed Little Dell Inflow (April - July)



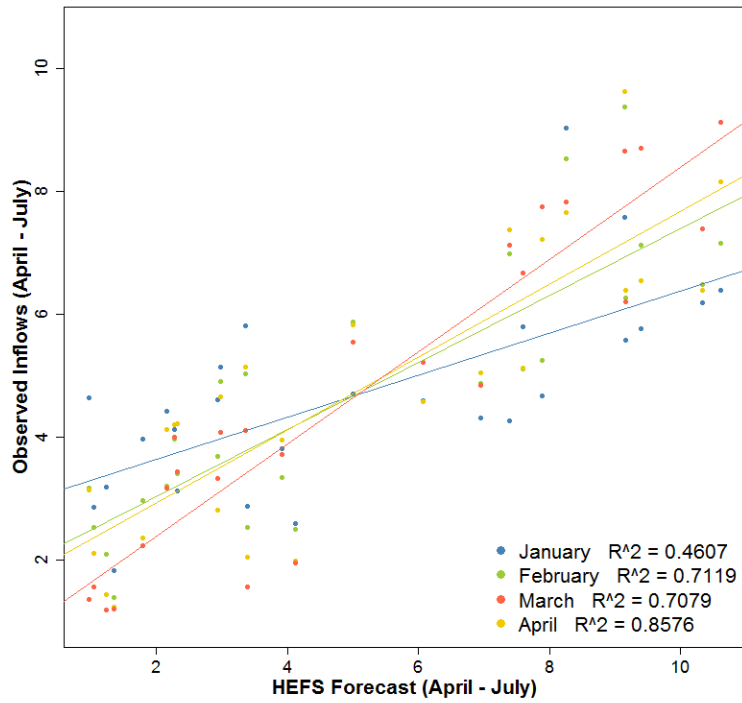
HEFS Forecast vs. Observed Little Dell Inflow (May - July)



HEFS Forecast vs. Observed Little Dell Inflow (June - July)



HEFS Forecast vs. Observed Lamb Inflow (April - July)



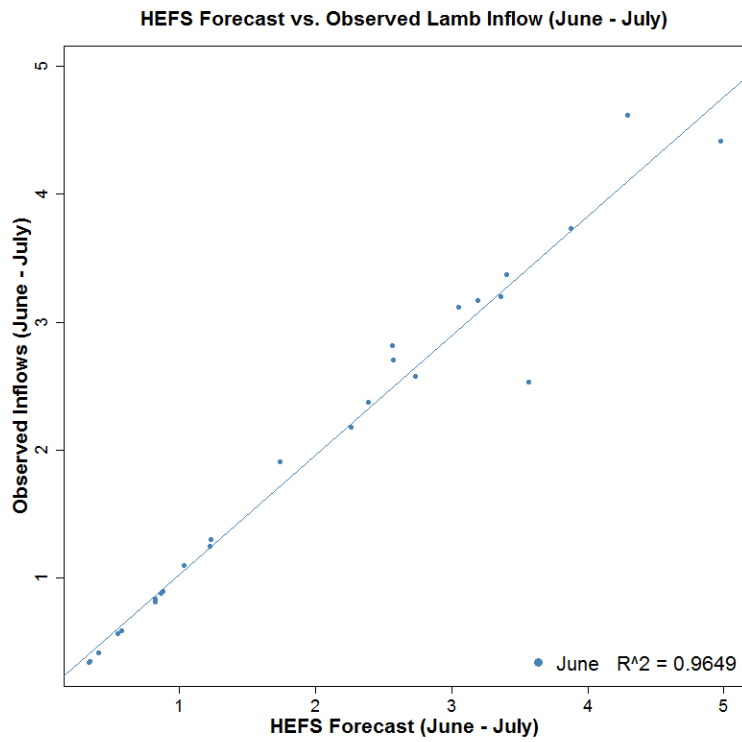
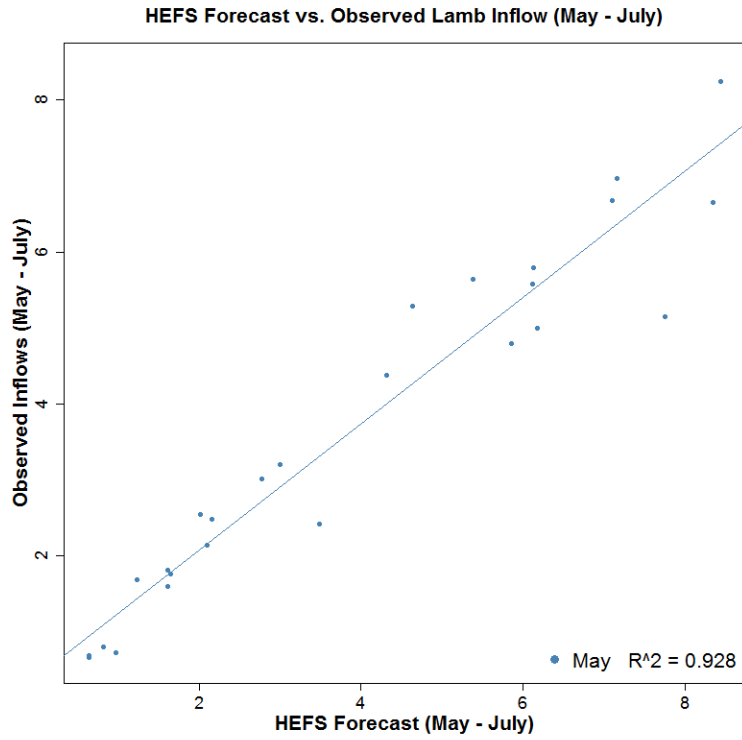


