POLITECNICO
MILANO 1863

SCHOOL OF CIVIL, ENVIRONMENTAL AND LAND MANAGEMENT ENGINEERING
MASTER OF SCIENCE IN ENVIRONMENTAL AND LAND PLANNING ENGINEERING

ASSESSING THE OPERATIONAL VALUE OF FORECAST INFORMATION UNDER CLIMATE CHANGE

Master Thesis by:
Henrique Moreno Dumont Goulart

Advisor:
Prof. Andrea Castelletti

Co-Advisor:
Prof. Matteo Giuliani
Prof. Jonathan Herman
Prof. Scott Steinschneider

Academic Year 2019

Acknowledgments

I would first like to thank my advisor Professor Andrea Castelletti, because he is the one that made all the work and development achieved in these last eleven months possible, in addition to the support and wise suggestions that improved the overall quality of the thesis.

I would also like to thank my co-advisors. Professor Matteo Giuliani has been fundamental to this work from the very first day in so many ways that it is difficult to demonstrate how grateful I am. His guidance, assistance and kindness shall not be forgotten. Professor Scott Steinschneider, who was kind to share his knowledge on synthetic forecasts and had lots of patience to teach me how to use his model. Finally, professor Jonathan Herman, who not only allowed me to stay with his research group in Davis for a couple of (intense) months, but also provided a great deal of instruction and information, besides being always available to the many times I found myself struggling. My most sincere thanks to all of you.

Special thanks to all the researchers and colleagues at the Natural Resources Management group at Politecnico di Milano and at the Herman Research group at UC Davis, it was a pleasure to cooperate with you.

The road here has not been easy, but with friends as good as mine everything is manageable. I am grateful to all of them, but especially to those who were somehow involved in this work, Ivan, Kaushal, Aaron, Gabriel, Nicolas, Alexandre, Kevin and Vinicius.

Finally, I would like to reserve a special place for my family, the foundation of it all. My parents and sister, for all the unconditional love, my grandmothers, whose care and zeal have been felt even from the other side of the ocean, my grandpa, who has had an active role in revising and improving my writing and Beatriz, who has stayed by my side even after the hardest obstacles. This has been a tough yet satisfying journey. Thank you all.

Abstract

As consequence of global warming, extreme events are expected to increase in magnitude and frequency at global and local scales. In the water resources field, on top of increasingly uncertain hydrological regimes, more recurrent and intense cases of flood and drought are projected, which causes reservoirs to lose efficiency and even to fail at their purpose of supplying water and preventing floods. Forecast has been proven to be a useful tool in improving water management, but little is known about its capability in future time periods under climate change. Therefore, it is the aim of this thesis to investigate the forecast value in the future and its contributions in mitigating the projected climate change impacts. For that, a study case at the Folsom reservoir is used, where 97 future climate scenarios are analyzed and selected according to their forecast potentiality. The selected scenarios are then simulated to establish the absolute and relative Expected Value of Perfect Information (EVPI). Given the nonexistence of data for future forecasts, a novel statistical framework is used to generate synthetic forecasts for future time periods whilst preserving the accuracy of the existing model currently used. When combined the selected future projections with the synthetic forecast framework, the future forecast ensembles for different climate change scenarios are created. In order to simulate these scenarios, a policy search framework is adopted, which applies an innovative method by controlling operation decisions with binary trees, defined by states and actions. Results indicate the use of forecast can improve water supply and prevent future floods from happening, in spite of an overall deterioration of the water operations in the future. The absolute value of forecast is projected to increase for all selected scenarios, while the relative forecast value has its evolution conditioned by the type of future scenario. In general, the relative value is expected to increase in wet scenarios but to decrease in dry scenarios. The thesis also found that forecast-based policies optimized over the past can improve the water supply levels over future time periods at the cost of increasing the flood risk. Moreover, given the concern about future uncertainty due to climate change, results show that forecasts allow for a wider range of future

scenarios to be contemplated by a single policy, granting more flexibility to the operation.

Riassunto

Come conseguenza del riscaldamento globale, si prevede che gli eventi estremi aumentino in ampiezza e frequenza su scala globale e locale. Nel campo delle risorse idriche, oltre a regimi idrologici sempre più incerti, vengono proiettati casi più ricorrenti e intensi di alluvione e siccità, che fanno perdere alle infrastrutture le infrastrutture responsabili di attività vitali, come fornitura d'acqua e la prevenzione delle inondazioni, fallire al loro scopo. Le previsioni hanno dimostrato di essere uno strumento utile per migliorare la gestione e il funzionamento delle risorse idriche, ma si sa poco sulla sua capacità in periodi di tempo futuri sotto il cambiamento climatico. Pertanto, lo scopo di questa tesi è studiare i contributi previsionali per mitigare gli impatti previsti sul cambiamento climatico nella gestione delle risorse idriche e valutarne il valore rispetto alle condizioni attuali. Per questo, vengono analizzate 97 proiezioni di afflusso per il serbatoio Folsom, basato su più modelli di circolazione generale (GCM) e percorso di concentrazione rappresentativo (RCP), quindi selezionate in base alla loro potenzialità di previsione. Gli scenari selezionati vengono simulati con una programmazione dinamica per stabilire politiche operative di base e perfette e vengono calcolati i valori potenziali assoluti e relativi. Data la inesistenza di dati per le previsioni future, viene utilizzato un quadro statistico per generare previsioni sintetiche per periodi di tempo futuri, preservando l'accuratezza del modello esistente a breve termine attualmente utilizzato nel serbatoio. Se combinate le proiezioni future selezionate con il quadro di previsione sintetico, vengono creati gli insiemi di previsioni future per diversi scenari di cambiamento climatico. Al fine di ottimizzare e simulare questi scenari, viene adottato un framework di ricerca delle politiche, che applica un metodo innovativo controllando le decisioni operative definiti da stati e azioni. I risultati indicano un generale deterioramento delle operazioni idriche, ma l'uso delle previsioni può migliorare l'approvvigionamento idrico e impedire che si verifichino inondazioni future per una serie di scenari futuri diversi rispetto a un'operazione di base. Si prevede che il valore assoluto della previsione aumenti con il tempo per tutti gli scenari selezionati in conseguenza dei maggiori costi delle oper-

azioni in futuro, mentre il valore relativo ha la sua evoluzione condizionata dal tipo di scenario futuro. In generale, il valore relativo dovrebbe aumentare negli scenari bagnati ma diminuire negli scenari asciutti. La tesi ha anche scoperto che una vecchia politica operativa supportata dalle previsioni può migliorare le prestazioni fino a un livello simile di una futura politica ottimizzata senza previsioni, aumentandone la durata. Inoltre, data la preoccupazione per l'incertezza futura dovuta ai cambiamenti climatici, i risultati mostrano che le previsioni consentono di contemplare una gamma più ampia di scenari futuri da un'unica politica, garantendo una maggiore flessibilità all'operazione.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The context

The planet earth has been experiencing a constant increase in its temperature along the past decades, both on land and on the ocean surface. It is estimated that an average warming of 0.85 °C has taken place since 1880 to present-day (*IPCC*, 2018). The increase in temperature has had reflections on acidification of the oceans, loss of ice sheets on both poles, rise of mean sea level, among other consequences (*IPCC*, 2014).

There are no records of any similar event to the present one and there is a plethora of evidences that relates global warming to an abrupt increase of greenhouse gas (GHG) concentration levels, which in turn have their roots at human-related activities, especially from 1750 onwards, with half of the cumulative anthropogenic CO2 emissions taking place in the last 40 years (*IPCC*, 2014). Emissions come mainly from electricity and heat production, industry, transport and agriculture, forestry and other land use. In addition, two factors stand out as the main drivers of CO2 emissions: human growth and economic growth, the latter presenting an ever-increasing contribution (*IPCC*, 2014).

Changes observed in temperature and climate carry consequences into the environment. Different species have altered their natural behaviour as a reaction to climate change, while hydrological systems are being affected due to snow melting, oceans are suffering from acidification, and certain crops have

had their yields negatively disturbed (*IPCC*, 2014). Moreover, it is expected that water scarcity and stress cases will grow in many regions around the world. The tendency is to observe increased occurrences of floods in certain regions, as a result of more frequent extreme precipitation events and to observe harsher droughts for longer periods (multi-decade) in other regions (*IPCC*, 2014; *Damania et al.*, 2017).

On top of that, the global population is expected to keep growing during this century. In spite of the variability of growth projections, it is predicted that a net value of one billion people will be added to the world by 2030, which makes for a total of 8.6 billion people. Furthermore, by the end of the century the total population should be around 11.2 billion people, increasing by 3.6 billion with respect to the current population (*United Nations*, 2017). While developed countries are growing at slower rates, especially in Europe, the main population growth will take place in countries in Africa and South Asia. At the same time, a different process is taking place, the rural flight. Migrations from rural areas to urban areas all over the world will add 2.4 billion people to cities by 2050 (*FAO*, 2017). This increase in human population implies naturally a greater demand for food and the expansion of cities translates into a bigger demand for sanitation and energy, all of which will accentuate even further the stress on water resources, given their role in agriculture, sanitation and power generation (*FAO*, 2017). Thus, the demand of energy keeps rising, concurrently with a need for decarbonising its sources, which contributes to the expansion of clean and renewable sources. Hydropower is a key supply in the energy production, being the major source of renewable energy, corresponding to 58% of the total global renewable energy generation (*IRENA*, 2018).

In this context, the importance of water-control structures, such as water reservoirs is growing (*Zarfl et al.*, 2014; *Culley et al.*, 2016; *Ehsani et al.*, 2017). However, simply expanding the infrastructure capacity of dams might prove costly and not-necessarily effective, since the infrastructure investment is irreversible and might not meet the future water supply needs. This is due to the amplified uncertainty of future projections, which prevents an accurate comprehension of the climate conditions in the long-term future (*Fletcher et al.*, 2019). Thus, as an alternative to the construction of water reservoirs, it is possible to act on the operation of the reservoirs, representing a low-cost and flexible solution that can improve the resilience of the system against the increased variability of future conditions (*Castelletti et al.*, 2008; *Ramos et al.*, 2013; *Giuliani et al.*, 2016).

Reservoir operations are traditionally managed based on historical records, such as precipitation, temperature, day of the year and daily demand. However, because of climate change, the patterns for hydrometeorological states change in multiple and unprecedented forms. Hence, the traditional approach loses reliability and performance, since it cannot effectively control the releases of water in a scenario where the hydrological conditions behave differently from what was historically defined and is more prone to system failures (*Georgakakos et al.*, 2012; *Turner et al.*, 2017). Because of that, an adaptive management strategy is needed. This approach consists in constantly receiving new data, either observations from a physical state or future predictions based on models, related to the hydrological conditions at the moment of the operation and using it for informing operational decisions. Studies (*Georgakakos et al.*, 2012; *Culley et al.*, 2016; *Giuliani et al.*, 2016) showed that different concepts of the adaptive method did improve the operation with respect to the traditional method. Among them, forecast systems excels as one of the most studied tools in water operations. The forecast value has been verified in different occasions and circumstances, be it for flood prevention using short-term models (*Adamowski*, 2008; *Cloke and Pappenberger*, 2009), water supply supported by long-term forecast information (*Hamlet and Lettenmaier*, 1999; *Mahanama et al.*, 2011; *Yuan et al.*, 2015) or for different types of reservoirs (*Turner et al.*, 2017; *Anghileri et al.*, 2016). Its flexibility allows for large ensembles of physical forecasts to be coupled for more accuracy (*Pappenberger et al.*, 2008) or to rely on statistics to generate the predictions (*Toth et al.*, 2000; *Block and Rajagopalan*, 2007). Building on these mentioned studies, the main interest of this work concerns moving the assessment of forecast from the present to the future time period, in which the uncertainty is intensified and extreme events are more likely to occur. Particularly, to understand if a forecast system is able to keep the properties and contributions seen in the current time and, furthermore, if it is capable of mitigating the negative impacts of climate change by reducing the occurrences of system failures and extending the lifespan of the structure.

1.2 Objectives of the thesis

As previously explained, even though there are plenty of studies supporting the benefits of using forecasts in water-control operations (*Georgakakos et al.*, 2012; *Denaro et al.*, 2017; *Turner et al.*, 2017; *Nayak et al.*, 2018; *Herman and Giuliani*, 2018), they usually focus on the use of forecast for past or current scenarios. In such cases, in spite of the benefits already being significant to the improvement

of regular operations, little can be learned or applied to the incoming future, where projections heavily differ from historical records and the states of the climate variables are harder to predict.

This study has the primary goal of exploring the forecast value in future conditions affected by climate change and to assess in which ways it differs from the current contributions estimated over historical conditions. Specifically, the research questions that guide this work are:

1) How water systems operations and their performance will be impacted by climate change?

2) Can forecast contribute to design adaptive operating policies?

3) Is the forecast value expected to increase under future and more variable hydroclimatic regimes?

4) How sensitive is the forecast value across an ensemble of uncertain scenarios?

5) Can an adaptive policy over a historical conditional provide satisfactory results when evaluated in the future? Is the addition of a forecast enough to extend its usefulness?

The case study of the Folsom reservoir, near the Sacramento city in California (United States of America) is used to test the proposed methodology. California is a state that has historically faced water-related problems, such as extreme shortages of water that can exceed the duration of years and severe floods affecting its many cities and inhabitants. Folsom reservoir is an artificial reservoir constructed in 1955 and located on the American River. Nowadays, the reservoir serves multiple objectives, but its main tasks are still flood control and water supply operations.

In order to achieve the ultimate goal of evaluating the benefits of forecast in a future scenario affected by climate change, some intermediate and specific objectives need to be set as well, which are:

1) Analyzing the projected climate and understanding the main trends in terms of variability and predictability, frequency and duration of extreme events.

2) Quantifying the maximum space of improvement of the historical operations with respect to an ideal solution designed assuming a perfect (deterministic) knowledge of the future, also known as Expected Value of Perfect Information (EVPI; *Giuliani et al.* (2015)).

3) Generation of a synthetic forecast ensemble for projected reservoir inflow based on the work of *Nayak et al.* (2018). The synthetic generation allows for a reliable simulation of the existing forecast models over projected inflow scenarios.

4) Quantifying the actual forecast value as the performance gain generated by informing the Folsom reservoir operations with the synthetic forecast over the historical and future scenarios.

1.3 Outline of the thesis

Chapter 2 describes the state-of-the-art on the topic of metereological and hydrological forecasting in water systems. It starts with the definition of value of exogeneous information for water systems operations. It then describes the types of existing forecasts, discussing skill, lead time, uncertainty, and the relation between forecast skill and forecast value. A brief explanation of reforecast, hindcast and synthetic forecast is also provided.

Chapter 3 provides a comprehensive description of the Folsom reservoir case study. Firstly, the morphologic and hydro-climatic features of the California river basin are presented. Then, a summary of the stakeholders and the main interests involved in the system is provided. Finally, the data utilized in the study are described.

Chapter 4 illustrates the proposed methodology, which is composed of the following three main blocks: 1) Scenarios selection based on a statistical analysis of the climate projections coupled with the quantification of EVPI; 2) Synthetic generation of inflow forecast for the selected scenarios; 3) Assessment of historical and future forecast value.

Chapter 5 reports the numerical results obtained in this study. According to the proposed 3-block framework, it first discusses the scenairo selection based on the projected climate statistics and the values of EVPI; the synthetic forecast generated for the selected scenarios are then reported; finally, it analyzes the

historical and projected forecast value.

Chapter 6 sums up the conclusions and suggests some starting points for further research about the topic.

2

State-of-the-art

2.1 Value of observation

Today there are many technologies, tools and methods that improve the management and operation of water resource systems. In a world with a surfeit of data available, one of the most common practices is to harness real-time data and observations, which can be achieved by using a variety of equipment, such as radars, sensors, satellites or even touristic webcams (*Giuliani et al.*, 2016), in order to gather additional useful information that allows for an enhanced operation. For water resources operations, relevant observable information can be for instance precipitation, upstream flows or snow pack depth (*Denaro et al.*, 2017). Similarly, *Giuliani et al.* (2016) developed a low-cost, real-time and observation-based method for extracting snow-related information from pictures on the web taken from users and webcams. After an automatic process where the pictures are localized, processed and analysed, a virtual snow index is computed to represent the snow-covered area. This index is embedded into the system operations and it produced an improvement of 10% in relation to the baseline operations (*Giuliani et al.*, 2016).

Even though direct observations cannot foresee future trends of inflow with great accuracy, the potential for directly using hydro-meteorological data is becoming relevant as indicated by *Denaro et al.* (2017). Their study showed that by knowing solely the accumulated snow state on the mountains during winter, it was possible to anticipate the amount of water to be melted during spring,

ultimately permitting a larger amount of water to be stored for providing reliable irrigation over the summer. The higher water levels of the lake led to an improvement of the water supply index of 10% in relation to the baseline scenario, in spite of an increased risk of flooding. In another study concerning flood events in the north of Italy, *Castelletti et al.* (2008), added to a traditionally-designed policy real-time information, e.g., precipitation and temperature measures. While the results presented modest improvements with respect to the regular policy performance during the flood season, this solution largely improves the system performance during the rest of the year.

However, even though better informing the operations of water systems has its benefits, it also carries some drawbacks, such as observational error and estimation biases, possibly negatively affecting the system performance. To support the selection of the most useful information for improving the system operations, *Giuliani et al.* (2015) developed the information selection and assessment (ISA) framework. Figure 2.1 illustrates the process and it works as following: first the expected value of perfect information is determined as the difference between an ideal, perfect solution designed under perfect knowledge of the future and a baseline solution relying on a basic set of information. Then, multiple candidate information are processed by an input variable selection algorithm to automatically identify the ones that are expected to be more valuable for informing operational decisions. Lastly, the selected variables are used in designing informed solutions. The operational value of this information is quantified as the the performance gain of the informed solutions with respect to the baseline.

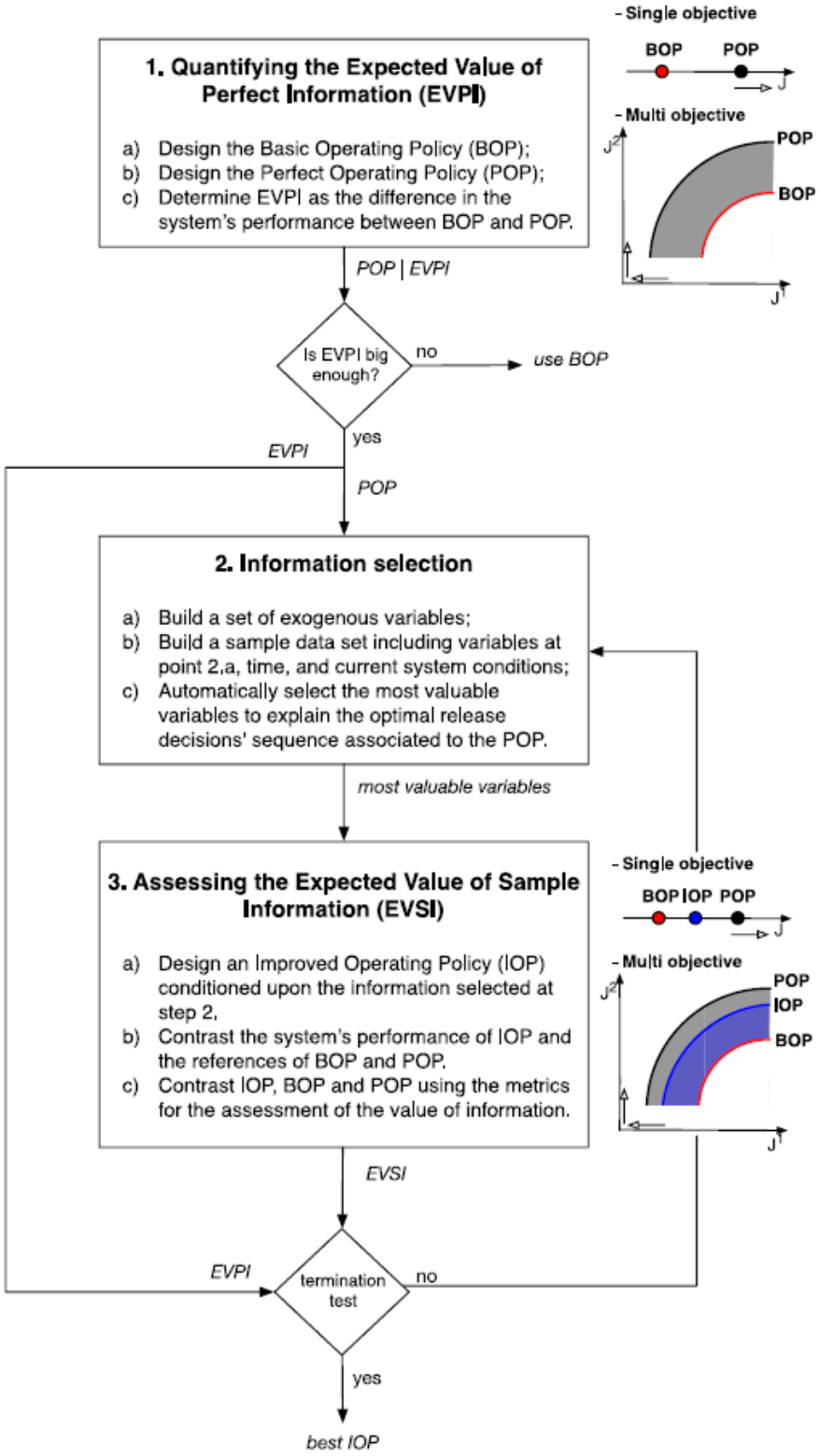

Figure 2.1: *Description of the steps of the Information Selection and Assessment (ISA) framework. Figure by Giuliani et al. (2015).*

Whilst the original intended use of the ISA framework was to identify and select the most valuable information from a set of observational data, it can also be exploited for forecasts identification and selection. The process and steps would remain the same, but the candidates in the second step of the framework would instead be different forecast models, different forecast variables, or different forecast lead times.

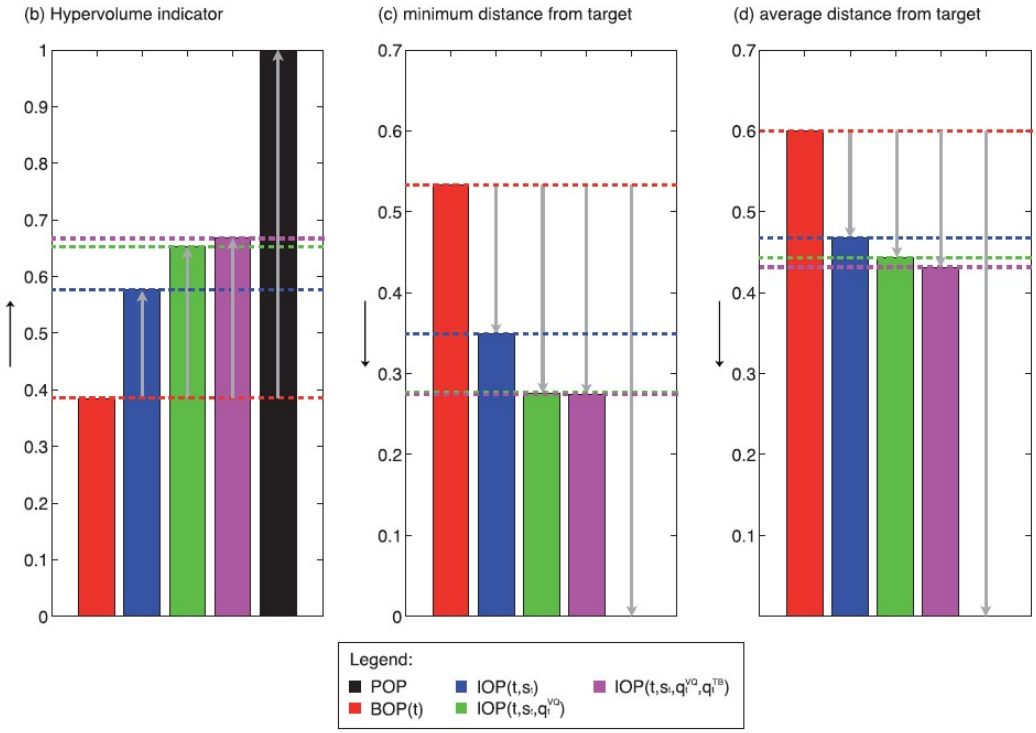

Figure 2.2: *Contribution of additional information to the overall performance of the operation. Figure by (Giuliani, Pianosi and Castelletti, 2015).*

2.2 Forecasting

2.2.1 Forecasting in water resources operations

Although it is possible to predict future conditions, such as flood events, only by relying on weather observations, there are situations in which a forecast model is necessary. It might happen that the equipment fails or it might be that the basin is located in an area where the conditions are hard to be measured (*Toth et al.*, 2000). Furthermore, studies have shown that with the help of forecast models, water supply failures were reduced and the life-spam of hydraulic infrastructure was increased, all at a relatively low cost, when compared to new

investments in new infrastructure (*Turner et al.*, 2017).

Along the past fifty years, there was a considerable improvement in weather forecasting, with great advances in lead times, forecast skill and value. This progress is directly linked to the considerable increase in computing power, instrumentation and understanding of the atmospheric dynamics (*Lynch*, 2008). Plenty of researches have been published analysing the multiples contributions of forecasting to the water resources operations, including different objectives like flood control, water supply, hydropower generation and different combinations of them together. Some of these studies are recalled and briefly explained in this section in order to provide evidences of the importance of forecasting in this domain.

A study by *Faber and Stedinger* (2001) presented the implementation of forecast into a Sampling Stochastic Dynamic Programming (SSDP) model and how it performed in comparison to a standard operation. In fact, two alternatives were developed: A SSDP model calibrated on the historical conditions, but incorporating current hydrologic information forecasts to update the probabilities and a SSDP model using ensemble streamflow prediction (ESP) updated on a weekly basis. Both presented a superior performance in relation to the baseline case, the same system and model but without the forecast components. Plus, some differences in the output of the two forecast models were found, with the one using the ESP outperforming the solution based on historical conditions but updated with current hydrologic information forecasts (*Faber and Stedinger*, 2001).

Anghileri et al. (2016) recently studied the value of long-term forecasts in the design of water supply operations, by using a forecast-based adaptive control framework. The first results showed that season-long ESP forecasts improved operations with respect to the baseline, remaining 35% below the perfect forecast value. The system considered in this study (i.e., the Oreville reservoir, California) allows storing water from one year to the other and would have here gained much more from forecast over longer lead-time than the seasonal one. In fact, the inter-annual component of the forecast contributed from 30 to 45% of the total forecast value. Moreover, it was possible to analyse in which circumstances each type of forecast generated most value with respect to the baseline (no forecast): for high demands with respect to the storage capacity or inter-annual carryover, only a perfect forecast would be able to properly anticipate the streamflow; medium-high demands but a small storage was suffi-

cient for the seasonal ESP forecast to perform and finally, inter-annual forecasts worked best for big storages, in spite of not being useful for small reservoirs (*Anghileri et al.*, 2016).

It is also worth noticing that the gains to the water operations provided by forecasts depend also on the final objective of the water reservoir, as shown by *Turner et al.* (2017). For flood-prevention operations, for instance, increasing the quality of forecasts directly implies in an improvement to the system's performance. However, for water supply purposes, the correlation is not as clear, because some reservoirs are capable of buffering the variability of the inflow. In addition, smaller reservoirs, which have less buffering capacity, are the ones that would profit the most of long-term forecast (*Turner et al.*, 2017).

Another example demonstrating the value of forecast is provided by *Nayak et al.* (2018) for the operations of Folsom Reservoir (California), in which it was possible to considerably improve the reliability of water supply without increasing the risk of floods due to the prediction of major flood events by a forecast model. However, the study found that forecasts with a skill of thirty days did not present significant improvements in relation to forecasts with a three-day skill for the river basin studied. Finally, it was suggested that forecasts can also bring advantages to conjunctive use, such as groundwater storage in water reservoirs, because of the longer times required for water banking and withdrawing from underground storages (*Nayak et al.*, 2018).

2.2.2 Forecast lead time

Forecasts are usually divided in two categories: short-term and long-term or seasonal. The first ranges from one day to five days, rarely exceeding one week, while the second ranges from months to multiple years. Each category has its own objectives, because it is still a challenge to integrate the different timeframes in a single model (*Yuan et al.*, 2015). The dynamics in each case change substantially and so do the main factors contributing to the forecast skill. In short-term predictions, the forecast horizon, which quantifies the number of days in advance that can be foreseen, is the most important factor. Generally, the longer the forecast horizon the better is the performance of the system. However, for long-term forecasts, the forecast uncertainty plays the main role, as the forecast can become too uncertain at long lead times. At medium forecast horizon, the performance of the system is equally affected by both factors (*Zhao et al.*, 2012).

Short-term forecasts are mostly aimed at alerting flood occurrences and at mitigating the impacts related to them. The few days of anticipation allow for releasing water from reservoirs, so the flood can be partially diminished, contributing to a reduction of the damage to people, buildings, agriculture, and infrastructure (*Pappenberger et al.*, 2008; *Elsafi*, 2014). Short-term forecasts can be based on direct observational data through the use of radars and gauges to capture the precipitation state or on more advanced models, such as numerical weather prediction and ensemble prediction systems that ensures accuracy, quickness, reliability and robustness of the forecast system (*Adamowski*, 2008; *Cloke and Pappenberger*, 2009).

On the other hand, long-term forecasts address concerns that are not as immediate or urgent as the short-term, but still of great importance to the environment and society, such as water supply and power generation. For these, it is important that the forecast can predict the future states at longer time leads, such as months, seasons, and years. Long-term forecast mostly rely on slow climate dynamics and, consequentially, require different weather information to be acquired. Among them, snow state information and soil moisture data have been proved to be useful at extending the forecast skill to longer lead times for specific regions. As an example, a study by *Mahanama et al.* (2011) found that in some regions of the United States of America, the contributions of soil moisture and snowpack initialization allowed for skillful forecasts with 95% confidence level of up to 9 months of lead time during certain seasons (*Mahanama et al.*, 2011).

Nonetheless, such correlations are limited to specific regions and seasons. At the global scale the understanding of ocean-atmosphere teleconnections is paramount and nowadays, most long-term forecasts rely on understanding teleconnection patterns (*Yuan et al.*, 2015). The term describes a naturally-occurring climate variation or alteration, such as the sea surface temperature (SST) fluctuation of the oceans, that affects different regions globally, usually at long distances and for multi-year periods of time (*Nigam and Baxter*, 2014). Among teleconnection patterns, the El Niño Southern Oscillation (ENSO) is one of the major ones, affecting the whole planet in different ways. It involves sea surface temperature (SST) oscillation along the years, with a warm-phase called el Niño and a cold-phase called La Niña, as seen in Figure 2.3. The return frequency varies from 4 to 6 years and lasts for 2 years. Forecasts based on ENSO were shown to keep the skill from 6 to 9 months of lead time for some regions along the coastline of the Pacific Ocean (*Hamlet and Lettenmaier*, 1999).

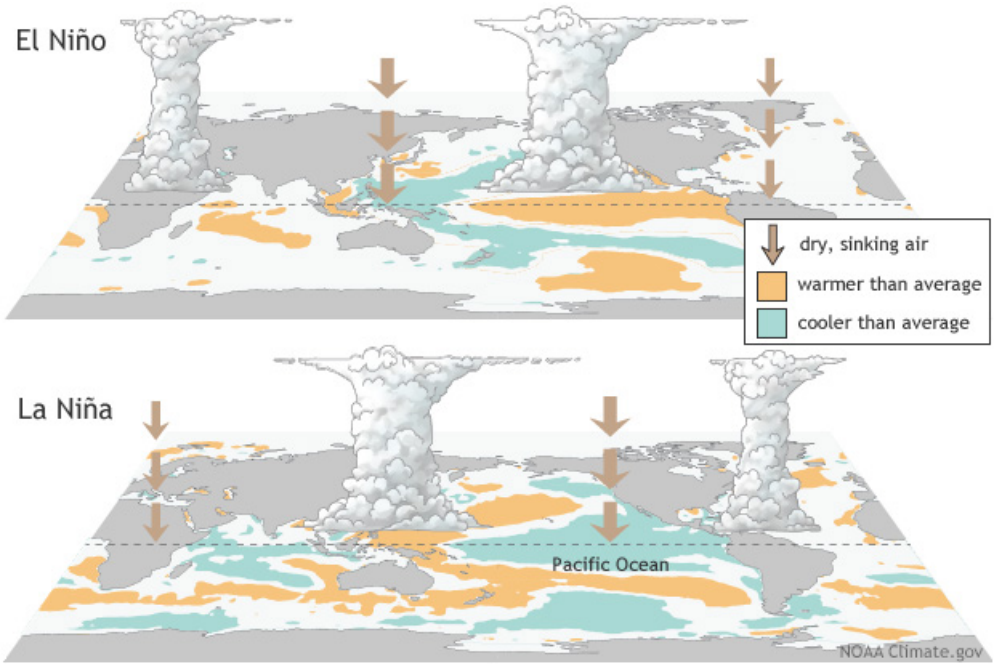

Figure 2.3: *Representation of both El Niño and La Niña, figure by National Oceanic and Atmospheric Administration.*

Yet, some regions are not substantially affected by the ENSO, as is the case of the European continent. During the colder months of the Northern Hemisphere, another phenomenon becomes predominant in the atmospheric variability, which is known as North Atlantic Oscillation (NAO). It is a large-scale reorganization of the air masses located in the Arctic and subtropical Atlantic. Every movement alters the average speed, direction, temperature and moisture of the air currents in the region and even storms are regulated by the NAO. The phenomenon has two phases, a positive one and a negative one and the consequences can be observed for instance in fish population, agricultural harvests, water management and energy supply (*Hurrell et al.*, 2003). Furthermore, when basing forecasts on NAO, it is possible to predict winter discharges with relatively high skill in some regions of Europe and also subsequent summer discharges, but at lower skill (*Bierkens and van Beek*, 2009).

One of the first researches studying teleconnections at water resources operations was the work by *Hamlet and Lettenmaier* (1999), which aimed at exploring the potentiality of long lead climate forecasts to improve river operations in the United States of America. By coupling the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO), to the streamflow forecast ensemble, it was verified that a 6-month forecast was achieved for the Columbia River.

Furthermore, the study showed that this addition of forecast would bring benefits to hydropower generation, water supply for agriculture activities, fisheries management and flood control (*Hamlet and Lettenmaier*, 1999).

2.2.3 Types of forecast models

A model is a simplified copy of a real system that has as purpose to mimic it, in order to represent as realistically as possible the experiments conducted and to generate reliable results, which can then be used as base for other experiments, tests and researches (*Soncini-Sessa*, 2007). There are many types of models, which are duly described in the study by *Jajarmizadeh et al.* (2012). In the realm of hydrological forecasting, however, there are two main approaches, the first relying on physically-based models and the second on data-driven models.

Physically-based models, also known as dynamic, process-based climate models aim at reproducing specific climate patterns and dynamics, as atmospheric pressure, temperature, precipitation, so a future state can be foreseen. A classical example is the numerical weather prediction (NWP), one of the most traditional models for the forecasting of climate states, such as precipitation. The outcome of the NWP can then be used as input for hydrological models, which will allow ultimately to predict a river flood event. While the dynamic models usually lead to better responses of a water system, its uncertainty is still considerable, with high chances of missing flood events or providing false alarms, especially at longer lead times (*Zhao et al.*, 2012). A way to solve the vulnerability is to conceive an ensemble of forecasts or an ensemble prediction system (EPS), in which multiple predictions of possible states are put together, providing different values of forecasting and thus giving a broader idea of the upcoming climate state, as can be seen in Figure 2.4 which will serve as information for a hydrological model. Studies showed EPSs contribute to the value of forecasting at medium-range, presenting a better performance at identifying floods events for longer lead times (*Cloke and Pappenberger*, 2009; *Pappenberger et al.*, 2008). Albeit, while EPS improved results, relying on a single Ensemble Prediction System might still offer some vulnerabilities motivating the use of multiple EPS together, also known as grand-ensemble. *Pappenberger et al.* (2008) used the THORPEX Interactive Grand Global Ensemble (TIGGE), which is composed of seven ensembles, to reduce the vulnerabilities of the system and attained an improved performance with respect to a single EPS, forecasting up to 8 days in advance flood events (*Pappenberger et al.*, 2008).

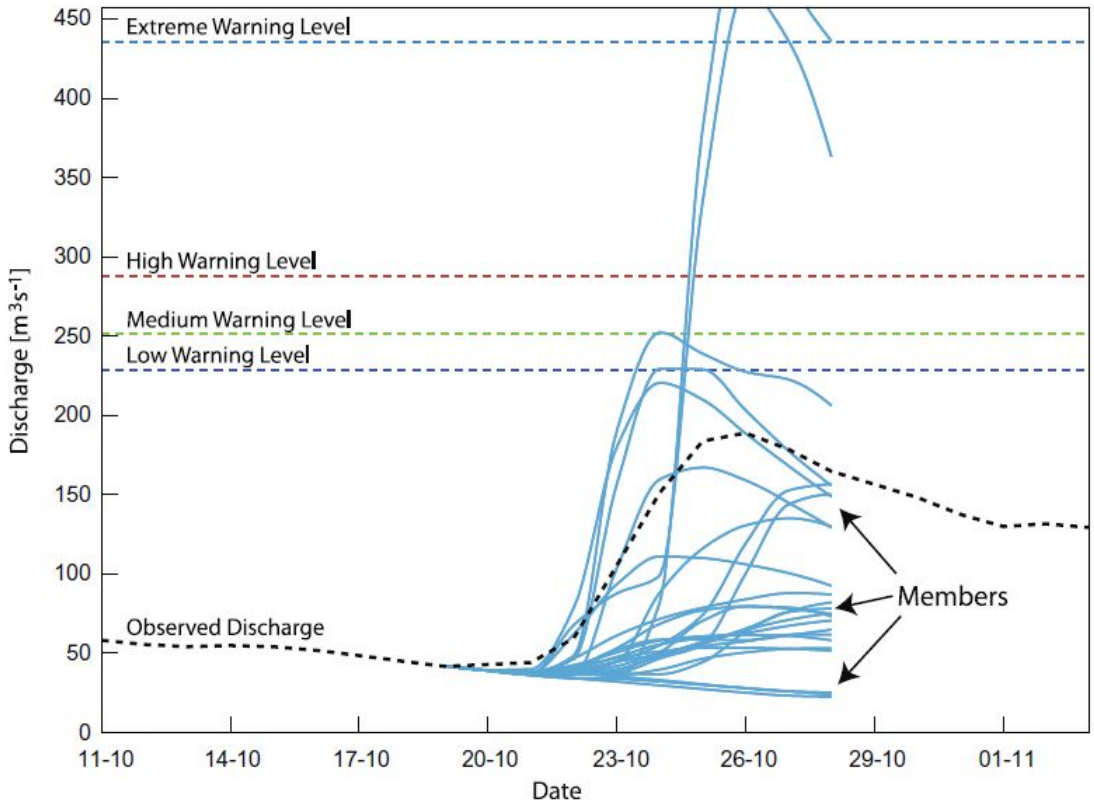

Figure 2.4: *Example of an ensemble prediction system, where different forecasts are taken together and a better understanding of the future scenario can be obtained. Figure by Cloke and Pappenberger (2009).*

The concept of ensembles can be applied as well for forecasting river inflows, which is called ensemble streamflow prediction (ESP), developed by the National weather service (NWS) since 1970 (*Yuan et al.*, 2015). It consists in combining the physical-based model of a river at the given moment, where current conditions are taken into calculations, with the climate data from many historical years, culminating into a set of patterns or traces that could possibly take place, each one based on a different year (*Faber and Stedinger*, 2001). Later, with the increase of computing capacity and better understanding of globally occurring phenomena, such as teleconnections, the general circulation models (CGCMs) were conceived, describing globally the behaviour of atmosphere, ocean and land in an integrated way. Alongside the traditional ESP, this allowed for the creation of the climate-model-based seasonal hydrologic forecasting (CM-SHF). It benefits from the ensembles of trajectories of the ESP for precipitation and temperature, but at longer lead times, because of the higher capacity of the CGCM at predicting season-long climate states (*Yuan*

et al., 2015).

The second class of forecast models are the data-driven models, also known as black box models. The idea behind data-driven models is to determine correlations between data, by applying different mathematics operations, without necessarily explaining the physical and chemical processes involved. The most common and basic type is the linear regression model, in which different data are coupled and correlated in a linear way. Among the drawbacks of linear regression, it is worth mentioning the impact of outliers and the inability to capture non-linear correlations between the data (*Block and Rajagopalan*, 2007). Another data-driven model commonly used is the autoregressive moving average (ARMA) model, which establishes a linear relation between different timesteps of the same data, predicting future values based on past observations (*Toth et al.*, 2000).

An alternative proposed by *Block and Rajagopalan* (2007) was to utilise a local polynomial-based nonparametric approach, overcoming some of the limitations of linear-regression models, being more flexible and able to capture other trends. The method allowed for the selection of the nearest neighbours of a specific point of interest, then locally parametrizing the data with a polynomial function and finally adding random deviates with the same standard deviation of the polynomial to generate an ensemble. The advantages of this method are the minimization of outliers' disturbance, detection of local correlations and elimination of multicollinearity, when one variable is correlated to multiple other variables (*Block and Rajagopalan*, 2007).

In addition, two other data-driven models deserve to be briefly discussed. The K-Nearest-Neighbour Method (KNN) explores the closest data to a specific observation in a data sample and establishes the correlation between them. It is a non-parametric approach and does not establish any structural relationship between the data sets (*Toth et al.*, 2000). The second model, Artificial Neural Network (ANN), is a computer intelligence trained at recognizing relationships between inputs and outputs. This method does not require any a priori hypothesis and works with any sort of relationship between the data sets, including non-linear and complex ones. Many studies showed ANN overcoming the more basic methods such as ARMA and linear regression (*Toth et al.*, 2000; *Adamowski*, 2008; *Elsafi*, 2014).

In spite of the evolution of both physical and data-driven models along the

past decades, there is still no clear evidence of outperformance of one model over the other, with each one presenting different benefits and disadvantages to the other. Physical models are precise, give better understanding of the natural processes involved and can simulate a variety of flow situations, but require more data to be set, which is not always possible and cannot take exogeneous data into account (*Elsafi*, 2014). Data-driven models are usually quicker, simpler and easier to operate, but are restricted to stationary data, which makes it particularly limiting for the current climate change context (*Adamowski*, 2008).

2.2.4 Forecast uncertainty

Meteorological and hydrological forecasting can contribute to the improvement of water systems in different ways, such as alerting for floods and preparing for droughts, by anticipating the upcoming conditions and extreme events. However, forecasts cannot provide flawless predictions and there is a need to assess and measure the uncertainty of the predictions, fully understanding the uncertainty of a forecast can actually further improve its value and reduce mislead decisions (*Ramos et al.*, 2013).

Better understanding of forecast uncertainty means incorporating additional information into the expected forecast prediction. Instead of only providing what the future value is expected to be, the uncertainty allows communicating the confidence interval of such value. Results by *Ramos et al.* (2013) suggest that decisions by operators were taken based on both forecasted values and uncertainty information show how the higher the uncertainty of a prediction, the more conservative the decisions taken were. Moreover, the same study pointed out that with uncertainty information, more optimal decisions were taken, improving the overall performance of the system (*Ramos et al.*, 2013).

A model can have many sources of uncertainty. Generally, the major source of uncertainty of a long-term forecast is the one related to the meteorological input (*Cloke and Pappenberger*, 2009). Observational uncertainty, which is related to the data assimilation, is also relevant because of the temporal and spatial uncertainties that can alter the system. Among the other sources of uncertainty, there are the geometry of the system and the characteristics of it, which are represented as parameters in the model and can be wrongly defined. For a more extensive discussion on the different types of uncertainty, refer to *Cloke and Pappenberger* (2009).

2.2.5 Forecast skill and value

There are two important factors when assessing the usefulness of forecasts: skill and value. The forecast skill is defined as the ability of the model of accurately predict the streamflow that will actually occur within given upper and lower bound (*Hamlet and Lettenmaier*, 1999). Thus, a skillful forecast implies that the value observed is within the estimation boundaries of the forecast model. The skill relies solely on the capacity of the forecast model and because of that, over the past 50 years, there has been a continuous improvement with the developing of new technologies and models (*Lynch*, 2008).

Skill alone is not enough to determine the quality of a forecast. For example, two forecasts with similar skill can generate different results for a water operation and there are cases in which, even if there is a perfect forecast with considerable skill, the operation might not change substantially because of other factors, such as structural limitation. An example is the case of a reservoir with a small storage capacity with respect to the incoming flood volume, which will not be able to buffer the flow peak even if this is perfectly forecasted (*Anghileri et al.*, 2016). The relationship between the forecast and the actual improvement to the operation is called forecast value. It is computed by comparing the system performance with the forecast and without it.

In a study on farmers decisions based on forecasts, *Li et al.* (2017) identified a clear non linear relationship between skill and value; moreover, they showed that most of the state-of-the-art weather forecast products still presented a limited accuracy for long-term forecasts. This inaccuracy generates doubts among farmers about forecast reliability and based on that, the quality of forecast did not necessarily improve proportionally the operation performance. On the other hand, farmer behaviour was shown to be a relevant variable influencing the estimated forecast: while a risk-neutral profile presented an increase of 3% with respect to the baseline case, a risk-prone profile experienced 10% of increment, since the operator was taking a bigger risk and relying more on the forecast result (*Li et al.*, 2017).

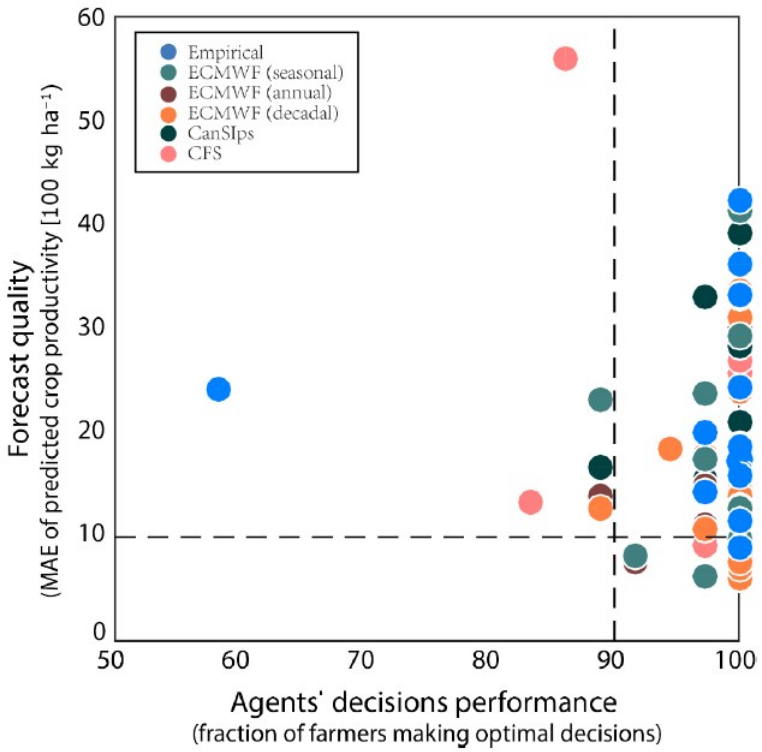

Figure 2.5: *Difference between accuracy and quality of a forecast and the actual improvement. Figure by Li et al. (2017).*

In a similar study exploring forecast skill and value for water supply, *Turner et al.* (2017) suggest that in order to minimize the risks of a misleading forecast and to potentialize its value, it is recommended to have long records of data and a large number of reforecasts (*Turner et al.*, 2017).

3

Study site

3.1 California

The state of California is located on the west coast of the United States of America and bordered by Mexico to the south (Figure 3.1. Globally recognized because of its many industries, it is the third largest state of the country in terms of area, the most populated one with roughly 39 million people and represents the largest economy in the country. California is also well known for its accentuated environmental awareness, and especially for advanced water management strategies. In fact, California has always had a deep dependency on water resources. They proved to be essential for the development of the state along the centuries in many areas, but especially for water-intensive activities, like mining and agriculture. Nonetheless, the state of California is now exposed to growing vulnerabilities generated by the recent occurrence of both extreme drought and flood events.