Figure A1: Regressions making up R-Squared Table

APPENDIX B

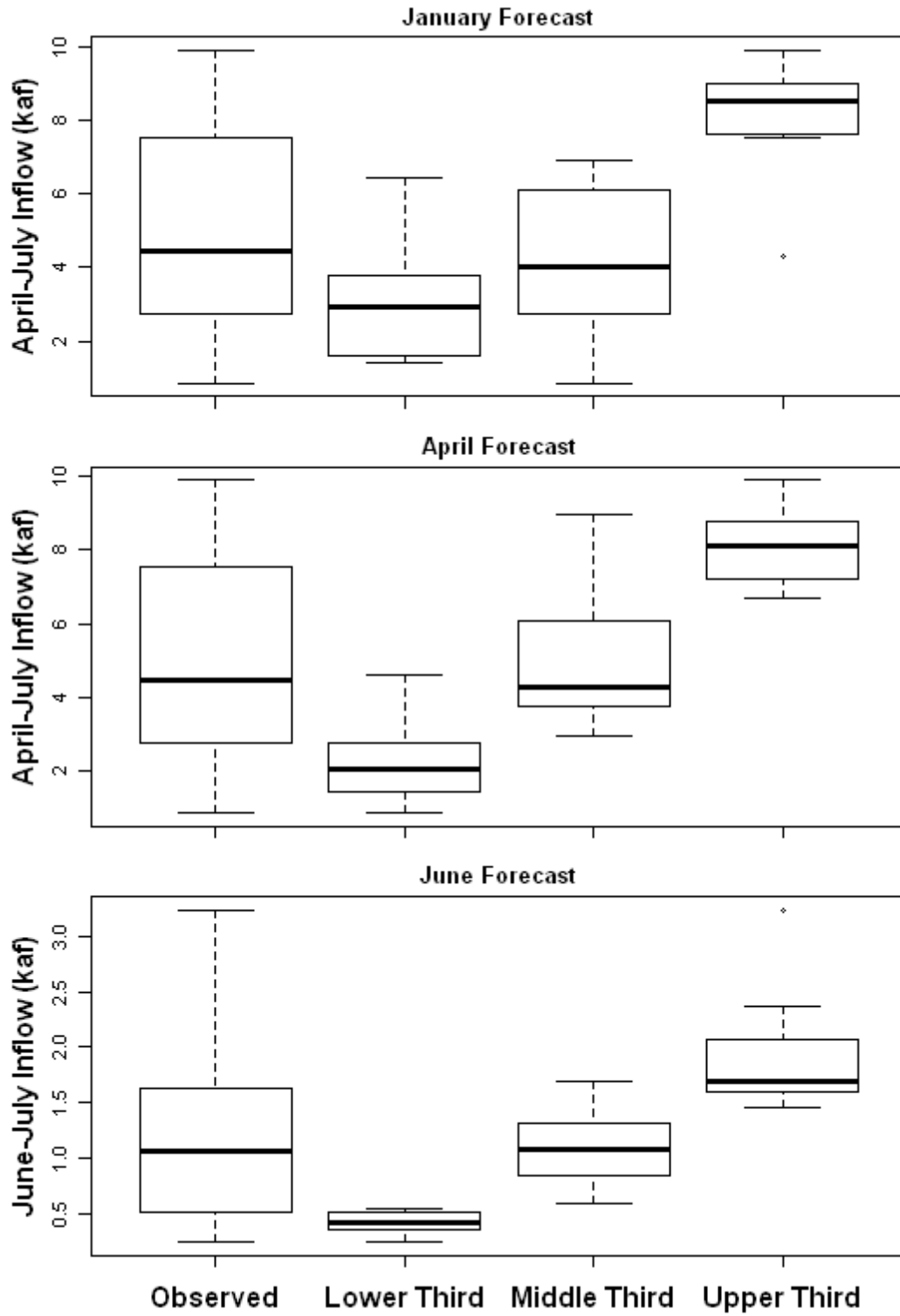


Figure B1: Annual April-July Inflow Volume – Little Dell HEFS (1985-2010)