Figure 3.1: *Location of the State of California and its main cities. Adapted from Bing Maps (2019).*

3.1.1 Climate and water resources

Given its considerable size, the state presents a complex range of climate conditions, from deserts in the south to mountains in the north, including temperate rainforests, coastal and Mediterranean climates as well. All of this variety contributes to a diverse range of water availability and runoff with respect to the location, as illustrated in Figure 3.2. In addition, it is worth noting that the water variability not only varies geographically, but temporally as well, with a considerable variability across seasons and years. The majority of the state is considered to be located in a semiarid region, which entailed recent disputes over water sources between multiple users (*Escriva-Bou et al.*, 2016; *Hanak et al.*, 2011). Despite presenting a general symptom of water scarcity throughout the state, there is also a great risk of floods in parts of it and most of them happen to be exactly where cities and urban areas have been raised (*Hanak et al.*, 2011).

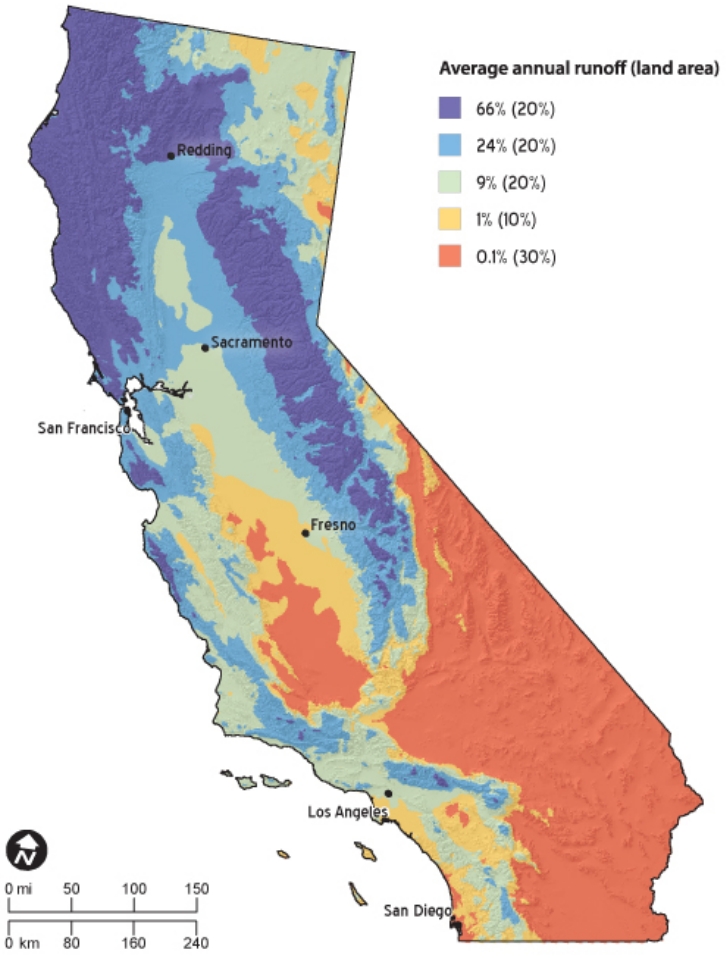

Figure 3.2: *Distribution of runoff. Figure by Hanak et al. (2011).*

However, even though the water systems of the state are already under stress, extra demand for high-quality water supply is expected, as the population is likely to keep growing and urbanizing. On top of that, climate change will also affect California's water management, either through temperature increase, sea level rise or precipitation changes, increasing the variability and unpredictability of hydrologic regimes and water availability. Without timely and effective adaptation strategies, these new circumstances are expected to negatively impact on California water resources in multiple ways, including the declination of native fish species, ultimately resulting in extinctions; more frequent and intense floods in urban areas; longer periods of drought and deterioration of water standards (*Hanak et al.*, 2011). A new and improved management of the water resources is deemed crucial for the future quality of life of the local population and for a sustainable development of the state.

3.1.2 Water network and management

In order to handle the multiple water demands from industries, farms and 40 million people, a complex system for water supply exists, consisting of a large network of reservoirs, groundwater basins, aqueducts, and pumping stations, most dating back to before 1960. Naturally, the management of such broad network of systems and infrastructures is a challenging task and only in the last decade an integrated approach including most parts of the system was conceived. This approach is supported by the CALVIN (California value integrated network) model originally developed by *Jenkins et al.* (2001), which simulates the dynamics of both surface and groundwater, while also considering the water demands and the operations of water facilities. The model aims at maximizing the economic value generated to the whole state (*Jenkins et al.*, 2001). A schematization of the CALVIN model, including 51 surface reservoirs, 28 groundwater basins, 18 urban economic demand areas, 24 agricultural economic demand areas, is illustrated in Figure 3.3.

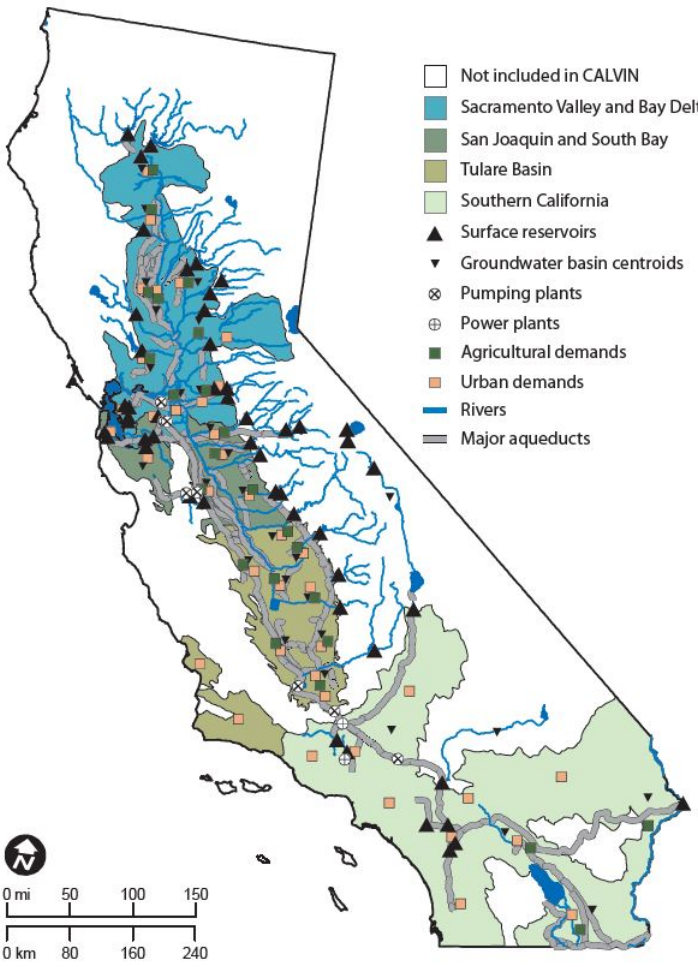

Figure 3.3: *California water supply system represented in the CALVIN model scheme. Adapted from Hanak et al. (2011).*

3.2 Folsom reservoir

The Folsom Dam and Reservoir was constructed by the U.S. Army Corps of Engineers and finished in 1956. Its primary objective is to prevent floods in the region of Sacramento, a highly vulnerable area to catastrophic floodings (*Hanak et al.*, 2011). The Folsom reservoir is currently managed by the U.S. Department of the Interior's Bureau of Reclamation and is part of the Central Valley Project, a water management project designed to supply water for the central region of California. Besides flood prevention, nowadays the reservoir is used also for hydroelectricity generation, water quality control, water supply, environmental preservation and recreation purposes. It is located in the central valley of California, between the city of Sacramento and the Sierra Nevada mountain range (figure 3.4), and is part of the American River Basin. The river

basin, considered to be one of the most important in the region because of its role in providing water fed by the melting snowpack of the Sierra Nevada to the central valley, covers an area of 1850 m^2 and ends up in the confluence with the Sacramento River, 48 km after the Folsom reservoir. Downstream of the Folsom dam, there is the lake Natoma, which helps adjusting the dam's regulations and fluctuations as a consequence of daily operations for power generation (*Maher*, 2008; *U.S. Bureau of Reclamation*, 2008).

Figure 3.4: *Location of the American River Basin and Folsom Reservoir. Adapted from Herman and Giuliani (2018).*

The dam's storage capacity is of 977 thousand acre-foot (TAF), covering a surface area of 11,450 acres. It is the main storage and flood control point of the river, with its dam walls standing 427 m tall (*Maher*, 2008; *U.S. Bureau of Reclamation*, 2008). The reservoir follows a historical operating strategies, mainly divided in two sections that varies according to the season: From November to February, which is the winter season, the conservation pool drops from 977 to 400 TAF, while the rest is used for flood control (*Nayak et al.*, 2018). After that period, the flood pool decreases and the conservation pool equals the maximum capacity at the beginning of the summer, as shown in Figure 3.5. More recently, an update was taken on the flood control policy, allowing for a smaller flood pool of 375 TAF, because the system got updated and is now benefiting from a short-term streamflow forecast, which enabled the anticipation from one to five days of the inflow (*U.S. Army Corp of Engineers and U.S. Bureau of Reclamation and California Department of Water Resources and H. I. Consulting*, 2017). For a better understanding, Figure 3.6 depicts the historical release of water per each day of the year from 1995 to 2016 and the mean daily values, which

are used as the daily demands value in this work. The release policy show a clear increase in demand during the irrigation period in California, the time of the year in which most of the water is used for agricultural supply, since there is no rain during this period and it is when the crops need it most to be able to grow. This policy, although recently updated, is likely to suffer from a decrease in performance, as the higher temperature in the future will provoke a decrease of the snowpack volumes and an earlier melting of the snow on the mountain range, deeply affecting the pattern of inflows of water into the reservoir (*U.S. Bureau of Reclamation*, 2008).

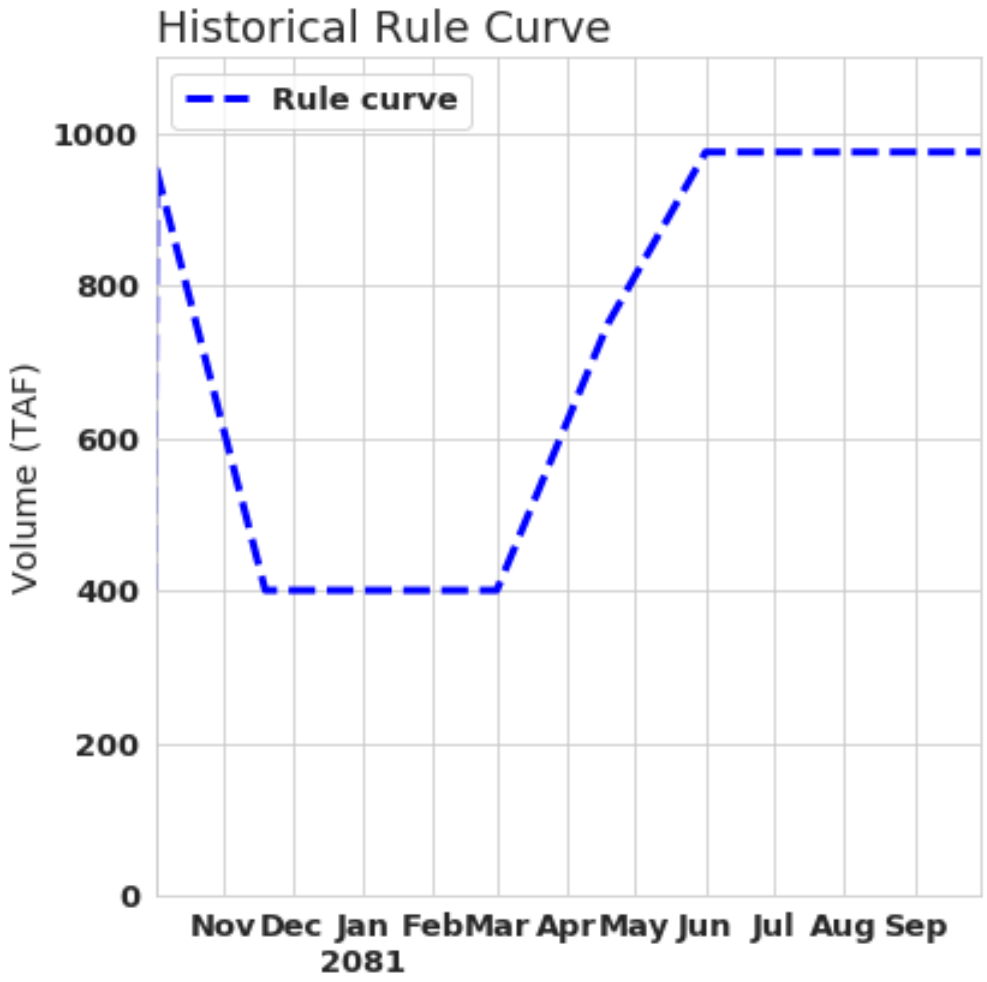

Figure 3.5: *Historical rule curve of the Folsom reservoir. During winter the conservation pool needs to be reduced to 400 TAF for flood prevention.*

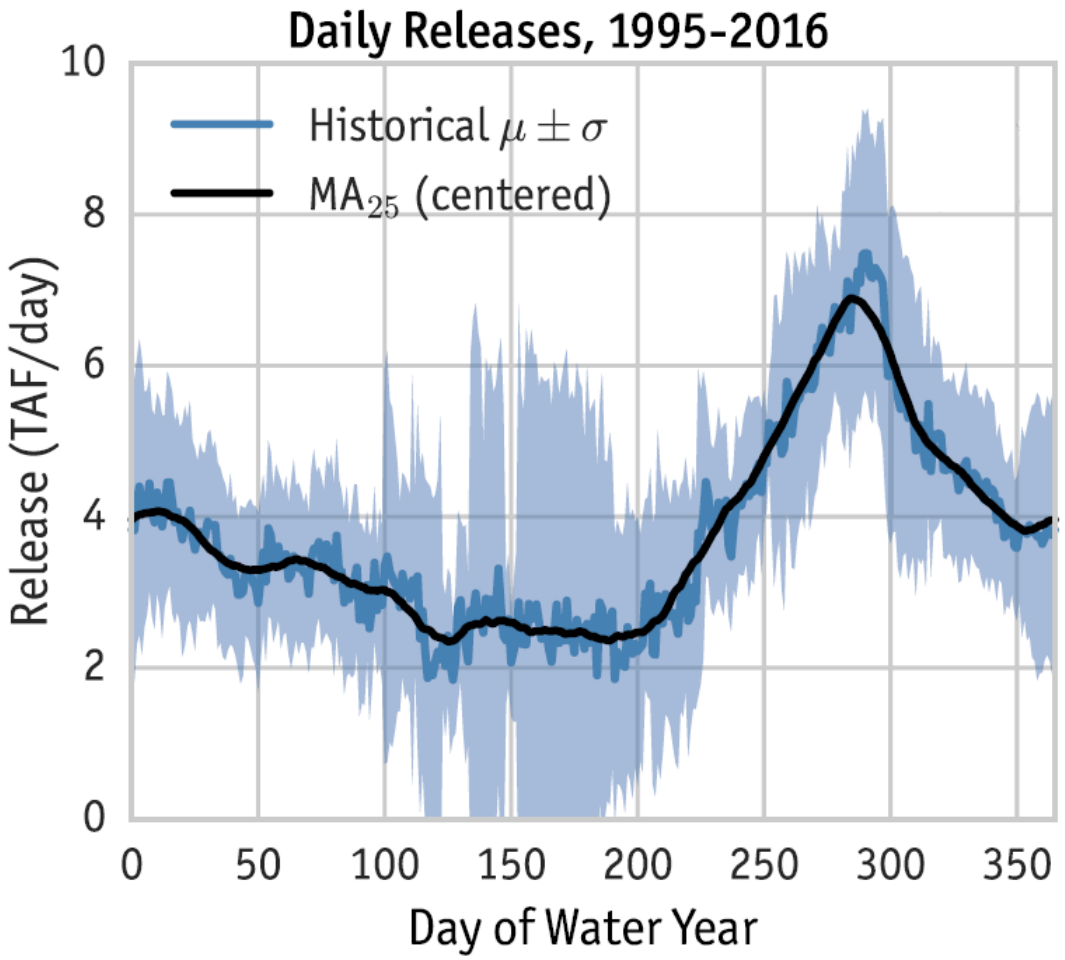

Figure 3.6: *Historical register of release of water per day. Adapted from Herman and Giuliani (2018).*

One relevant characteristic of the Folsom reservoir is that it had release capacity limitations. Nowadays, while its maximum release capacity is about 34,000 cubic feet per second (cfs) through the regular outlets, when the level of water reaches the spillway, the release can grow up to 115,000 cfs. In case of extreme floods, there are three emergency spillway gates that allow for a total release of 160,000 cfs for a limited amount of time (*U.S. Army Corp of Engineers and U.S. Bureau of Reclamation and California Department of Water Resources and H. I. Consulting*, 2017). The channel was also improved, with its levees reinforced, guaranteeing 160000 cfs of discharge capacity (*Maher*, 2008).

The dynamics of the Folsom reservoir are modeled by a mass balance equation of the water storage using a daily timestep, i.e.:

$$s_{t+1} = s_t - r_t + q_{t+1}, t \in [0, H] \tag{3.1}$$

where the variables s, r and q stand for storage, release and inflow, respectively. The multipurpose operations of the dam, targeting water supply and ensuring flood control, is modeled by means of the following objective function:

$$J = \frac{1}{H}\sum_{t=0}^{H} max(D_t - r_{t+1}, 0)^2 + \sum_{t=0}^{H} c * max(r_{t+1} - r_{max}, 0)^2 \tag{3.2}$$

In this case, r_{t+1} is the total release of water (including spills and constraints), r_{max} is the limit to a safe release, which for this work is $r_{max} = 130,000$ cfs and $c = 1000$ and represents a constant to make any flood event so costly that the evolutionary algorithm is expected to automatically avoid any case of flood. Therefore, this equation aims at reducing J by minimizing the deficits of water, especially large ones, since it is a quadratic function, while preventing releases above the limit to persist. In the case a final policy does result in a release value above the limit, it needs to be examined, as it implies that in the given scenario avoiding flood events is impossible, since all experimented solutions still suffered from flood events.

3.3 Data collection

The standard procedure for a long-term simulation of climate change requires both Representative Concentration Pathways (RCPs) and General Circulating Models (GCMs). Whilst the GCMs provide the proper mechanisms to process the climate along the years, the RCPs are entitled to guide the models through a pathway towards a specific target, providing all the information required. Naturally, different RCPs will generate different results, even if the model is the same for all of them. Likewise, running different models for a single RCPs will generate different results as well. In addition, for impact assessment, it is usual to downscale the GCM to a RCM so local results can be more reliable, as described in the previous subsection.

The data set used in this study includes 97 climate change scenarios of Folsom reservoir inflows over the time horizon from 2000 to 2100. These projections are the result of the combination of 31 GCM models with all four Representative Concentration Pathway (RCPs). Figure 3.1 provides further information on the models and their centres.

Modeling center	Institute ID	Model Name
Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Bureau of Metereology (BOM), Australia	CSIRO-BOM	ACCESS1.0
Beijing Climate Center, China Metereological Administration	BCC	BCC-CSM1.1 BCC-CSM1.1(m)
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
National Center for Atmospheric Research	NCAR	CCSM4
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(BGC) CESM1(CAM5
Euro-Mediterranean Center on Climate Change	CMCC	CMCC-CM
Commonwealth Scientific and Industrial Research Organization Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	LASG-CESS	FGOALS-g2
The First Institute of Oceanography, SOA, China	FIO	FIO-ESM
NASA Global Modeling and Assimilation Office	NASA GMAO	GEOS-5
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H-CC GISS-E2-R GISS-ES-R-CC
National Institute of Meteorological Research/ Korea Meteorological Administration	NIMR/KMA	HadGEM2-AO
Met Office Hadley Centre	MOHC	HadGEM2-CC HadGEM2-ES
Institute for Numerical Mathematics	INM	INM-CM4
Institute Pierre-Simon Laplace	IPSL	IPSL-CM5A-MR IPSL-CM5B-LR
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies	MIROC	MIROC-ESM MIROC-ESM-CHEM
Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology,	MIROC	MIROC5
Max Planck Institute of Meteorology	MPI-M	MPI-ESM-MR MPI-ESM-LR
Meteorological Research Institute	MRI	MRI-CGCM3
Norwegian Climate Centre	NCC	NorESM1-M

Table 3.1: *Table illustrating the costs of each scenario for different types of policy. The columns show the type of model used and the rows define the type of scenario for each time period.*

The observed inflow data for the Folsom Reservoir were acquired from the United States Bureau of Reclamation (USBR) and consist in daily observations of the inflow for the time period ranging from January 1915 to December 2015 as in *Nayak et al.* (2018). The data related to the forecasted inflows are also obtained from *Nayak et al.* (2018). In order to generate the hydrologic inflows, they used a forecast model that is a lumped version of the conceptual Sacramento Soil Moisture Accounting (SAC-SMA) model, which is the same one used for current operations at the Folsom Reservoir and simulates rainfall-runoff pro-

cesses in the upstream basin. The model uses rainfall, snowmelt and temperature as input and returns the inflow as output and the forecasts produced by the SAC-SMA model are 30 traces of daily prediction of the average inflow for the upcoming three days from January 1980 until December 2015. For more information and details about the processes and methods to generate the forecasted inflow, refer to *Nayak et al.* (2018).

3.4 Experiment settings

The policy tree for the Folsom case is designed so that there are two types of models, the rule curve and the three-day forecast cases, as mentioned in section 3.2. The algorithm optimization requires the decisions to be conditioned by the variable states, which depends on the type of model. For the first case, the rule curve, the two states available are the storage level S_t and the day of the water year *dowy*, measured in days, whilst for the forecast case, the 3 days ahead predicted storage is also condidered as reported in table 3.2. Given the probabilistic nature of the forecast (i.e., 30 member ensemble), it is also necessary to define which statistics of the ensemble to use for conditioning the operating policy. In this work, both the maximum and the 90th percentile are analyzed. The reason behind adopting upper-level percentiles is to be conservative with respect to the flood cases, since it is the main penalty factor and the forecast used is of short range, generally linked to flood prevention (review subsection 2.2.2).

State variable	Range	Unit	Policy
Day of the water year (*Dowy*)	0-365	Days	All
Storage (S_{t-1})	0-975	TAF	Rule curve
Storage + inflow + prediction ($S_{t-1} + Q + P3$)	0-975	TAF	Three-day forecast

Table 3.2: *State variables used for the policy tree algorithm.*

The actions available for the policy tree to choose from are "Release the demand", "Hedge p%" and "Release Excess". Release demand is the action where the operator is able to release the exact amount of water required for that day, without major problems. Hedge p% is a set of actions that allows for the releasing of a percentage from the demand, varying from 50% to 90%. The last action release excess is aimed at tackling the flood events, so the release of water is not ruled by the demand, but rather by the possibility of flood occurrence. All actions are summarized in table 3.3

Action variable	Target release rule	Policy
Release demand	$u_t = D_t$	All
Hedge p%	$u_t = (p/1000) * D_t$	All
Release excess	$u_t = K_f(Q_t + S_{t-1})$	Three-day forecast

Table 3.3: *Actions that can be taken for each policy. The rule curve is limited to either releasing demand or hedging a variation of p and three-day forecast policy has an additional action to anticipate floods.*

The different parameters used in this work are shown in table 3.4. There are two types of forecast use, a model without forecast and the other with a three-day one; three types of scenarios are used for the experiments, dry, intermediate and wet. Two time periods are considered, being the first one from 2000 to 2020, labelled historical and the the other from 2080 to 2099, the future one. The forecast ensemble is considered in two ways, a 90th percentile and the maximum value among the 30-trace ensemble. Finally, ten different random seeds are used for each experiment to minimize possible initialization issues.

Parameter	Alternatives	Number of alternatives
Forecast type	rule curve and three-day forecast	2
Scenarios	dry, intermediate and wet	3
Time period	historical and future	2
P3 function	max value and 90th percentile	2
Initialization	random seeds	10

Table 3.4: *Changes required in the parameters to allow for the proper simulations of each case.*

Finally, the settings of the Genetic Programming optimization are the following ones (as in *Herman and Giuliani* (2018)): each optimization was run for 100,000 function evaluations; the maximum tree depth is set equal to 5, the size of the population equal to 96, crossover and mutation probabilities equal to 0.7 and 0.9, respectively; due to the stochastic nature of the evolutionary algorithm (which can be affected by random effects in initial populations and runtime search operators), each optimization was run for 10 random generator seeds, as showed in table 3.5.

Parameter	Definition	Value
Max. NFE	maximum number of nodes per case	100,000
d_{max}	maximum depth allowed for the tree	5
μ	size of the parents pool	10
λ	size of the population	96
p_c	crossover probability	0.7
p_m	mutation probability	0.9

Table 3.5: *Common parameters for the entire simulation.*

In total, 1800 optimization trials were run, requiring approximatively 5760 computing hours. The computational experiments were parallelized and run using the High Performance Computing (HPC1) cluster located at University of California Davis, California. The cluster contains 60 nodes with 32 cores each that reach up to 2.4GHz.

4

Methods

The aim of this work is assessing how beneficial can the use of forecast models be to future water systems operations under climate change scenarios. In order to properly achieve this goal, a three-step methodology as illustrated in figure 4.1 was developed: Scenario selection, synthetic forecast generation and assessment of historical and future forecast value. All three blocks are explained in the coming sections.

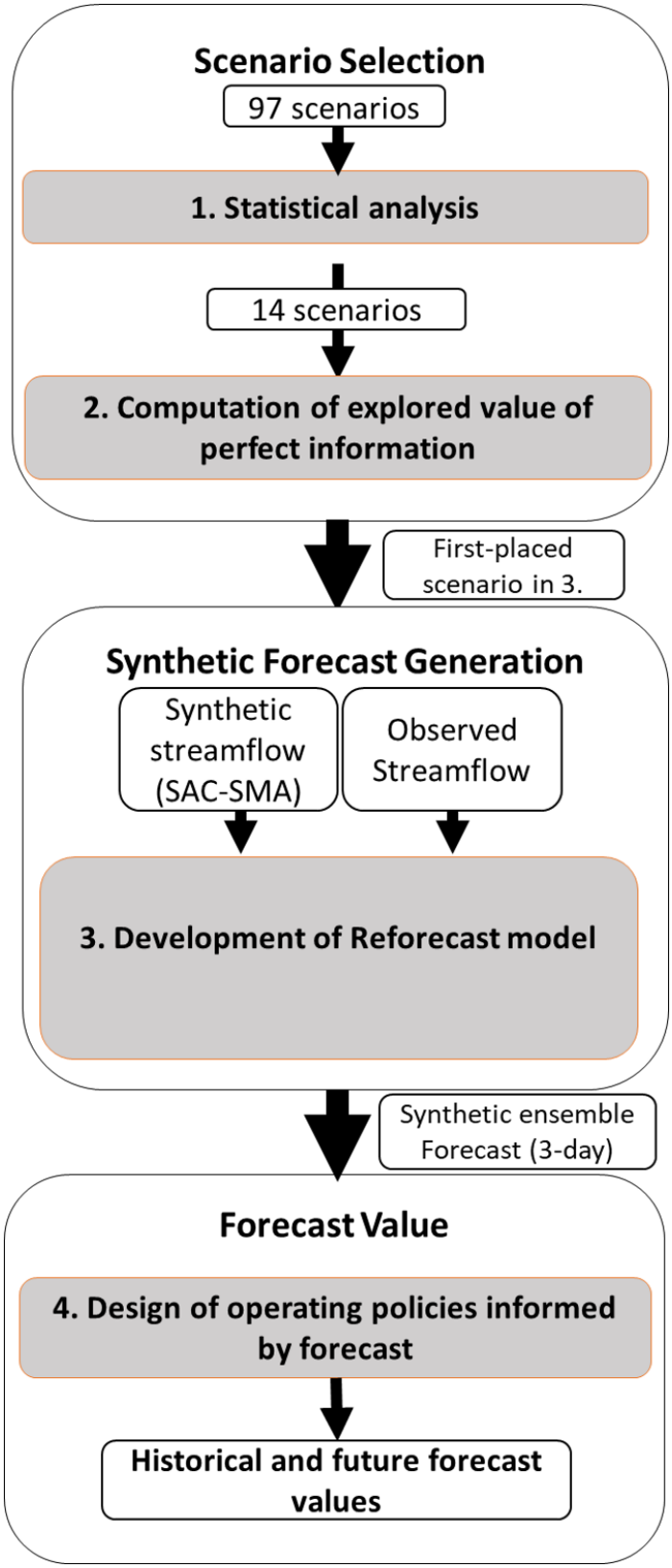

Figure 4.1: *General structure of the methodology used in this work.*

4.1 Scenario Selection

The first block of the proposed methodology aims at analyzing all inflow scenarios to the Folsom reservoir obtained from the work by *Herman and Giuliani* (2018) for selecting a subset providing diverse projections (e.g., drier or wetter conditions) that are expected to benefit from the use of forecast information. This is due to the fact that the longer the Expected Value of Perfect Information is (i.e., the difference in performance between the basic policy and the policy with perfect forecast) the bigger the potential forecast value is.

4.1.1 Statistical analyses

The first step to find promising candidate scenarios was applying basic statistical indexes: the arithmetic mean, variance, minimum and maximum values of the daily inflow series along the 100 years of prediction for each scenario. The trend of these statistics is also analyzed, such as an increase of variance or maximum values over time. The statistics indicate the possibilities of observing higher or lower volumes of inflows and increased unpredictability, which are expected to challenge the current operations of Folsom Reservoir.

In order to support the indices conceived so far, it was decided to directly apply the inflow data into the RClimDex library to use its values as a performance benchmark (*Zhang and Yang*, 2004). RclimDex is package in R (a free programming language) that calculates the extreme climate indices as defined by the Expert Team on Climate Change Detection and Indices (ETCCDI). For this work purpose, the library was adjusted for streamflow and the indices considered were: Annual total wet day precipitation, very wet days, extremely wet days and number of days above a given threshold. Beside standard statistics, the annual number of days when the inflow exceeds the 90th percentile and the ones below the 10th percentile is computed for each scenario, to capture the projected trends in extreme floods and droughts. The percentiles used as reference level were calculated and defined over the first two decades of the century (2000-2020). Again, greater occurrences of extreme events are expected to challenge systems without forecast. Moreover, based on the climate indexes defined by (*Zhang and Yang*, 2004), the same index was also computed using a different reference value for each day of the year in order to account for possible changes in the timing of future hydrologic regimes.

The last index considered is the length of extreme events. For each scenario,

the trend in the annual longest flood or drought spell is the analyzed. Also for this case, a second index was introduced to account for seasonality.

4.1.2 Computation of explored value of perfect information

With the most promising scenarios selected, they are taken to the second step of the first block, for deepening the analysis by computing the Expected Value of Perfect Information of each one. To calculate the perfect operating policy (POP), the deterministic dynamic programming (DDP) was considered, which allows for an optimal policy because it minimizes the objective function (eq. 3.2) by recursively computing the optimal cost-to-go as in:

$$H_t(x_t) = min_{u_t} \Psi_{\epsilon_{t+1}} \Big[\Phi \Big[g_t(x_t, u_t, \epsilon_{t+1}), H_{t+1}(x_{t+1}) \Big] \Big] \tag{4.1}$$

where H_t is the optimal cost-to-go with respect to the state x_t, Ψ is a statistic used to filter the disturbance, Φ an operator for aggregation over time, g a step-cost function generated from one time-step to the other, u_t decision or control vector and ϵ the disturbance vector.

For the calculation of the basic operating policy (BOP), the baseline operation of the policy tree for the Folsom Reservoir was used. This is due to the fact that in *Herman and Giuliani* (2018) the policy tree is used to design both the policies informed by the forecast as well as the baseline solutions that do not use forecast information. In fact, an optimal policy tree dependent on day of the year and reservoir storage represents a good approximation of the historical operations of Folsom Reservoir, which makes it the most suitable candidate for the basic operating policy (BOP). After obtaining the BOP and POP performance, the EVPI can be measured in two ways: the first one is a simple subtraction of the lower boundary from the upper boundary, which is considered to represent the absolute performance increase; the second one is instead the ratio of the two performance that measures the relative performance increase.

Given the many statistical indices introduced in subsection 4.1.1, 19 scenarios are selected and the EVPI is computed for each scenario. Some scenarios can then be discarded looking at BOP and POP performance. Specifically, in case the BOP performance (i.e. water deficit) is equal to zero, there is no need of introducing forecast for improving the baseline operations. At the other extreme, if the POP generates a flood, the scenario is also excluded because the solutions informed by the forecast cannot outperform the POP.

4.2 Synthetic Forecast

After having selected a subset of climate scenarios, the second block of the proposed methodology implements a synthetic forecast generation model to produce a forecast ensemble for each climate scenario. The synthetic forecasts allow quantifying the actual improvement in system performance obtained by operating the system with a policy informed by the forecast with respect to the BOP, which does not use any forecast. This value is known as Expected Value of Sample Information as it does not account for perfect information as the EVPI, but rather relies on the available data at the time of the operation.

4.2.1 Type of residual and preliminary tests

Among the different approaches for generating a synthetic forecast, the one used in this work resides in comparing forecasted inflows to the observed ones over the historical period in order to model the forecast errors. A crucial aspect in this process is to determine which type of residual should be used because it can have multiples forms, such as the difference or ratio between observed and forecasted values and in different scale, like normal scale or logarithmic scale. Some tests need to be run to select the best approach for modeling the forecast error.

The first test is the histogram of the residuals, which is a graph that allows the comparison between the frequency of the residuals and the range of the residual values. Residuals that present a bell-shaped distribution are usually an indicator of a normal and random distribution, whilst other formats suggest there is some sort of bias in the residuals. The second test, normal quantile-quantile (Q-Q) plot, compares the residuals to theoretical values in order to define if there is a normal distribution of the data. Should it be normal, there ought to be a straight line roughly at an angle of 45 degrees, otherwise it suggests once more that there is some sort of bias or non-random pattern in the residual data. Then, there are two tests for memory, autocorrelation function (ACF) and partial autocorrelation function (PACF). ACF aims at evaluating if the residuals are in any way related along time between themselves. The PACF demonstrates the linear relationship between one value and the past one without considering the other past values. Both ACF and PACF indicate that the residuals are not random if there is a significant memory in the series of residuals. The last test is called heteroskedasticity. It measures how the residuals behave for different values of forecasted observations. If the data follow a vi-

sual pattern like a cone or a fan, it means the residuals vary according to the magnitude of the forecast, which suggests residual data are not random and can be represented by a mathematical function.

After establishing the best representation of the residuals and how they behave, the synthetic forecast model reproduces the error distribution and propagate the forecast uncertainty to any inflow trajectory, either back in the past when forecasts were not generated yet (*Nayak et al.*, 2018) or, as it is the case for this work, in the future for each climate change scenario.

4.2.2 Synthetic generation

The process of generating the synthetic traces of an existing forecast model consists on a two-step procedure, relying on the use of the K-nearest neighbor algorithm (KNN). First step requires two types of datasets, one is the simulation dataset, composed by observed and projected inflows and the second is the training dataset, composed by historical observations that have corresponding forecast data. For every t of the simulation, the algorithm measures the Euclidean distance d_j between the observed inflow from the simulation dataset, Q_t, and all the observed inflows from the training dataset, as it can be seen in the following equation:

$$d_j = \sqrt{(Q_t - \sum_{t^*=1}^{n} Q_{t^*})^2} \tag{4.2}$$

where Q_{t^*} is the observed inflow of the training dataset corresponding to the training day t^* and n stands for the length of the training data. This steps generates a pool with all the observed inflow from the training set and their corresponding Euclidean distance to the Q_{t^*}. To select the best matches, the size of the pool is reduced and the value K is responsible for that, which, according to *Lall and Sharma* (1996), should be $K = \sqrt{\frac{n}{n_s}}$, where n_s stands for the number of seasons. The size of the reduced pool means that only the K closest values are considered when selecting the suitable match, and these values do not have equal weight, as they follow a kernel density function:

$$f(d_j) = \frac{1/j}{\sum_{j=1}^{K} 1/j} \tag{4.3}$$

The closer to the observed data a value is, the higher its weight is and the more probable it is to be chosen. Then, with each observed value assigned a new forecasted value, the forecast error can be computed. For this study, it is

adopted as following:

$$\hat{\varepsilon}_{Q,t^*+1:t^*+3} = log\hat{Q}_{t^*+1:t^*+3} - logQ_{t^*+1:t^*+3} \tag{4.4}$$

where $\hat{\varepsilon}_{Q,t^*+1:t^*+3}$ is the error originated by the difference between the observed $Q_{t^*+1:t^*+3}$ and the forecasted $\hat{Q}_{t^*+1:t^*+3}$ inflows for the forecasted period. The^above the Q indicates that this is an estimate or forecasted value and not the actual one.

The second step consists in generating the forecast out of the observed inflows. As each observed value now has a corresponding residual, in order to generate the synthetic forecast, the forecast error should be added to the observed inflow, i.e.:

$$\hat{Q}_{t+1:t+3} = Q_{t+1:t+3} + \hat{\varepsilon}_{Q,t+1:t+3} \tag{4.5}$$

A brief summary of the process is that the observed inflows from the simulation dataset (Q_t) are assigned residuals values by adopting the residuals from corresponding observed inflows from the training dataset (Q_{t^*}). This matching process is performed by the KNN algorithm. Then, with every observed inflow from the training set coupled with a residual value, the synthetic forecast ($\hat{Q}_{t+1:t+3}$) is generated by adding the former two values.

After training and testing the synthetic forecast for the entire simulation dataset period, it is required to check the quality of the synthetic forecast with respect to the original forecast. To do so, a residual analysis is carried out, which involves:

- Comparison between forecasted series and observed series for both synthetic and historical values.
- Error distribution analysis between historical and synthetic values.
- Histogram of the residuals.
- Autocorrelation functions for the residuals.

Should it be considered satisfactory, then the model is properly calibrated and ready.

4.3 Forecast Value

The third and last block of the proposed methodology aims at assessing the actual improvements the use of forecasts can bring to the system operations under historical and future conditions. This step required the decision of operating policies informed by the synthetic forecast generated in the previous block and the comparison of their performance with the references provided by BOP and POP solutions.

4.3.1 Policy tree

The method used in this work to design an operating policy that can be informed by inflow forecasts is called policy tree *Herman and Giuliani* (2018). Policy tree is a direct policy search approach where the closed loop operating policy is defined as a binary tree optimized via genetic programming. Decision variables are discrete actions ($a_i \in A$), along with specific thresholds activating then on the basis of the observed system conditions (state variables x_t). The overall solution of the problem means finding the optimal policy tree structure and thresholds (T^*), i.e.:

$$T^* = argmin_T J(s_0, T, q_1^H) \tag{4.6}$$

During the simulation, at any time t, the optimal action to be taken depends on the system state x_t, i.e. $a_t^R = T^*(x_t)$. Whenever there is the need to include forecast information into the policy tree, the release decision has to consider the predicted storage three days in advance, which is based on the previous day storage, the current inflow and the predicted incoming inflow for the next three days, as the following equation demonstrates:

$$\hat{S}_{t+3} = S_{t-1} + Q_t + \hat{Q}_{t+1:t+3} \tag{4.7}$$

4.3.2 Genetic Programming

Tree-structured operating policies are developed by means of genetic programming, a sub-domain of evolutionary computation that searches and optimizes solutions mimicking the natural selection, a widely recognised mechanism of the evolutionary biology. The main steps involved in the a genetic algorithm are: 1) Survival of the fittest, 2) Crossover and 3) Mutation.

For a proper explanation of how the policy is optimized, an introduction to the components of the policy tree is required. A binary tree has only two types

of nodes: the indicator node, which displays a state variable (x_t) in comparison with a threshold value (k), and the action node (a_t), which instructs which decision to be taken given the indicator node state. The trees follow a binary structure: for each indicator node, there is always a true and a false option, which can then give place to another indicator node or to an action node, as seen in figure 4.2.

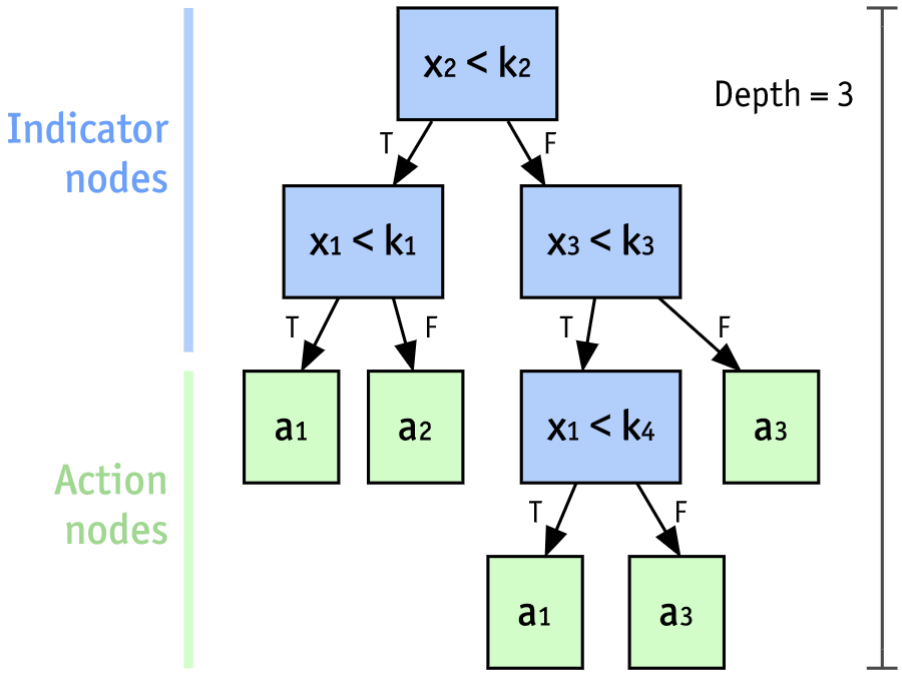

Figure 4.2: *Representation of a policy tree model. Figure by Herman and Giuliani (2018).*

At the very start of the simulation, there is no policy tree structure and in order to generate one, a process known as initialization is required. The initial population must be randomly generated and to do so a random indicator node with a random threshold is created. Then it follows a process of appending and assigning random indicators, thresholds and actions up to the maximum tree depth allowed (d_{max}, an hyperparameter to be set before running the optimization). Since the algorithm is basing itself on natural selection, only a set composed of the best performing solutions is as the "parents" for the next iteration. The underlying concept is to advance at each iteration configurations offering the most suitable solution, minimizing the equation 4.6. In order to avoid the local optimum issue, usually linked to the starting population or location, the algorithm is ran multiple times with different and random initialization data.

The second step, crossover, is the process entitled for the generation of the

next population, the "children" solutions. It consists in generating a child by swapping subtrees between two parents with a specific probability. A constraint is that crossovers are only allowed up to the maximum depth, to avoid exchanging an action node for a indicator node and thus increasing the true size. Child trees that are not the result of a crossover yet part of the new population are randomly changed by a new tree following the initialization procedure. Then, in the third step, mutation, each node of the tree, either indicator or action, has a probability of being altered randomly.

In addition to selection, crossover and mutation, the policy tree optimization implements also the pruning operator, with the purpose of eliminating redundancies by eliminating possible distortions generated by the crossover and mutation processes, such as both true and false branches resulting in the same result or unreachable branches. All these operators are illustrated in figure 4.3.

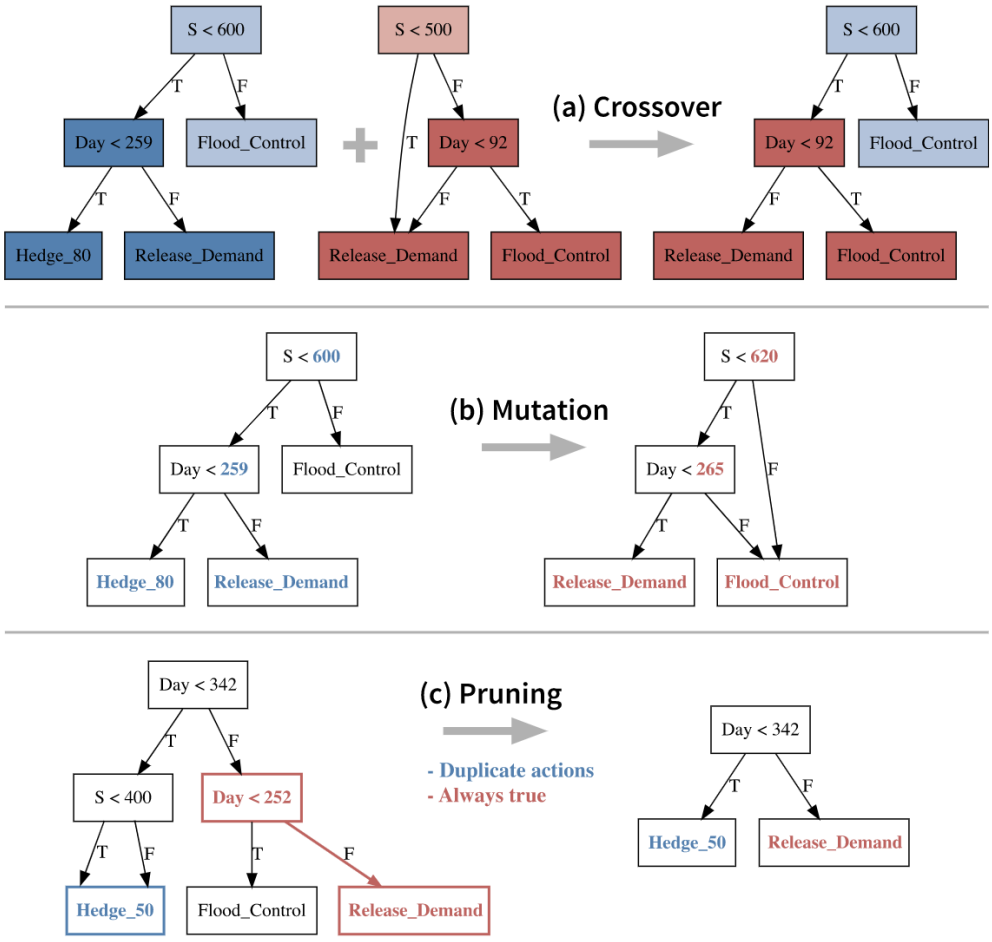

Figure 4.3: *Examples of processes in the genetic algorithm. a) Crossover, b) Mutation and c)Pruning. Figure by Herman and Giuliani (2018).*

5 Results

The numerical results are presented in this chapter following the three main blocks of the methodology illustrated in Chapter 4, namely scenario selection, synthetic forecast generation, and quantification of historical and future forecast value.

5.1 Scenario Selection

Starting from the 97 scenarios of future inflows to the Folsom reservoir, this section serves the purpose of a selection process in order to single out subsets of candidate scenarios that are expected to benefit the most from the use of forecast information. This section is further divided into two sub-sections, the first being a preliminary screening based on statistics, while the second runs a more profound simulation-based analysis to determine quantitatively the potential of each candidate. Two subsets of selected scenarios are generated, a larger pool containing the scenarios selected from the first step for experiments that require more data and a smaller pool with the scenarios obtained in step two. These are representative scenarios of different climate conditions that allow for a wider range of experiments to be tested with better readability.