APPENDIX C

Section C1: Baseline Simulation Rules

If the storage of the reservoir started out very low, before 10,000 acre ft, then the culinary pipeline experienced severe restrictions and only could get a 5 cfs release. The flood releases are 0 because at that stage, flooding would not be a concern.

```
if (combined.storage[t-1,(year-1980)] <= 10000) {  
  culinary.pipeline[t,(year-1980)] <- 5*86400/43560  
  parleys.pipeline[t,(year-1980)] <- min(max(0,(combined.storage[t-1,(year-1980)]-  
    culinary.pipeline[t,(year-1980)]-rule.curve[t]),200*86400/43560))}
```

If the storage is below 14,000 acre feet, flooding could be again is not likely a concern so bypass releases should be minimized. If the rule curve is below 14,000 acre feet and model storage is also, it should not be penalized for following the rule. So in that scenario the full water demand should be met. However, if storage is lower than 14,000 af and the rule curve is not, there could be an issue with low inflows in which case releases for water demand should be somewhat reduced, in this case half of the demand is being delivered.

```
else if (combined.storage[t-1,(year-1980)] <= 14000 & rule.curve[t] <= 14000) {  
  culinary.pipeline[t,(year-1980)] <- ideal.water.demands[t,1]  
  parleys.pipeline[t,(year-1980)] <- min(max(0,(combined.storage[t-1,(year-1980)]-  
    culinary.pipeline[t,(year-1980)]-rule.curve[t]),200*86400/43560))}
```

```
else if (combined.storage[t-1,(year-1980)] <= 14000){  
  culinary.pipeline[t,(year-1980)] <- ideal.water.demands[t,1]/2  
  parleys.pipeline[t,(year-1980)] <- min(max(0,(combined.storage[t-1,(year-1980)]-  
    culinary.pipeline[t,(year-1980)]-rule.curve[t]),200*86400/43560))}
```

If the storage is between 14,000 acre feet and 20,000 af this is the idea range to be in. Full water demands are met and bypass releases are made only when the storage exceeds the rule

curve. At these levels, bypass releases are capped at 200 cfs. This is the threshold under which flooding issues in town should not occur.

```
else if (combined.storage[t-1,(year-1980)] <= 20000) {  
  culinary.pipeline[t,(year-1980)] <- ideal.water.demands[t,1]  
  parleys.pipeline[t,(year-1980)] <- min(max(0,(combined.storage[t-1,(year-1980)]-  
culinary.pipeline[t,(year-1980)]-rule.curve[t]),200*86400/43560)
```

There is a high flood risk associated with the storage exceeding 20,000 af. In this case, as much water should be released as possible while still trying to reduce flooding. Full drinking water demands are met in this area. The bypass release is capped at 300 af.

```
} else if (combined.storage[t-1,(year-1980)] <= 21200) {  
  culinary.pipeline[t,(year-1980)] <- ideal.water.demands[t,1]  
  parleys.pipeline[t,(year-1980)] <- min((parleys.pipeline[t-1,(year-1980)] +  
20*86400/43560), 300*86400/43560)
```

The reservoirs combined capacity is 21,200 af. This level cannot be exceeded because there is no spillway. Exceeding this would cause major uncontrolled flooding downstream and damage to the dam. This scenario's releases are modeled as a full culinary release and the bypass release is not capped. As much water can continue to be released out of the parleys pipeline as needed to return to a proper volume. If bypass releases got extremely high and the storage volume was exceeding the capacity, further investigation would be required. This did not happen in the 29 year baseline run, so it was included on a basis that it would raise red flags if the simulation model was used in the future, outside of the baseline context.

```
} else {  
  culinary.pipeline[t,(year-1980)] <- ideal.water.demands[t,1]
```

```
parleys.pipeline[t,(year-1980)] <- (parleys.pipeline[t-1,(year-1980)] +  
20*86400/43560)}
```

The final storages were calculated in a very simple mass balance:

```
combined.storage[t,year]  
= combined.storage[t-1,year] + inflow[t] - culinary.pipeline[t,year]  
- parleys.pipeline[t,year]
```

Section B2: “Common-Sense” Simulation Rules

```
if (combined.storage.sim[month_new[month],(year-1980)] >= 20000) {  
  culinary.pipeline.sim[(month_new[month]+t),(year-1980)] <- ideal.water.demands[t]  
  parleys.pipeline.sim[(month_new[month]+t),(year-1980)] <- 200*86400/43560  
  
} else if (combined.storage.sim[month_new[month],(year-1980)] >= 21200) {  
  culinary.pipeline.sim[(month_new[month]+t),(year-1980)] <- ideal.water.demands[t]  
  parleys.pipeline.sim[(month_new[month]+t),(year-1980)] <-  
combined.storage.sim[(month_new[month]+t-1),(year-1980)] - ideal.water.demands[t] - 21299  
  
} else {  
parleys.pipeline.sim[(month_new[month]+t),(year-1980)] <- parleys.pipeline.avg[t]  
culinary.pipeline.sim[(month_new[month]+t),(year-1980)] <- culinary.pipeline.avg[t]  
}
```