5.1.1 Preliminary analysis

The 97 scenarios present different projections of inflow for the Folsom reservoir, and this is due to the distinct combinations of Representative Concen-

tration Pathways, Global and Regional Circulation Models used for projecting the climate conditions. Nevertheless, there are some similarities across these scenarios as illustrated in Figure 5.1, where all scenarios are averaged over a calendar year. The first four months of the year represent the wet period of the year, with special attention to the period ranging from beginning of January until end of February, with values up to 44 TAF/d (628,16 m^3/s). This indicates high chances of floods taking place at this period, but a lot of uncertainty is present, as seen in the figure, which makes it troublesome to quantify with precision the actual flood risk. Then, from March until the end of April, a progressive decrease of inflow is observed until summer arrives and the inflow of the reservoir reaches 0 TAF/d (0 m^3/s). Even though other periods of the year might suffer from a highly variable expected inflow value on account of the multiple projections linked to different scenarios, the summer season shows essentially no variation between the scenarios. This evidence indicates that acute droughts are expected to take place with a high degree of confidence. The last part of the year sees an increase in the inflow, but then again with substantial fluctuations, ranging from low to flood-level values.

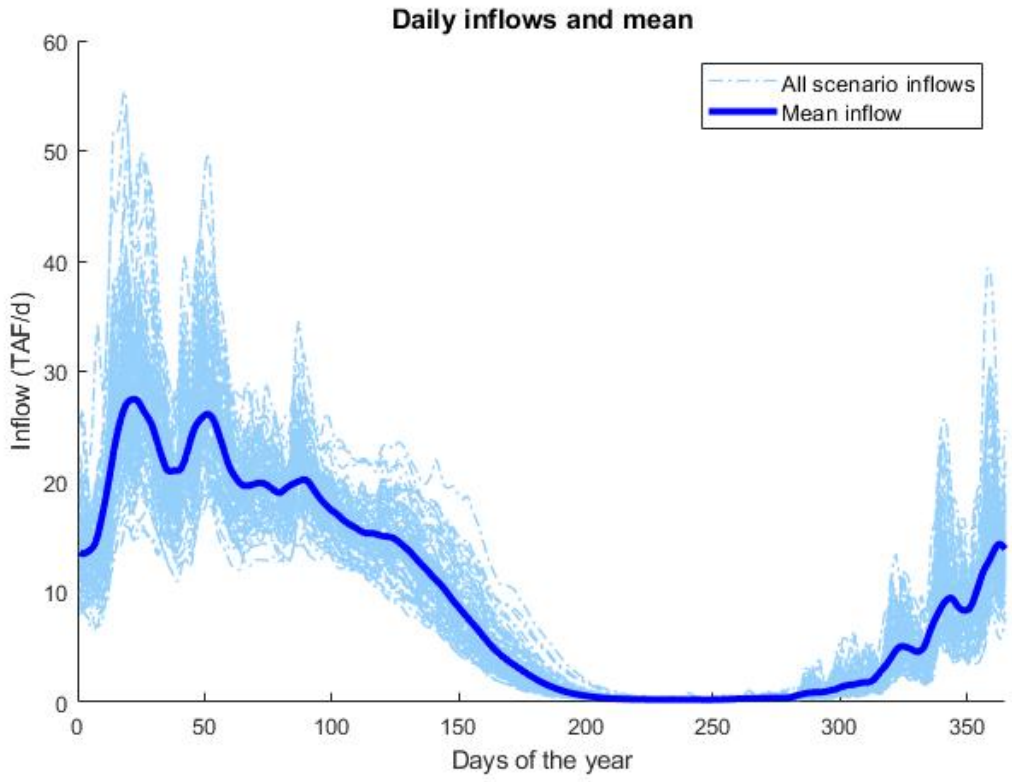

Figure 5.1: *Daily inflow for all 97 scenarios and the mean inflow among all scenarios for a calendar year.*

Climate change is a dynamic phenomenon which translates to impacts varying according to time. Therefore, the effects are not only influenced by the scenarios, but also by the progression of the years. Additional analysis can be done on the data in order to try to better understand the trends along the years, as illustrated by the boxplots in Figure 5.2, which show the average trend across the scenarios from 2020 to 2099. The figure shows that while the average values of the inflow along the time series slightly decreases, the lower quartile values are increasingly approaching the zero inflow. Conversely, the upper quartile keeps approximately the same levels, but the upper extreme are constantly increasing. This confirms that while the majority of the scenarios will see a decrease of the future inflows, generating long and intense droughts, extreme flood events are also projected to increase.

To better understand the projected inflow trends, a three-dimensional barplot is reported in Figure 5.3, in order to provide additional information about how the inflows (averaged for the 97 scenarios) behave with respect to the days and to the years. The trajectories are shown from January to May to highlight the most critical period for the operations of the Folsom reservoir during the transition from the wet to the dry season. Beside confirming the increasing frequency of high flow episodes, the figure also clearly shows the projected reduction of the summer inflow: taking as reference the day 130 in the graph, the inflows referring to the first years are around the 14 TAF/day (199.87 m^3/s) and they decrease smoothly and constantly until reaching approximately at 5 TAF/day (71,38 m^3/s).

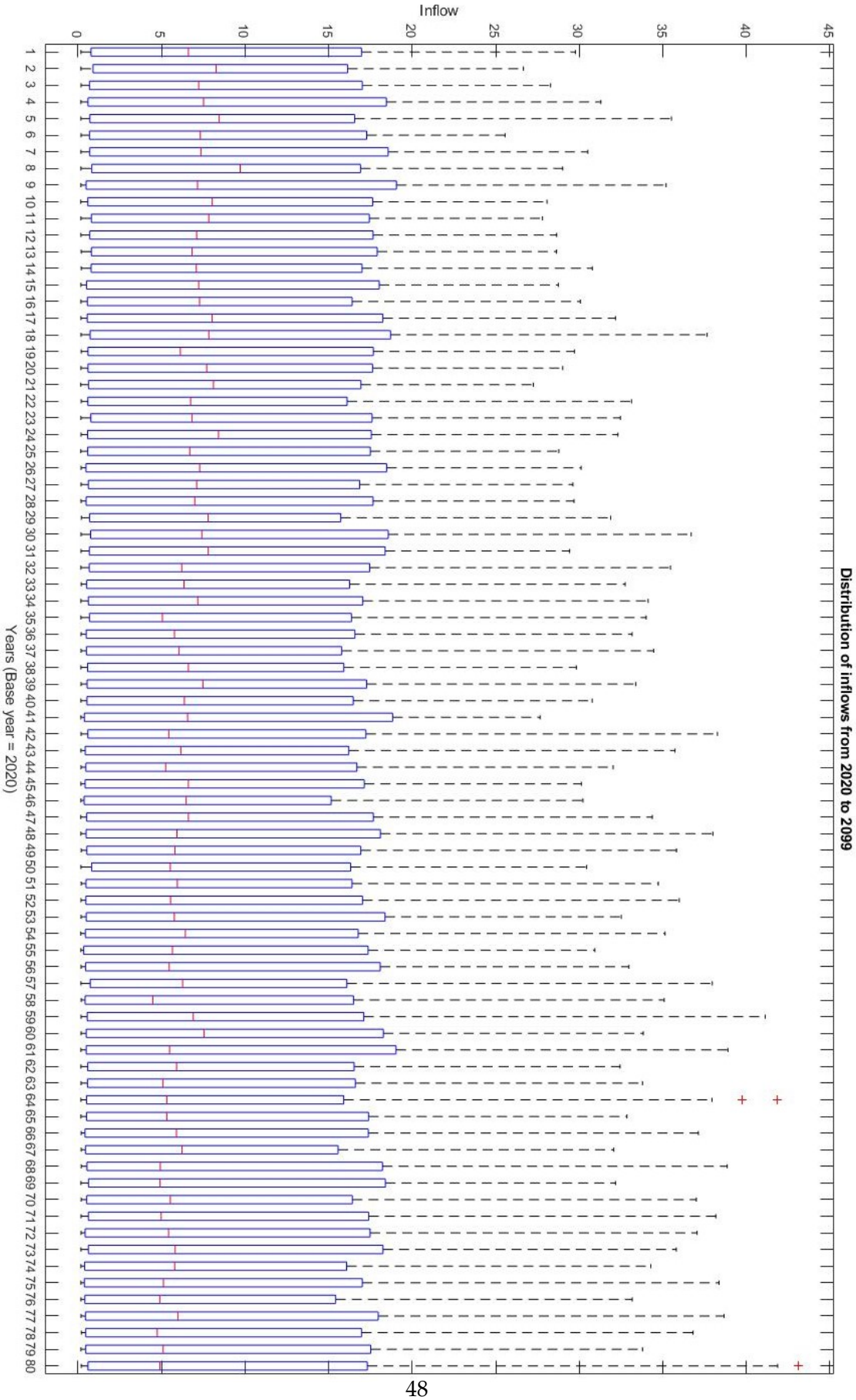

Figure 5.2: *Boxplot with the pattern of inflows along the years 2020 until 2099.*

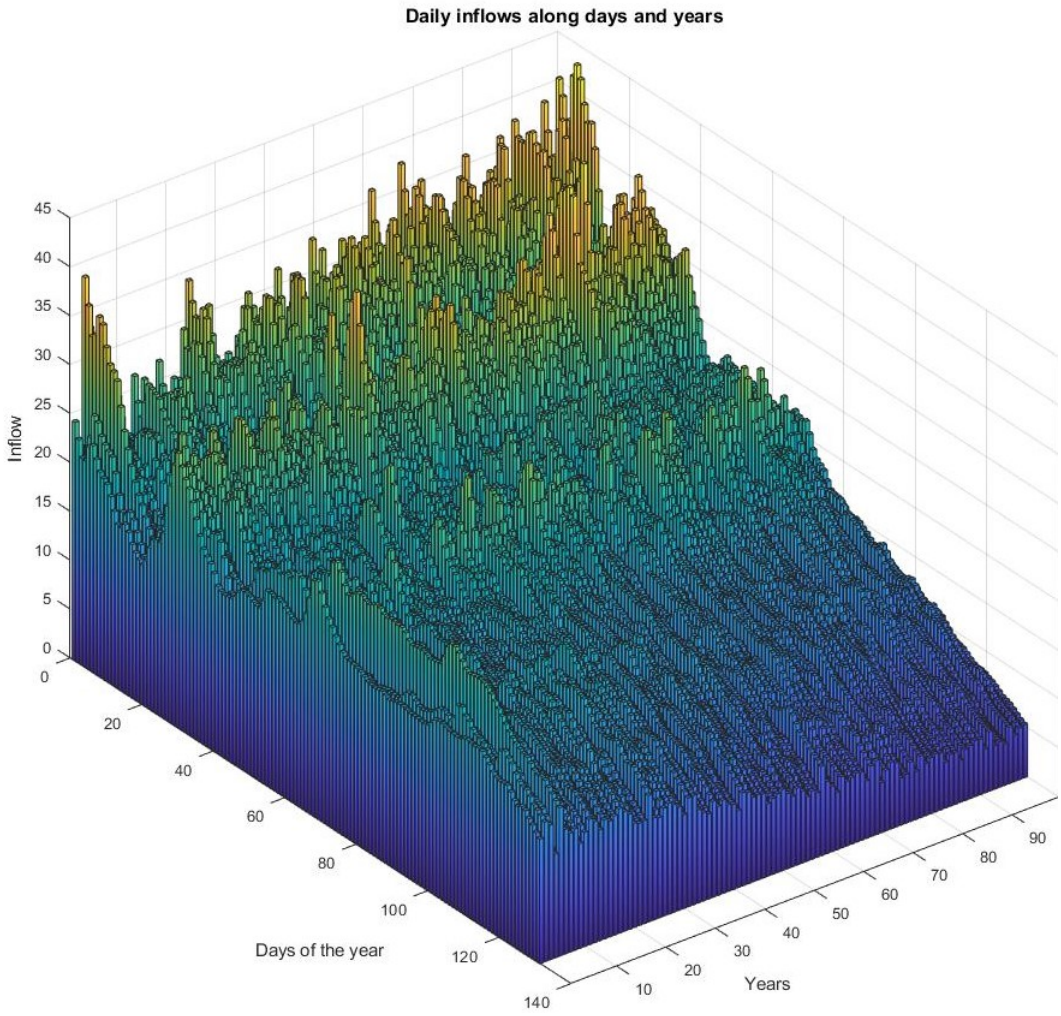

Figure 5.3: *Daily inflow variability with respect to the both days and years. The anticipation of the dry period can be seen along the years as well as an increase of the flood events during winter.*

5.1.2 Statistical analysis

Starting from the visual analysis presented in the previous section, the main characteristics of the future scenarios are then discussed by computing some statistics, such as mean, variation, maximum values and minimum daily values displayed in table 5.1, but also the number of drought and flood days (defined in section 4.1.1).

The same statistics, now computed over the 2070-2100 period which present the more evident changes in the projected inflows with respect to the historical ones, are illustrated in Figure 5.4. The red bars identify candidate scenarios to be selected, which are listed in Table 5.2. Not surprisingly, the majority of the these scenarios have in common the RCP8.5, the most extreme Radiative

Index	Scenario	Frequency
Max mean	11	9
Min Mean	79	7
Variance	11	10
Maximum	25	7
Minimum	82	5
Counter flood	93	6
Counter drought	80	7
Counter flood season	52	6
Counter drought season	5	6
Length flood	51	4
Length drought	40	6
Length flood season	52	4
Length drought season	5	7

Table 5.1: *Statistical characteristics to be assessed, the scenarios selected and the frequency of these scenarios as the most suitable for each index.*

Concentration Pathway.

Scenario	Model name	RCP
2	access1-0	RCP8.5
5	bcc-csm1-1	RCP2.6
11	canesm2	RCP8.5
17	cesm1-bgc	RCP8.5
20	cesm1-cam5	RCP6.0
24	cnrm-cm5	RCP4.5
25	cnrm-cm5	RCP8.5
32	fgoals-g2	RCP8.5
40	gfdl-cm3	RCP8.5
51	giss-e2-r	RCP2.6
52	giss-e2-r	RCP4.5
66	inmcm4	RCP8.5
70	ipsl-cm5a-mr	RCP8.5
79	miroc-esm-chem	RCP6.0
80	miroc-esm-chem	RCP8.5
82	miroc-esm	RCP4.5
84	miroc-esm	RCP8.5
93	mri-cgcm3	RCP8.5
95	noresm1-m	RCP4.5

Table 5.2: *All scenarios selected, their model names and RCPs.*

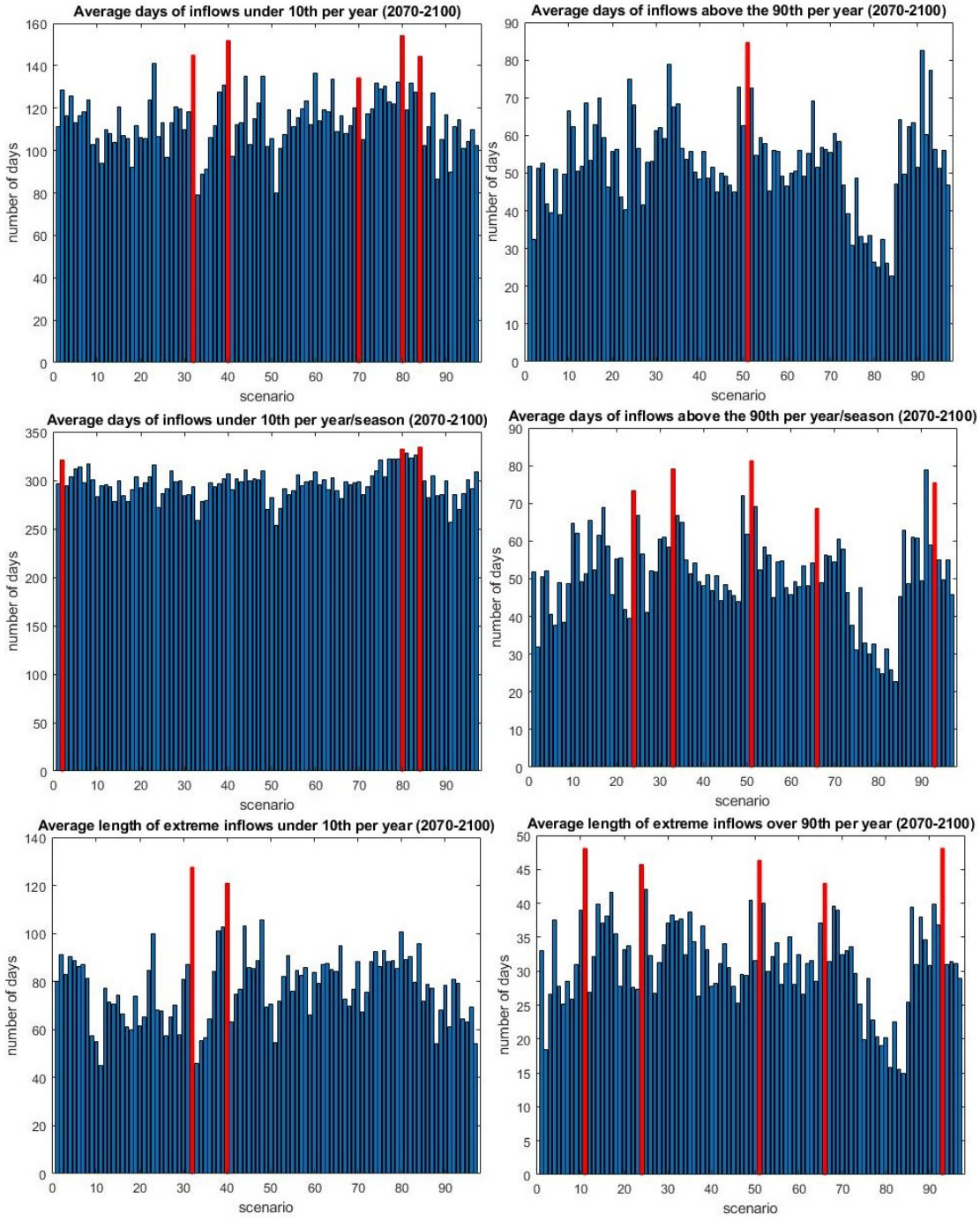

Figure 5.4: *Different criteria displayed in bar graphs and the selected scenarios in red.*

To complement this statistical analysis, four RClimDex indices, adapted from their original formulation for meteorological variables (*Zhang and Yang*, 2004) to be used for inflow scenarios, are analyzed. As it can be seen in table 5.3, the correlations between the statistical indices and the indices from the RClimDex library are generally significant.

Index	TOT. INFLOW	COUNTER R90	R95P	R99P
Mean	1.00	0.94	0.88	0.80
Var	0.73	0.47	0.92	0.95
Min	0.55	0.63	0.37	0.23
Max	0.41	0.15	0.56	0.65
Flood Legnth	0.71	0.69	0.62	0.58
Flood Length Seasonal	0.70	0.67	0.59	0.54
Flood Counter	0.94	1.00	0.69	0.58
Flood Counter Seasonal	0.95	1.00	0.71	0.60
Drought Length	-0.46	-0.36	-0.44	-0.55
Drought Length Seasonal	-0.67	-0.78	-0.39	-0.33
Drought Counter	-0.72	-0.75	-0.48	-0.45
Drought Counter Seasonal	-0.91	-0.99	-0.63	-0.52

Table 5.3: *Correlation between indices used at this work and the ones from RClimDex.*

5.1.3 Computation of the expected value of perfect information

After the statistical selection process, 19 scenarios were selected and then analyzed looking at the Expected Value of Perfect Information. The estimated values of EVPI are reported in Table 5.4 and illustrated by the figure 5.5.

Scenario	POP $(TAF/day)^2$	BOP$(TAF/day)^2$	Absolute$(TAF/day)^2$	Relative
Hist	0.127	0.309	0.182	0.59
2	0.05	3.72	3.67	0.99
5	0.47	3.59	3.12	0.87
11	404730000.00	976990000.00	flood	flood
17	0.01	242610000.00	flood	flood
20	0.31	4.05	3.74	0.92
24	0.01	3.29	3.28	1.00
25	0.14	55471000.00	flood	flood
32	0.11	2.54	2.43	0.96
40	0.64	4.94	4.30	0.87
51	0.00	0.00	0.00	NA
52	0.00	0.00	0.00	NA
66	480200.00	480200.00	flood	flood
70	0.79	2.55	1.76	0.69
79	1.31	2.69	1.38	0.51
80	2.50	5.49	2.99	0.54
82	0.52	0.83	0.32	0.38
84	1.07	1.40	0.33	0.23
93	0.01	0.18	0.17	0.93
95	106150000.00	545958495892.64	flood	flood

Table 5.4: *Perfect operation policy and basic operation policy performance; the absolute EVPI and relative EVPI computed over the selected future scenarios and over the historical period.*

The obtained values of EVPI characterize two types of results. Scenarios 51 and 52 present an operational cost of 0 $(TAF/day)^2$ for both deterministic and stochastic cases. This indicates that a forecast is not necessary for such scenarios as it is already possible to achieve perfect operation without it. These scenarios have been therefore not considered in the following steps of the proposed procedure. Conversely, scenarios 11, 66 and 95 show a much higher operation cost for both BOP and POP, where values over thousands of $(TAF/day)^2$ are a result of the flood penalty in equation 3.2. This result suggests that these scenarios predict critical flood events that cannot be controlled even with a perfect knowledge of the future. Since informing the operations with the synthetic forecast cannot outperform the POP solution, also these scenarios will not benefit from forecast information and are not considered in the following steps. In conclusion, only 14 scenarios are left to be used for the later stages of this work.

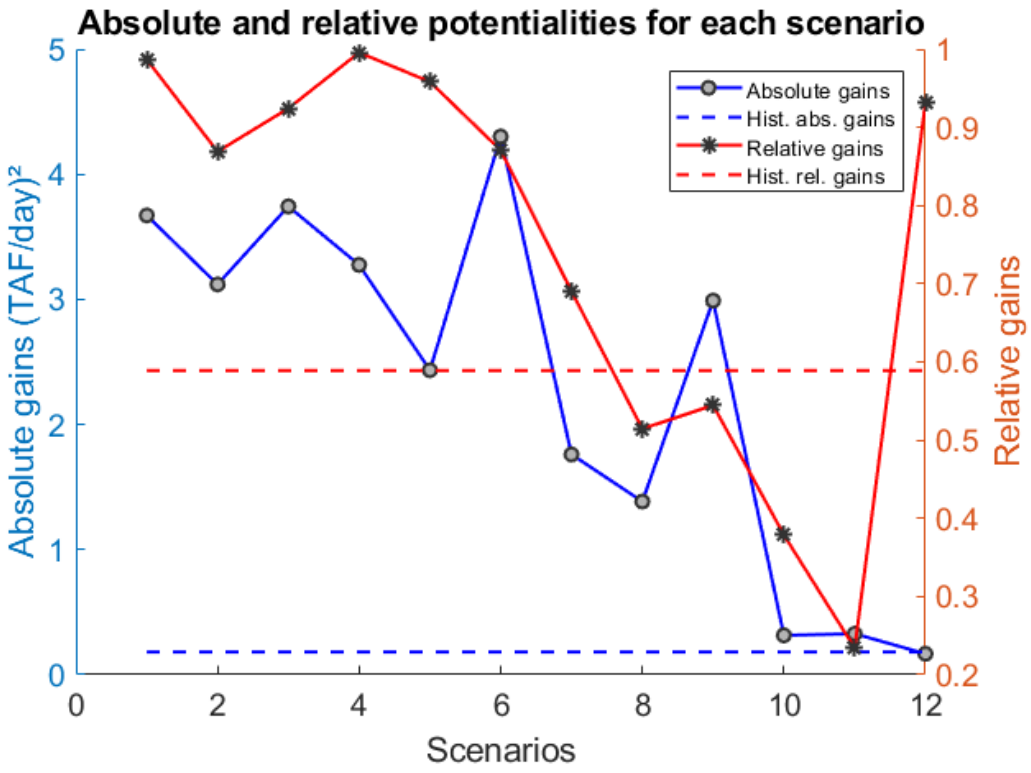

Figure 5.5: *Absolute and relative EVPI: Red colour stand for the relative gains and blue colour stands for the absolute gains; Dashed lines correspond to the historical values.*

The analysis of the remaining 14 scenarios provide interesting insights about the potential value of forecast for informing the future operations of Folsom reservoir, as illustrated in Figure 5.5. Scenarios 17 and 25 show the presence of flood for the basic operation but no flood with a perfectly informed operation.

This suggests that forecasts can in theory help prevent floods that otherwise are expected to happen (these scenarios are not included in figure 5.5 for readability reasons). Looking at the other scenarios in the figure, two main findings emerge: the first is that the absolute potential gains are not completely proportional to the relative potential gains. In fact, there are scenarios where a relative gain is high and the absolute gain is low and vice versa. Secondly, no scenario excels in both indices. To better understand the relationship between future scenarios and potential value of the forecast, the correlation between the absolute and relative EVPIs is investigated (see Table 5.5).

Index	Absolute	Relative
Mean	0.28	0.80
Variance	0.44	0.63
Min	-0.12	0.44
Max	0.09	0.14
Flood Legnth	0.36	0.75
Flood Length Seasonal	0.40	0.79
Flood Counter	0.17	0.77
Flood Counter Seasonal	0.21	0.77
Drought Length	0.18	-0.05
Drought Length Seasonal	-0.26	-0.83
Drought Counter	-0.09	-0.58
Drought Counter Seasonal	-0.13	-0.75

Table 5.5: *Correlation between the two types of performance gain and the various indices indicating different properties of the scenarios. Positive correlation for the absolute gain are variance and flood length. For the relative gains flood indices and mean are positively correlated and drought negatively correlated. Relative gains seem to be more influenced by the climate conditions than the absolute gain.*

Results suggest that there are differences in correlation between considering the absolute or relative EVPI. The absolute gain seems to be less correlated with the climatic indices. It has almost no negative correlations, except for a weak correlation with the drought and it has a medium positive correlation with the variance and the flood seasonal length, with a maximum of 0.4 of dependency. On the other hand, the relative gain seems to be more sensitive to the climatic indices. It has a strong positive correlation with the mean, 0.8, the variance, 0.6, and especially with the flood indices, where they range between 0.75 and 0.79. However, it has a strong negative correlation with the drought indicators, up to -0.83. This suggests that long spells of drought decrease the efficiency of the forecast.

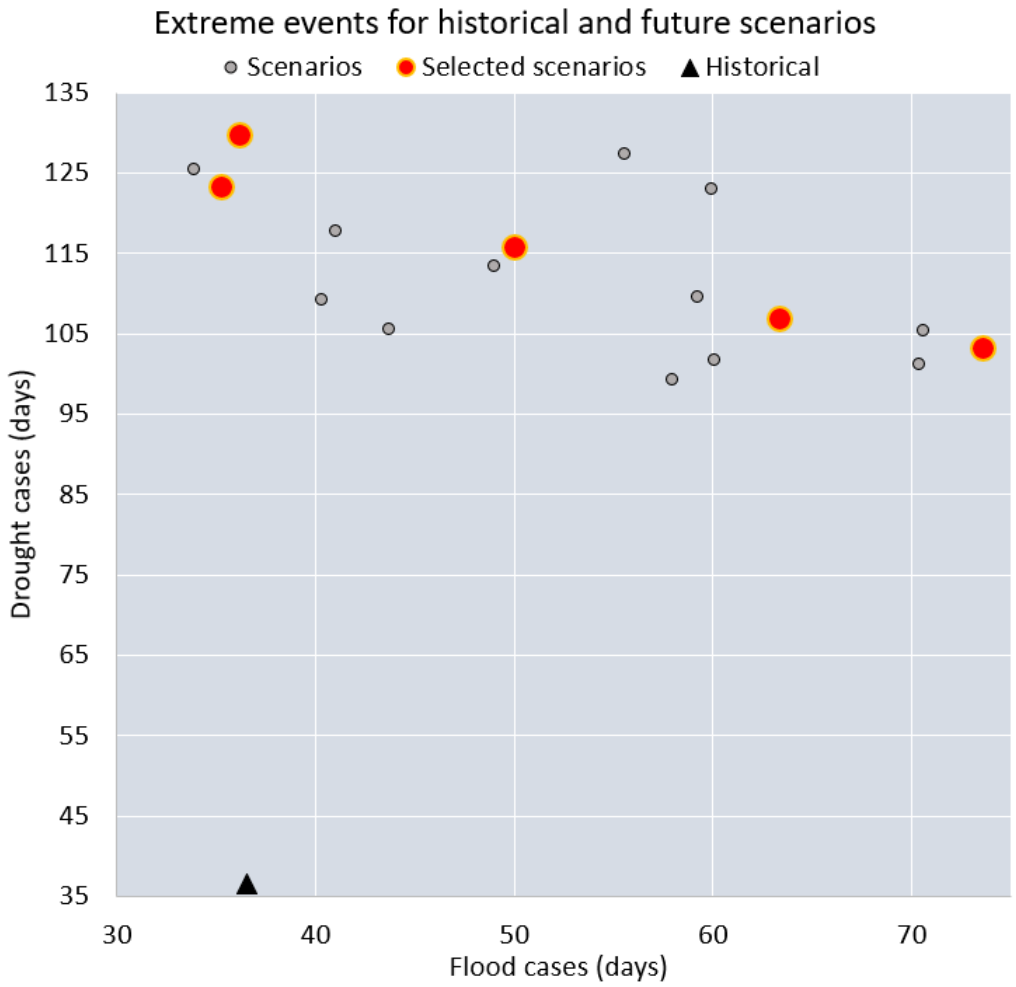

Figure 5.6: *Flood and drought occurrences for the historical scenario and for the future projections. The red dots represent the selected scenarios.*

Lastly, given these complex relationships between EVPI and the characteristics of the projected scenarios, it is more suitable to use the 14 subset of scenarios only for specific experiments that require more data. Simplifying the subset allows for better readability and still preserves the quality of the results. Because of this, the subset of 14 scenarios is processed via k-means clustering to identify three groups of similar scenarios. The clustering algorithm was run over the floods and drought indicators, so the groups could be divided according to the levels of wetness and dryness. Using this clustering analysis, three groups of scenarios were identified and classified as dry, intermediate and wet. The climate characteristics among the members of each group were averaged and the most similar scenario to the group average was selected as the representative of each group. They were named after the group they represent: dry, intermediate and wet. The dry scenario selected is the model MIROC-ESM-

CHEM based on the RCP 6.0 from the Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies; the second scenario, labelled as the intermediate one, is the model IPSL-CM5A-MR based on the RCP 8.5 from Institute Pierre-Simon Laplace and the wet scenario is the one from the Community Earth System Model Contributors, based on the model CESM1 (BGC), running on the RCP 8.5 as well. In addition to these three representative scenarios, two extra scenarios are included for some experiments, one presenting the driest scenario and the other representing the wettest scenario. The driest scenario is similar to the dry one, being developed by the same group and using the same model, but this time simulating the RCP 8.5. The wettest scenario is also based on the RCP 8.5 as simulated by MRI-CGM3 models. Figure 5.6 illustrates the characteristics of these scenarios with respect to the historical observation in terms of number of flood and dry days per year.

Scenario	Dry	Intermediate	Wet	Driest	Wettest
Mean (TAF)	6.84	9.19	11.23	7.04	11.17
Variance (TAF^2)	273.16	454.79	545.70	289.59	324.96
Min (TAF)	0.16	0.16	0.16	0.15	0.162
Max (TAF)	426.34	478.38	497.77	468.24	378.42
Flood length (days)	18.90	31.18	34.44	20.33	31.93
Flood counter (days)	35.32	50.05	63.41	36.22	73.61
Drought length (days)	83.75	73.93	69.88	78.67	71.19
Drought counter (days)	123.23	115.67	106.76	129.60	103.11

Table 5.6: *Table illustrating the properties of each selected scenario. The columns show the scenarios representing the three types of groups: dry, wet and intermediate and the two extra scenarios driest and wettest. The rows are for the main characteristics of each scenario.*

The properties of the scenarios are illustrated by Table 5.6. As expected, the driest and dry scenarios are the ones with longest spells of drought, 83 days in a row, and the largest amount of dry days, 129. The wet scenario, on the other hand, is less worrying in terms of drought, even though it still presents 106 days of drought in average per year. In terms of floods, however, it is expected to have 63 days of floods, being 34 of those during a consecutive streak, while the wettest achieves a total of 73 days of flood in a year. Finally, the intermediate scenario presents an equilibrium between the two other scenarios, with drought occurrences at 115 days per year and 50 days of flood cases per year. The variability characteristic to the intermediate scenario can also be seen in figure 5.7, in which the intermediate scenario not only shares values with both dry and wet scenarios along the century by constantly alternating between them,

but it is also the one with the largest confidence interval. The dry scenario is constantly at low inflows until the very end of the century, where it increases significantly. Lastly, the wet scenario also presents a large confidence level, but in general it stays above the 10 TAF/day threshold, with a constant increase in inflow along the century.

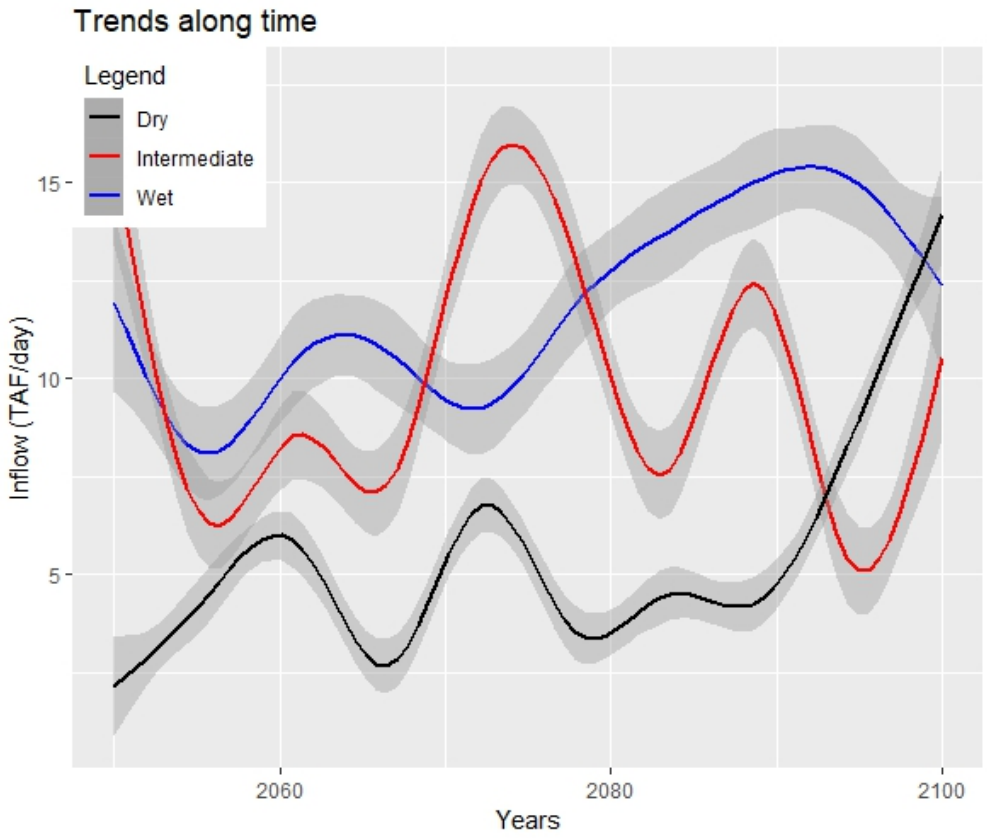

Figure 5.7: *Smoothed inflows for the entire future period, presenting trends for each scenario. The buffer around the inflow series is the confidence level of 95% for each scenario at a given time.*

5.2 Synthetic forecast generation

After the selection of the candidate scenarios, it is necessary to develop a synthetic generator that mimic the existing forecast model to produce ensemble forecast for the future scenarios. Following the methodology introduced in Section 4.2, the present section discusses the type of residual selected to be modeled, the synthetic model performance alongside its error distribution with respect to the original model and an example of the actual synthetic forecast in the future joined by a future scenario.

In order to generate synthetic forecasts, two data sets are required, observed and forecasted inflows, and the essential step here is to model the forecast residuals. It is worth recalling that this approach assumes stationarity in the forecast skill, meaning that it assumes the forecast model in the future is characterized by the same level of accuracy (i.e., same residuals) shown in the historical period. The forecast model is then calibrated over the 1980-2015 time period for which real forecast data are available. As in *Nayak et al.* (2018), the statistical properties of the synthetic forecasts are then validated over the period 1915-2015. The accuracy of the existing forecast model is shown in the scatterplot in Figure 5.8, with the trajectories of forecast residuals illustrated in Figure 5.9.

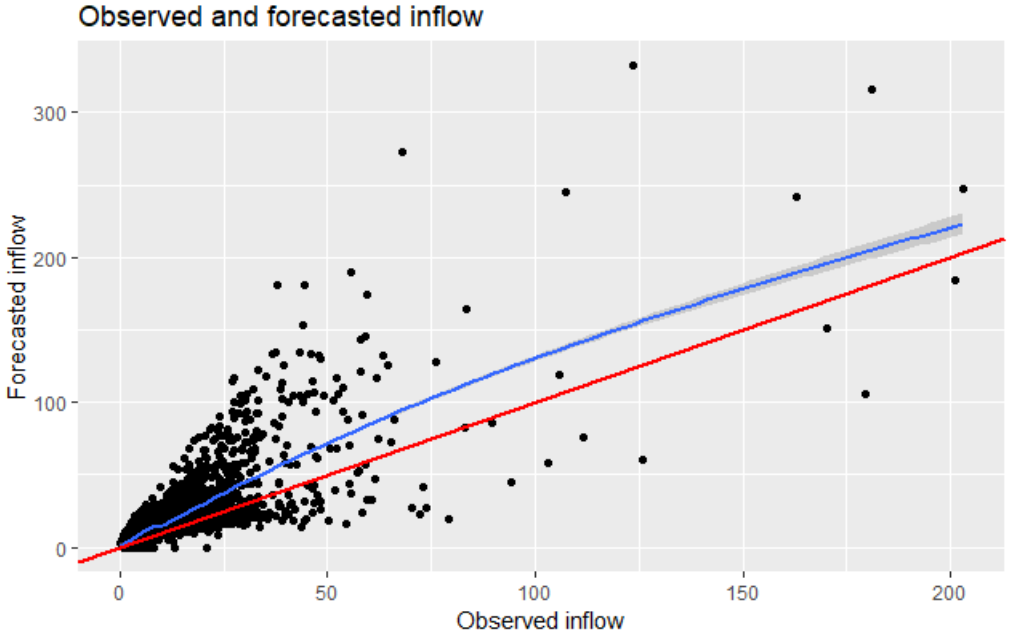

Figure 5.8: *Scatterplot between observed and forecasted inflow. Blue line is the pattern of the data distribution and the red line indicates the perfect fit between the data.*

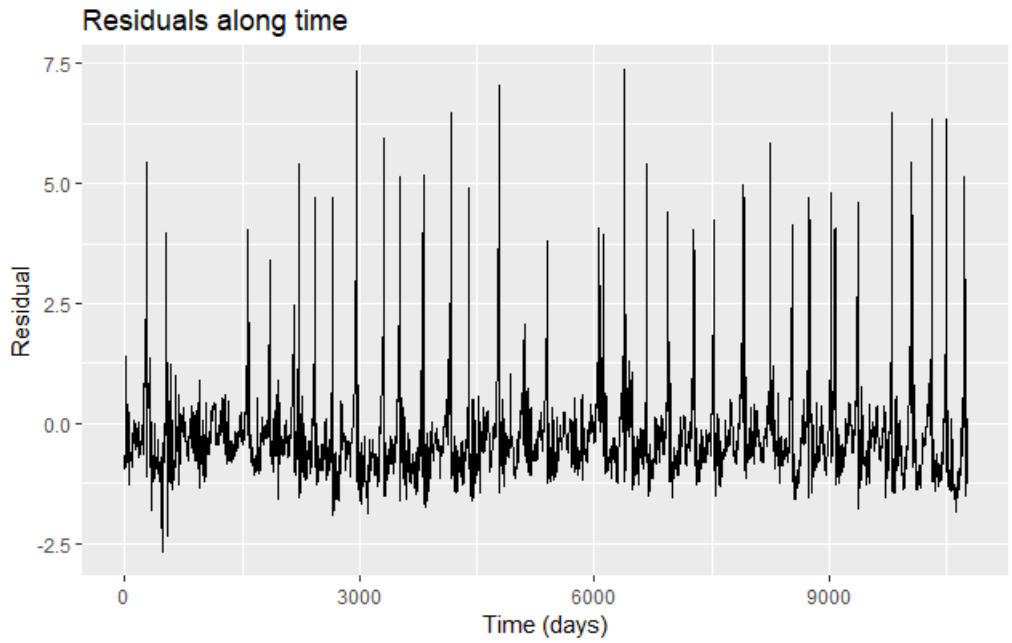

Figure 5.9: *Distribution of the forecasted inflow in relation to the observed inflow.*

The majority of the forecasts tends to overestimate the observed inflow, particularly for values up to 50 TAF/day; moreover, the greater the inflow gets, the larger the residual is. Figure 5.9 confirms these results and shows that the majority of the residuals are close to zero. Nevertheless, it is also verifiable that there is a periodic occurrence of extreme values along the years, ranging from 4 up to 7.5 TAF/day in the log format.

5.2.1 Residual testing and selection

The first step in generating the synthetic forecast is to define the nature and type of residual to be used. It can be either additive or ratio between the forecasted and observed values, and the space can be logarithmic or normal. This choice is supported by the tests reported in Figures 5.10-5.13, i.e.: histogram of the residuals, quantile-quantile (Q-Q) plot, autocorrelation (ACF) and partial autocorrelation (PACF) tests and heteroskedasticity.

The following figures present the performances of the four different types of residuals:

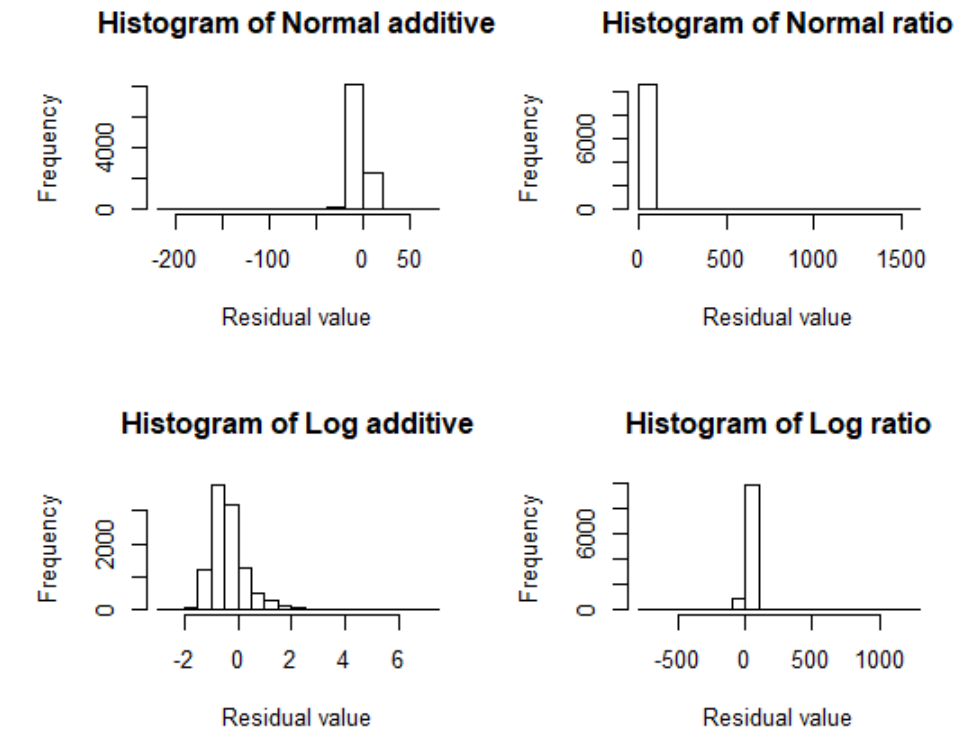

Figure 5.10: *Histogram for different types of residuals. Log additive residuals present a more uniform distribution than the others.*

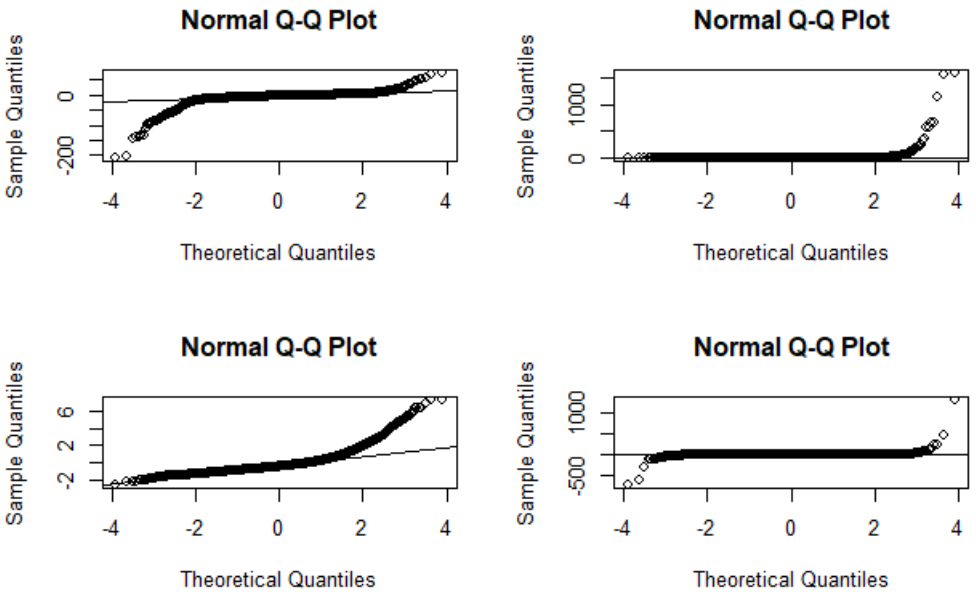

Figure 5.11: *Normal Q-Q plots for different types of residuals. Additive residuals present a more uniform distribution.*

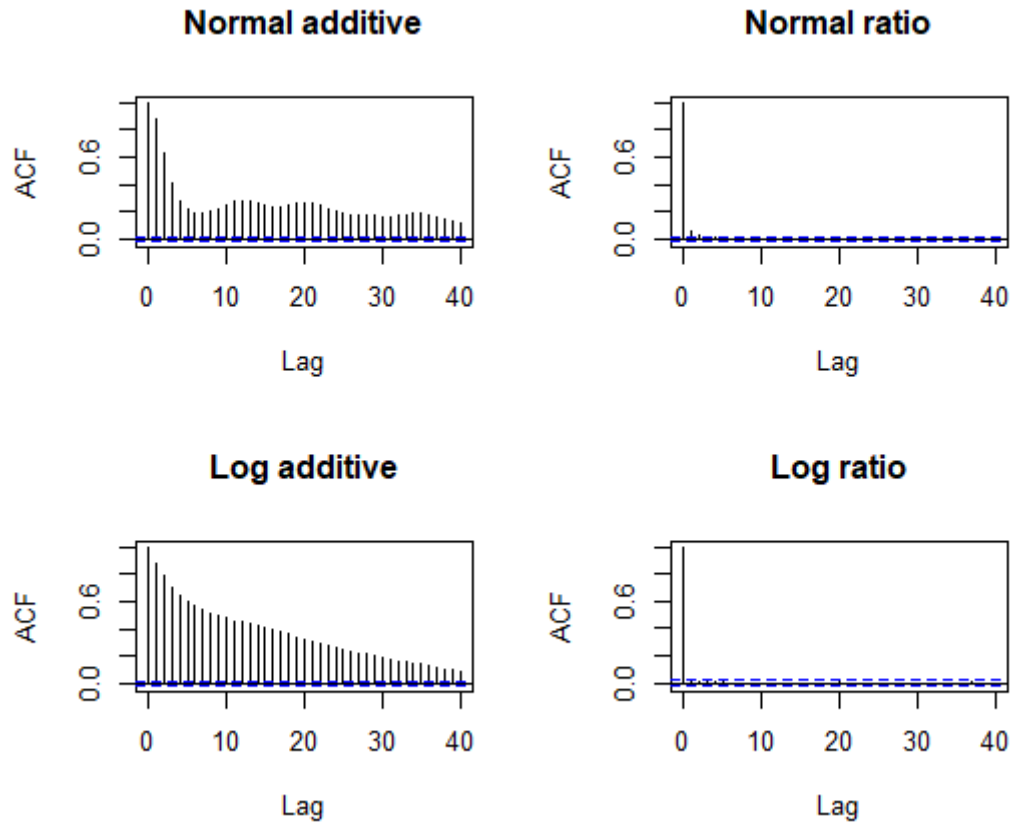

Figure 5.12: *Autocorrelation function for different types of residuals. Additive residuals show longer memory.*

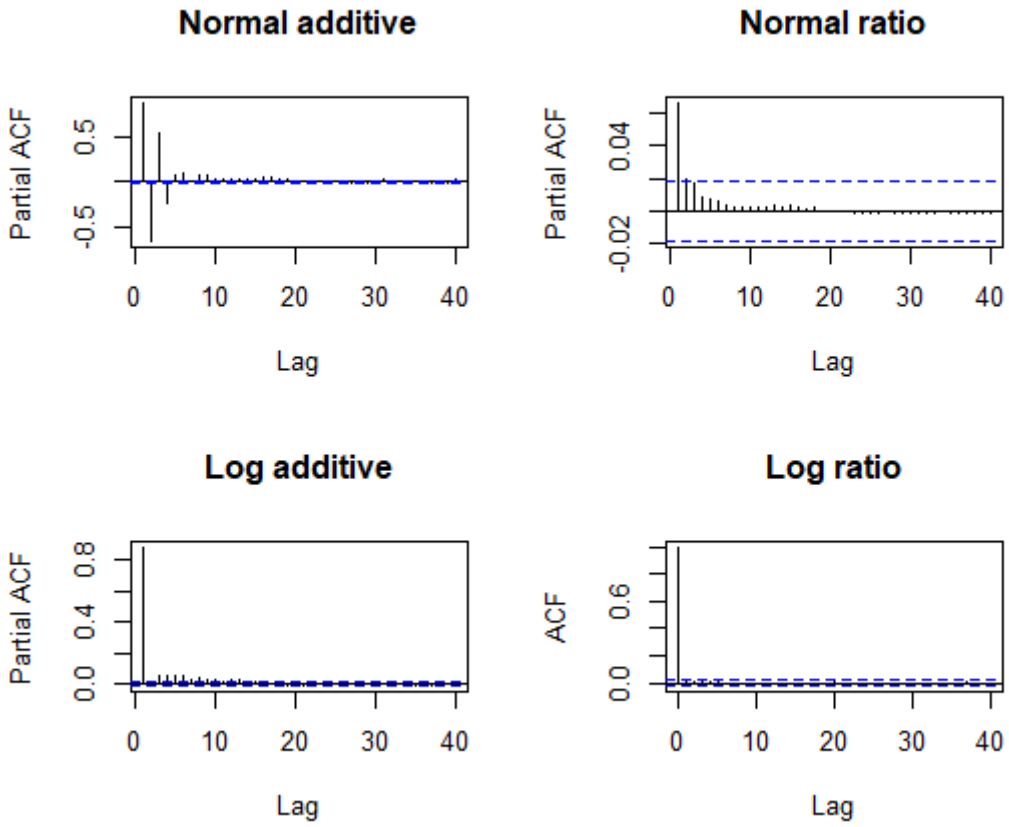

Figure 5.13: *Partial autocorrelation function for different types of residuals. Normal additive residuals are the only showing negative PACF values and longer memory.*

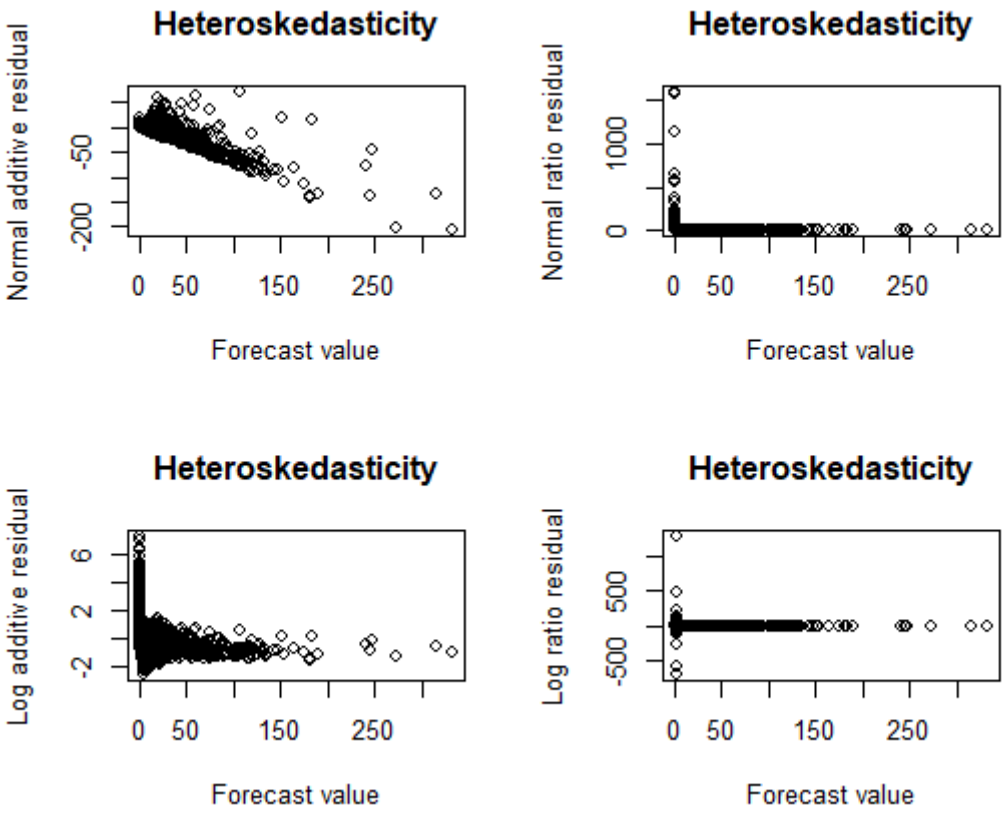

Figure 5.14: *Heteroskedasticity function for different types of residuals. Only additive residuals show heteroskedasticity.*

Figure 5.10 suggests that the only residual looking normally distributed is the additive residual in the logarithmic form, in spite of its slight skewness to the right. However, figure 5.11 shows that none of the types of residuals follows a normal distribution, but the closest to that is again the additive residual in logarithmic form. All figures present "heavy tails", indicating a considerable amount of extreme values, which are not expected in a normal distribution case.

There is a remarkable difference between additive residuals and ratio residuals when it comes to the memory. The autocorrelation, Figure 5.12, presents a long memory in both additives, while no memory for the ratio residuals. In this case, the space in which the residuals are measured did not seem to be of great impact. Conversely, the results of the PAFC test in Figure 5.13 show the space is dominating the type of results generated, with log residuals presenting a one-day memory, while normal space residuals presented longer memory. Lastly, the heteroskedasticity of the four residuals reported in figure 5.14 indicates that

only additives errors are heteroskedastic; for ratio residuals, the residuals' variation is instead independent from the magnitude of the observation.

These five tests helped in better understanding how the residuals behave and that the four types of residuals are not completely random processes. From these results, the additive residual in logarithmic form was selected for the forecast generation model (see next section) as it presents a histogram and an autocorrelation function that are easier to to reproduce synthetically.

5.2.2 Synthetic model

The synthetic forecast model generates a 30-member forecast ensemble by means of the kNN approach described in Section 4.2 working on the additive log-scale residuals. The model is calibrated over the training period 1980-2015 and validated over the 1915-2015. Four main analyses were used: Error distribution analysis, forecasted against observed values, histogram and autocorrelation functions. Figure 5.15 illustrates a scatterplot between observed and forecasted inflows, including both the real forecast and the synthetic one. It shows that the synthetic forecast successfully reproduce the skill of the real forecast. Figure 5.16 gives further evidence to the quality of the synthetic results by showing that the synthetic forecasts are again similar to the observed ones in terms of residual-observation relationship. Both sets of data are distributed according to a triangular shape with the majority of the errors being close to zero. The histogram presented in figure 5.17 also shows some similarities between real and synthetic forecasts.

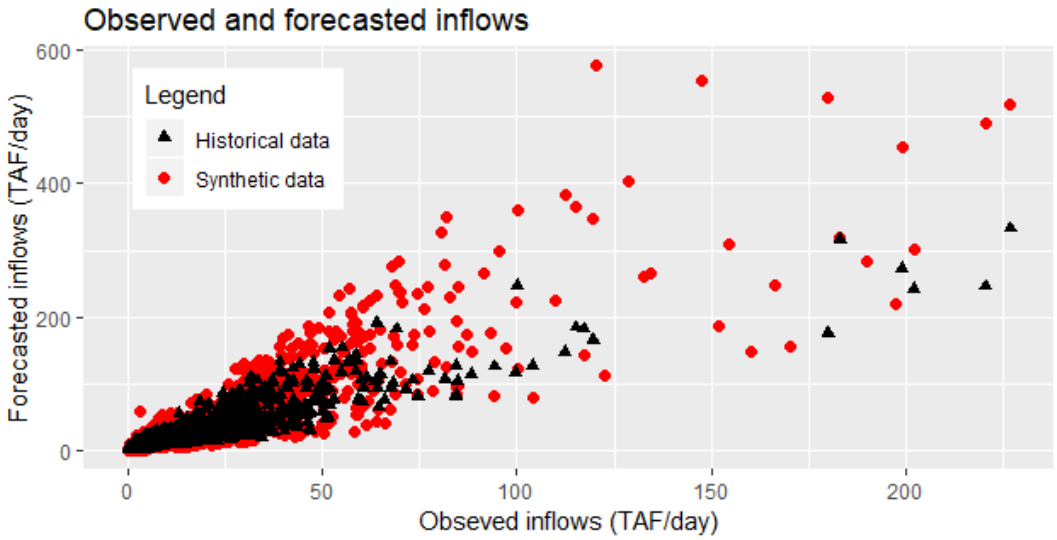

Figure 5.15: *Scatterplot presenting the forecasted inflow based on the observed inflow. Black triangles represent the historical data and the red circles the synthetic generated.*

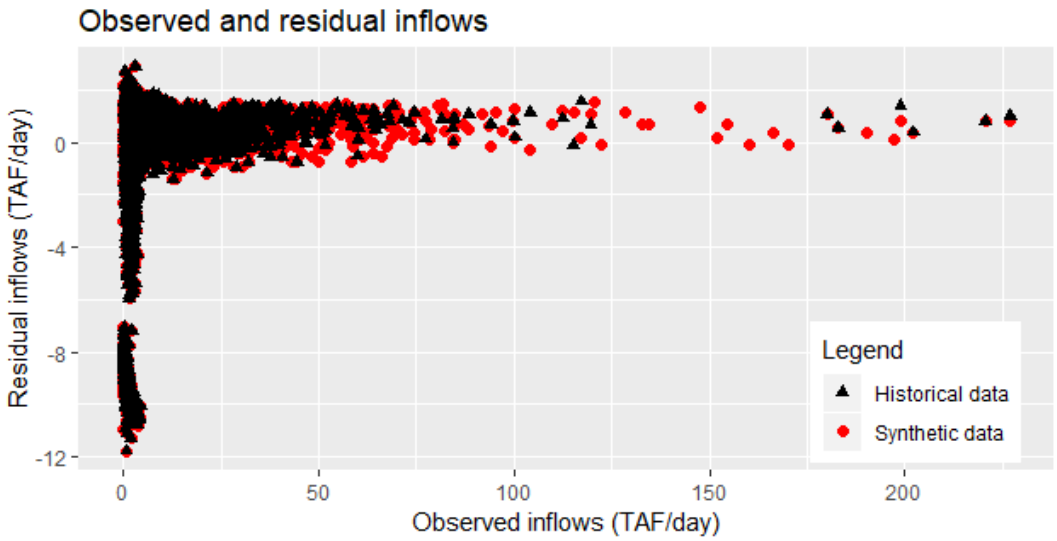

Figure 5.16: *Relationship between the residual inflow and the observed inflow. Black triangles represent the historical data and the red circles the synthetic generated.*

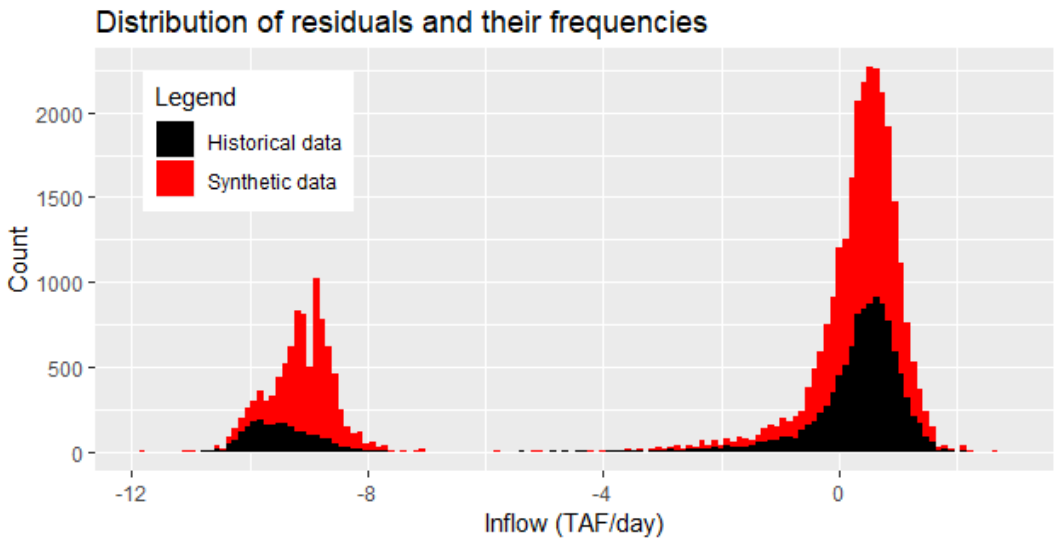

Figure 5.17: *Histogram with magnitude and frequencies of the residuals for both historical and synthetic cases. Black represents the historical data and red the synthetic generated.*

The autocorrelation functions of the synthetic forecast reported in Figure 5.18, however, show worse performance. The original residuals during summer had a long memory, up to 40 days, with a linear shape descending from 0.9 for the 1-day lag up to 0.1 for the 40-day lag. Also, the winter ACF had a similar shape to the summer one. The synthetic model reproduces instead a weaker ACF for the summer period. The same bias affects the winter period which exhibits weaker memory with respect to the case of original forecasts. In general, these differences in memory can however be considered acceptable, because the autocorrelation values are anyway above the statistical significance for the majority of the days.

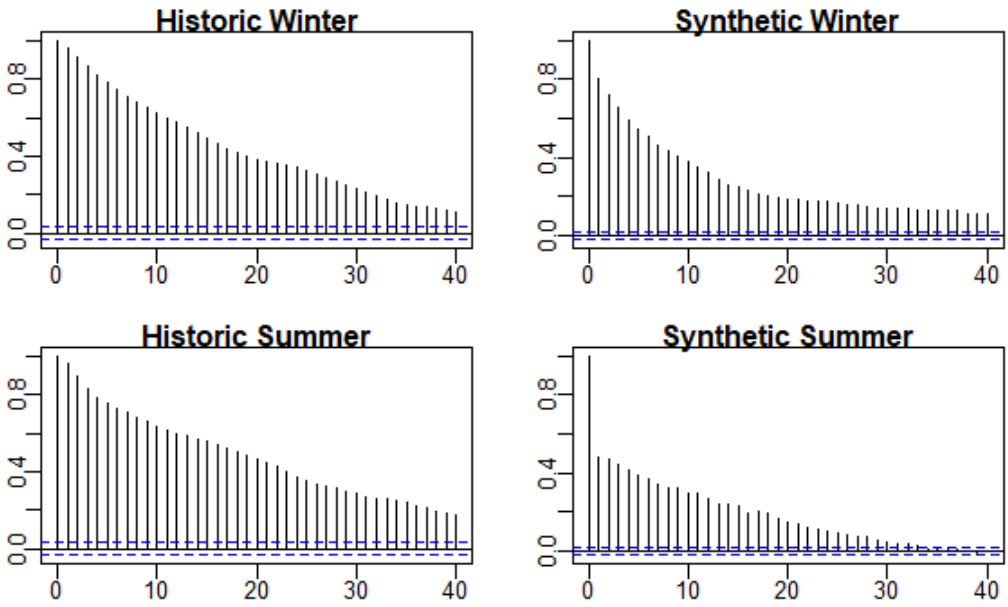

Figure 5.18: *Autocorrelation functions for observed residuals and synthetic residuals for both summer and winter seasons.*

When considering the main properties of the synthetic model, the analyses previously done allow for a general assessment of the model. In three out of four tests the synthetic properties present a high fidelity reproduction of the historical properties and only one, the autocorrelation of the residuals, was not as high as the original one. Nevertheless, its performance is deemed satisfactory, as, even though not with the same intensity, the memory is still accounted for the entire period. The model was then applied to each scenario of projected inflows in order to generate a 30-members ensemble of projected synthetic forecast for the time period 2070-2099.

5.3 Forecast value

The final section of the results is designed to provide answers to the main scientific questions of this work by quantifying the historical and future value of inflow forecasts for the operations of Folsom reservoir. The next step is to explore the behaviour of the system when ruled by different versions of the policy tree, such as the differences in storage dynamics, release decisions and in which ways the forecasts can contribute to a change in the operation. Lastly, the results of the water operations are displayed for different scenarios, time periods and models of policy tree, followed by multiple comparisons and analyses

with the purpose of investigating exhaustively the impacts and contributions of forecast models in future conditions. For more information on the validation of the results obtained by the policy tree, see annex A.

5.3.1 Historical and future forecast values

Firstly, a summary of the operating costs across different scenarios and operating policies is shown in figure 5.19, where all the scenarios are shown in the vertical axis referring to both future and historical time periods. The horizontal axis indicates the cost function as defined in eq. 3.2. The blue bar represents the perfect operating policy designed via DDP under the assumption of perfect knowledge of the future; the black bar indicates the baseline operation, without any use of forecast; and the red dot represents the operation informed by the synthetic forecast. Additionally, the black X indicates when there is an event of flooding. Results show that for all scenarios, the forecast-based operation reduced the cost in comparison with the baseline. Second, operations in general are going to have their costs increased for every future scenario when compared to their historical performance. It is interesting to observe that, for the dry scenario, the baseline operations attains a null cost over the historical period, meaning the water supply always met the demand also controlling any potential flood event; the same operations simulated over the future time period attains a cost exceeding 2.5 $(TAF/d)^2$. For the intermediate case, even though the historical period already presents costs in all operating policies, with almost no value generated by the use of forecast information, all of them get increased in the future period over which forecast allows reducing the cost of the baseline solution. Lastly, the wet scenario indicates small costs for all policies over the historical periods; in the wet future, however, the baseline operations becomes incapable of controlling floods, with forecast information becoming extremely valuable as they allow avoiding such extreme events.

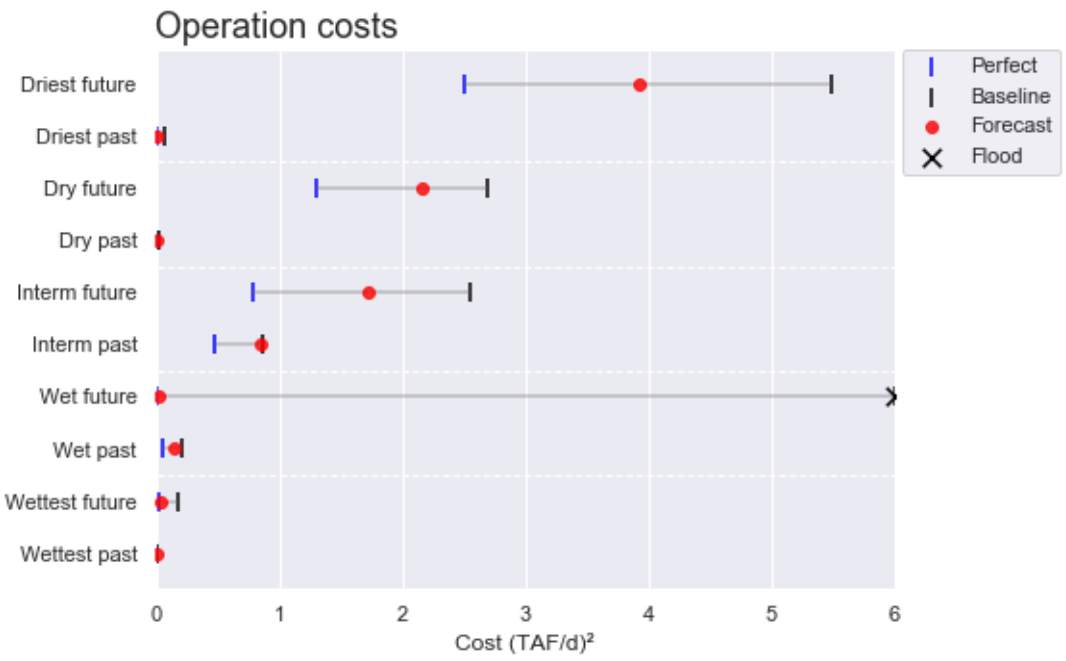

Figure 5.19: *Operating costs for different scenarios and time periods. The black bar represents the baseline operation without forecast, the blue bar stands for the perfectly-informed operation and the red dot shows the actual operation with the use of forecast. The black X represents scenarios that have flood events and penalties linked to it.*

The results in Figure 5.19 are then illustrated in Figure 5.20 adopting a normalization of the cost where the performance of the baseline performance is set equal to 1. The performance improvement attained by the perfect operating policies and the one informed by the synthetic forecasts is then shown as a fraction of avoided costs. The dry scenario over the historical period shows that the use of forecast information provides almost the same reduction in cost as the perfect case, suggesting that the existing forecast already provides all the information needed to perfectly operate the system. In future periods, however, the operating costs increase and this means the performance of the forecast-informed solution is lower than the perfect operations, suggesting the potential need of more accurate forecast to face the projected inflows under this scenario. For the intermediate case, the forecast operation is closer to the perfect operation in the future period than it is in the historical one, indicating that forecast value is expected to increase over time. This is also true for the wet scenario, where the future period presents a perfect and a forecast-based operations that are approximately zero percent of the baseline cost which is driven by the occurrences of flood events, whilst the historical period shows a less relevant contribution by the forecast.

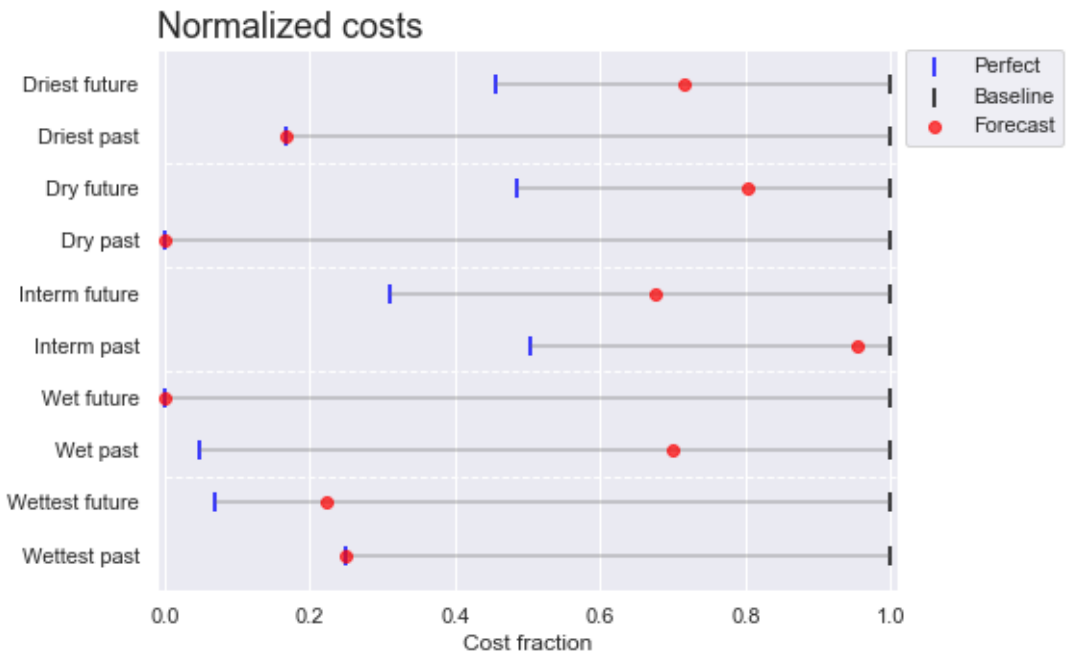

Figure 5.20: *Costs reductions and fractions in a normalized space for the different scenarios. The black bar represents the baseline operation cost without forecast as 1.0, the blue bar indicates the maximum reduction possible with a perfectly-informed operation and the red dot shows the actual reduction of the operation cost when using forecast.*

The direct comparison between historical and future forecast value is reported in Figure 5.21. The black squares represent historical absolute forecast value and the red triangles represent the forecast value in the future. The figure shows that the gain is expected to increase for every single scenario, no matter the conditions predicted by the different scenarios. As explained in the previous paragraphs, the wet scenario shows a occurrence of flood for the baseline scenario, which makes the actual absolute gain way larger than it is demonstrated in the figure, but for the sake of readability, it was decided to represent this fact with a sign and preserve the proportions for the other scenarios.

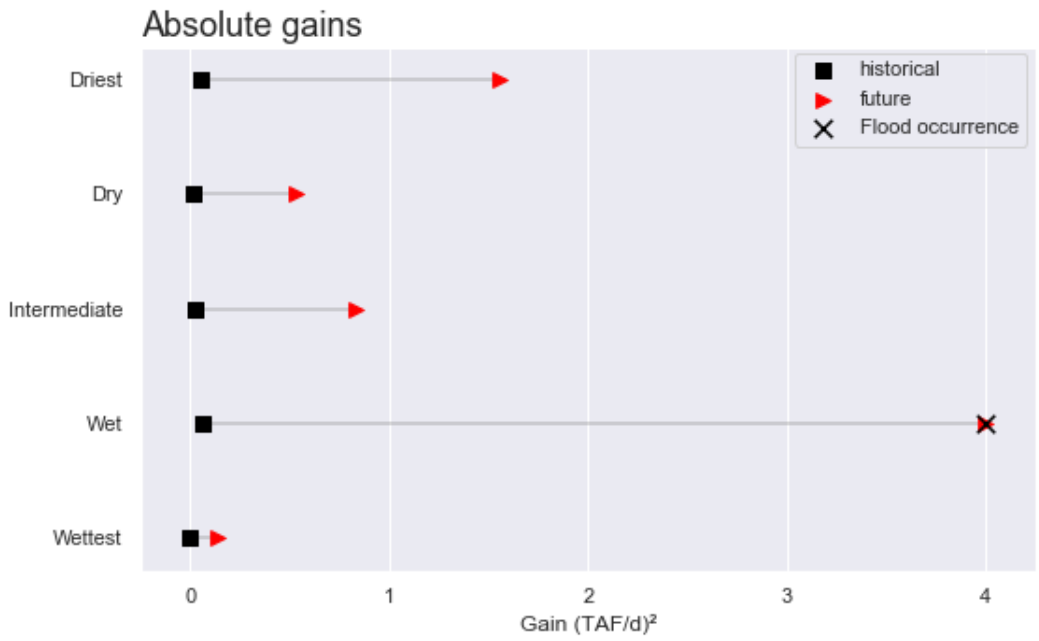

Figure 5.21: *Forecast gain for both historical and future time periods in absolute terms, $(TAF/day)^2$. Black square stands for the historical forecast improvement with respect to a baseline operation, red triangle stands for the operation in a future time period. Note that for the sake of readability, the gain of future wet scenario which allows avoiding the flood event is rescaled and marked with a black cross.*

In addition to these absolute gains, figure 5.22 presents the results for the relative gains. While the black square and red triangle still represent the historical and future forecast value, respectively, the figure reports relative values, ranging from 0 to 1, being the reference level the performance attained by the baseline policy without forecast. Differently from the absolute gain, there is no clear trend or pattern among the scenarios, with the intermediate and wet scenarios presenting higher forecast value in the future than in the past, while dry scenarios show a loss of gain with time.

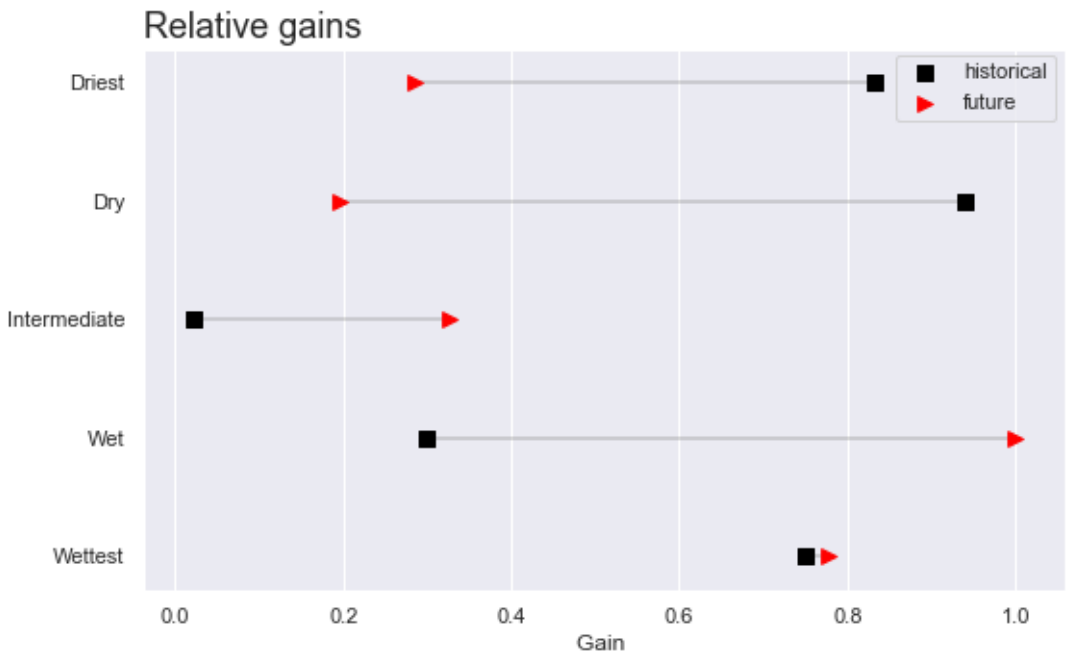

Figure 5.22: *Forecast gain for both historical and future time periods in relative terms, based on the baseline operation without forecast. Black square stands for the historical forecast improvement with respect to a baseline operation and red triangle stands for the operation in a future time period.*

One last experiment is taken, where the fraction of the maximum performance improvement (with perfect knowledge of the future) covered by using the synthetic forecast is analyzed. Figure 5.23 shows the fraction of the gain generated by the forecast for each scenario in both historical and future periods, where 0 means there is no gain with respect to the baseline operation and 1 means the forecast operation is performing as the perfect operation. Results seem to confirm the findings discussed on the relative forecast values varying with the underlying climate scenarios, with the two dry scenarios indicating a decreasing forecast value while the intermediate and the wet future scenarios present an increasing trend over time.

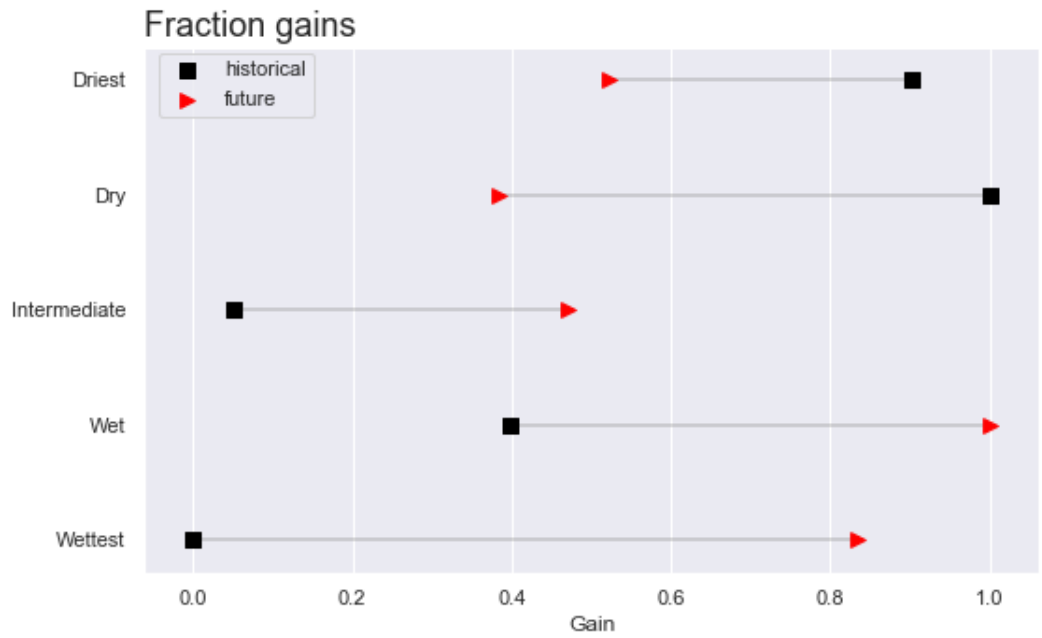

Figure 5.23: *Temporal evolution of the forecast gain between a historical time period and a future time period. The red triangle stands for the gain of forecast in the future, while the black square represents its gain in a historical time period. If the red triangle is closer to one than the black square, it means the value increased in time, otherwise, it lost value.*

These results and the way they are conditioned, just like previously seen with the relative case, suggest an analysis involving more scenarios could give additional contributions to the study. Therefore, all the 14 scenarios obtained in Sub-section 5.1.2 are used for the experiment shown in Figure 5.24, where the gains for every single scenario are measured and the scenarios are ranked from driest to wettest. This figure confirms the hypothesis that the increasing/decreasing trend in forecast value over time is linked to the characteristics of the considered climate projections. On the lower part of the figure, which encompasses the driest scenarios, the forecast value decreases with time. Then, in the middle section, which holds the intermediate scenarios, gains and losses are observed; and, finally, in the upper section with the wettest scenarios, an increasing forecast value is observed. Wet scenarios are therefore expected to be associated with increasing forecast value in the future, whereas dry scenarios are expected to show decreasing values.

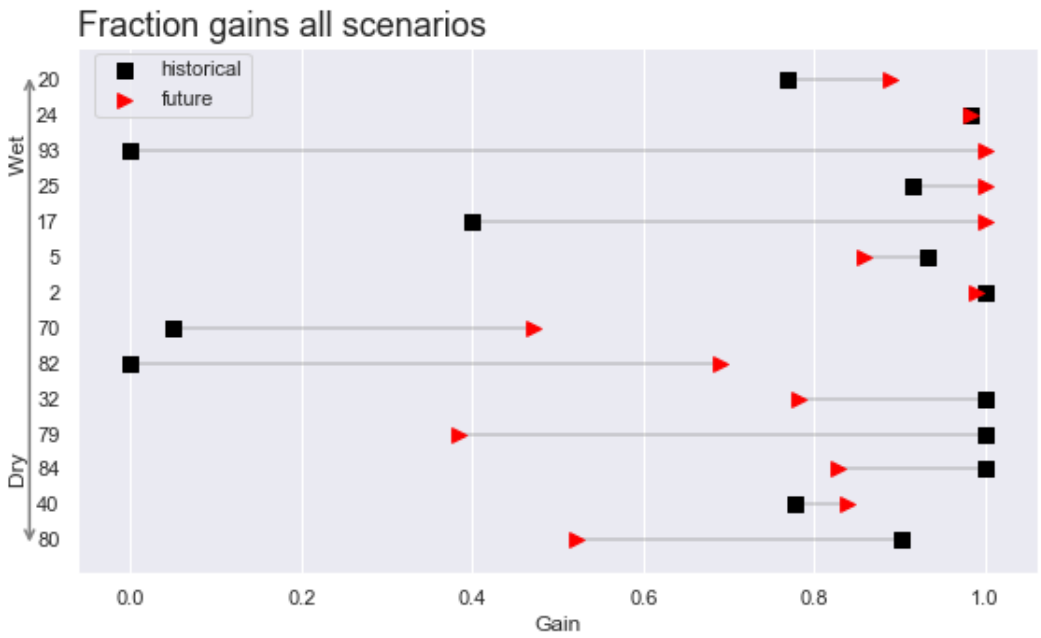

Figure 5.24: *Assessment of the forecast gain for different scenarios along time. The red triangle stands for the gain of forecast in the future, while the black square represents its gain in a historical time period. If the red triangle is closer to one than the black square, it means the value increased in time, otherwise, it lost value.*

To provide further support to these evidences, Figure 5.25 (a) presents the temporal gain or loss of forecast value according to the number of flood days (blue triangles) and to the number of drought days (red circles) for the different climate scenarios. Despite the large variability probably due to the limited number of scenarios, the figure shows that there is a positive correlation between the forecast value and the quantity of flood, and there is a negative correlation when taking the number of drought days. Figure 5.25 (b) reports a scatter plot between the temporal deviation of the drought days (vertical axis) and flood days (horizontal axis) over future and historical periods, respectively, with the color of the circles representing the trend of forecast value over time. The scatterplot confirms the two main patterns obtained so far: the greater the increase in the frequency of drought, the the more negative is the forecast value change in the future, and the wetter the scenarios get, the higher is the increase in the future forecast value.

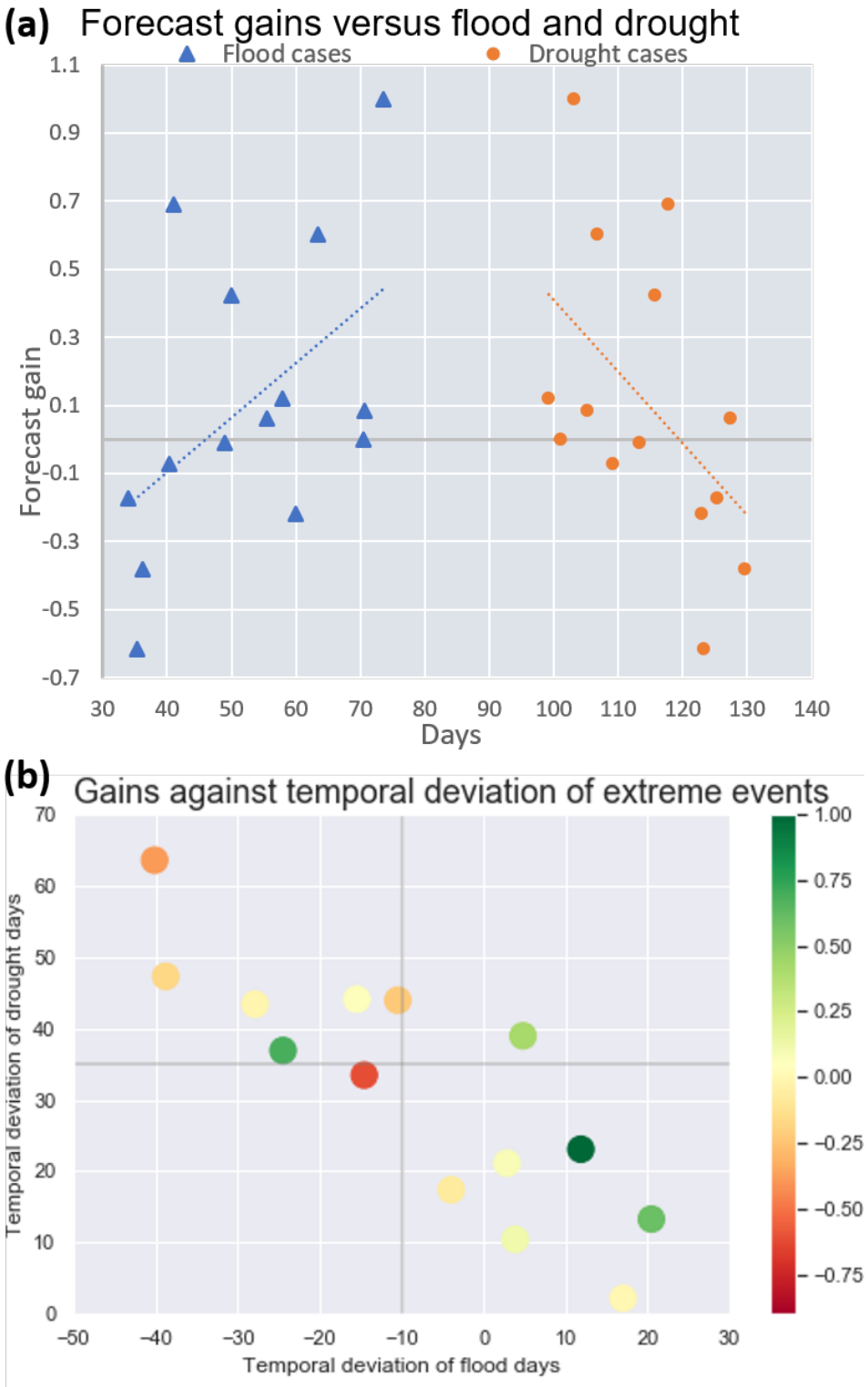

Figure 5.25: *(a) Analysis of the forecast gains when compared to extreme events such as floods and droughts. Blue triangles stand for the number of days in a year of floods and red circles represent the days in a year with a drought. The two lines stand for the linear regression of the two data sets, blue for the flood cases and red for the drought cases. (b) illustrates the forecast gain in time, varying in color, with the temporal deviation of the two types of events, flood and droughts, between the past and future time periods.*

The pattern observed in this work for the fraction and relative forecast values are likely linked to the type of forecast model used. A short-term forecast model with a 3-day leadtime is designed to anticipate quick and intense events, such as floods, but is not ideal for long-term operations. Therefore, a future scenario that presents a higher level of water inflow offers more possibilities for the forecast model to operate, either by hedging water or preventing floods. Conversely, for scenarios that are expected to become drier with time and the drought spells to get longer, the 3-day leadtime forecast system becomes less efficient in improving the water management, since it is incapable of foreseeing far into the future and to act accordingly. It needs to be seen, however, the contributions of a long-term forecast to the future dry scenarios, as it is the appropriate type of forecast for dealing with long lengths of drought and improving water supply over a longer time period.

5.3.2 Dynamics of the forecast

Given the results illustrated in the previous section showing the benefit of using forecast for informing the operating policy, it is crucial to understand how the decisions are influenced by the forecast information. The first analysis focuses on the wet season dynamics in the reservoir. During this season, there is a considerable amount of water entering into the reservoir, usually higher than the amount demanded downstream, and the main concern is to prevent floods. Figure 5.26 compares storage and release trajectories for different policies over the 2080-2100 time period under the dry scenario. Results in panel (a) show that the forecast operation allows for higher levels of water to be stored than under the baseline solution. The forecast-based operation, which is not constrained by the flood control pool, is able to store more water along the year without facing any flood event through the entire period of time, thus improving the amount of water available for water supply during the rest of the year. Figure 5.26 (b) presents the corresponding release trajectories, which confirm the superiority of the forecast-based operation in satisfying the downstream water demand. Finally, Figure 5.26 (c) compares the storage dynamics of policies informed by perfect and synthetic forecasts. By using a perfectly-informed prediction, during the wet periods, the level of water is similarly managed to the actual forecast-based operations.

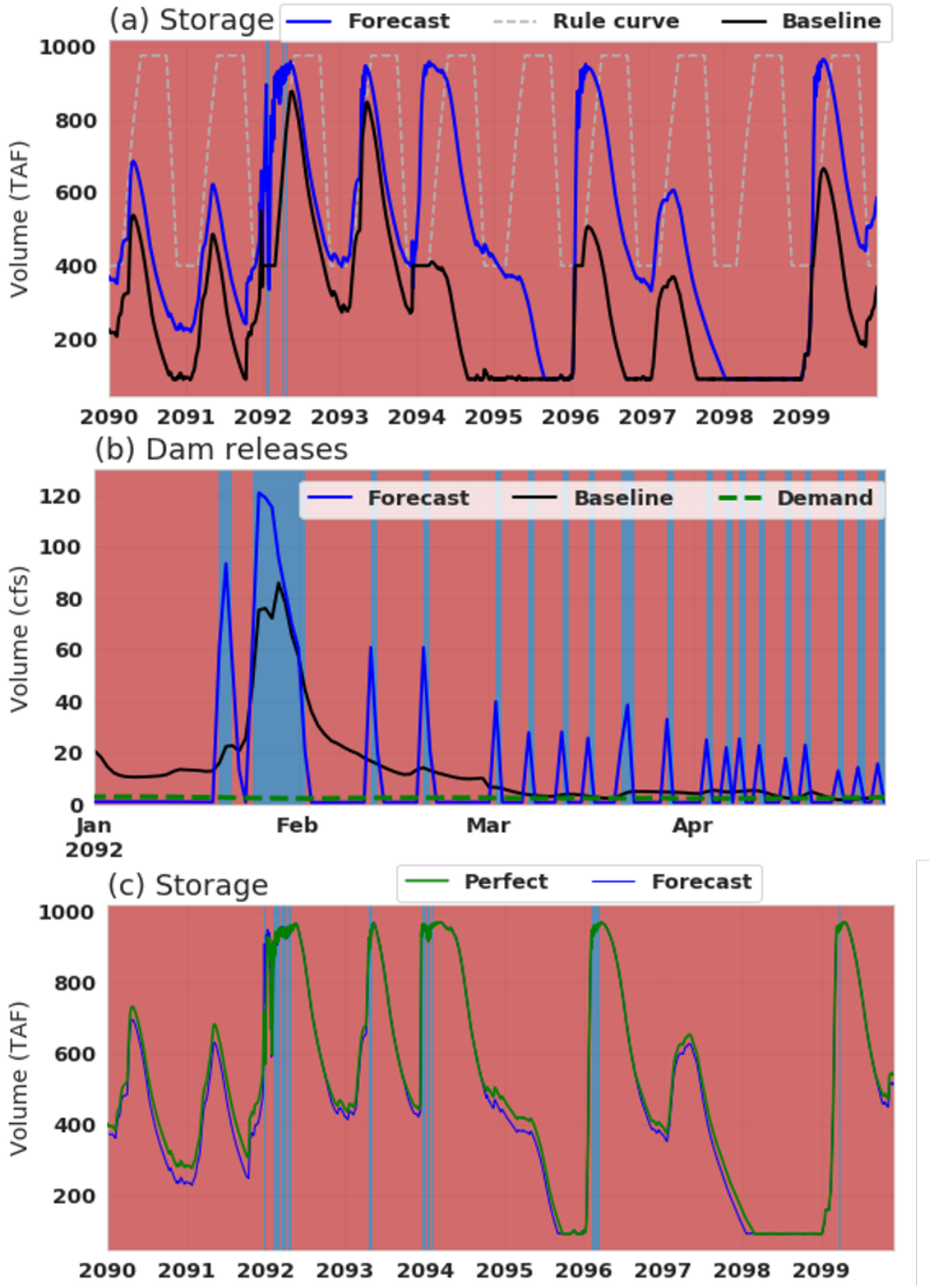

Figure 5.26: *Storage levels for a dry scenario in the future. (a) Comparison between the storage levels managed by the baseline operation, the black line, and the forecast operation, blue line. Rule curve stands for a static ruling policy; (b) Release scheme during a wet season, being the dashed line the water demand; (c) Comparison between the storage levels of the actual forecast (blue line) and the perfect forecast operation (green line). The background colors represent the type of action taken by the policy: red for hedging, blue for releasing excess and light yellow for releasing the demand.*

Figure 5.27 analyzes the role of forecast under a wet scenario, under which the baseline solution generates some flood events (note the change in action in panel b, by the change of colours from release demand, light yellow background, to release excess of water, the blue background) despite the presence of the flood control pool. On the contrary, the forecast-based solution is able to maintain higher average storage values (panel a) and, at the same time, anticipate the inflow peak and successfully control the flood (panels b-c). When considering the perfect forecast, Figure 5.27 (d), the two trajectories follow a similar dynamics, but the perfect one manages to keep the level of storage at its full capacity for longer periods than the synthetic forecast-based operation.

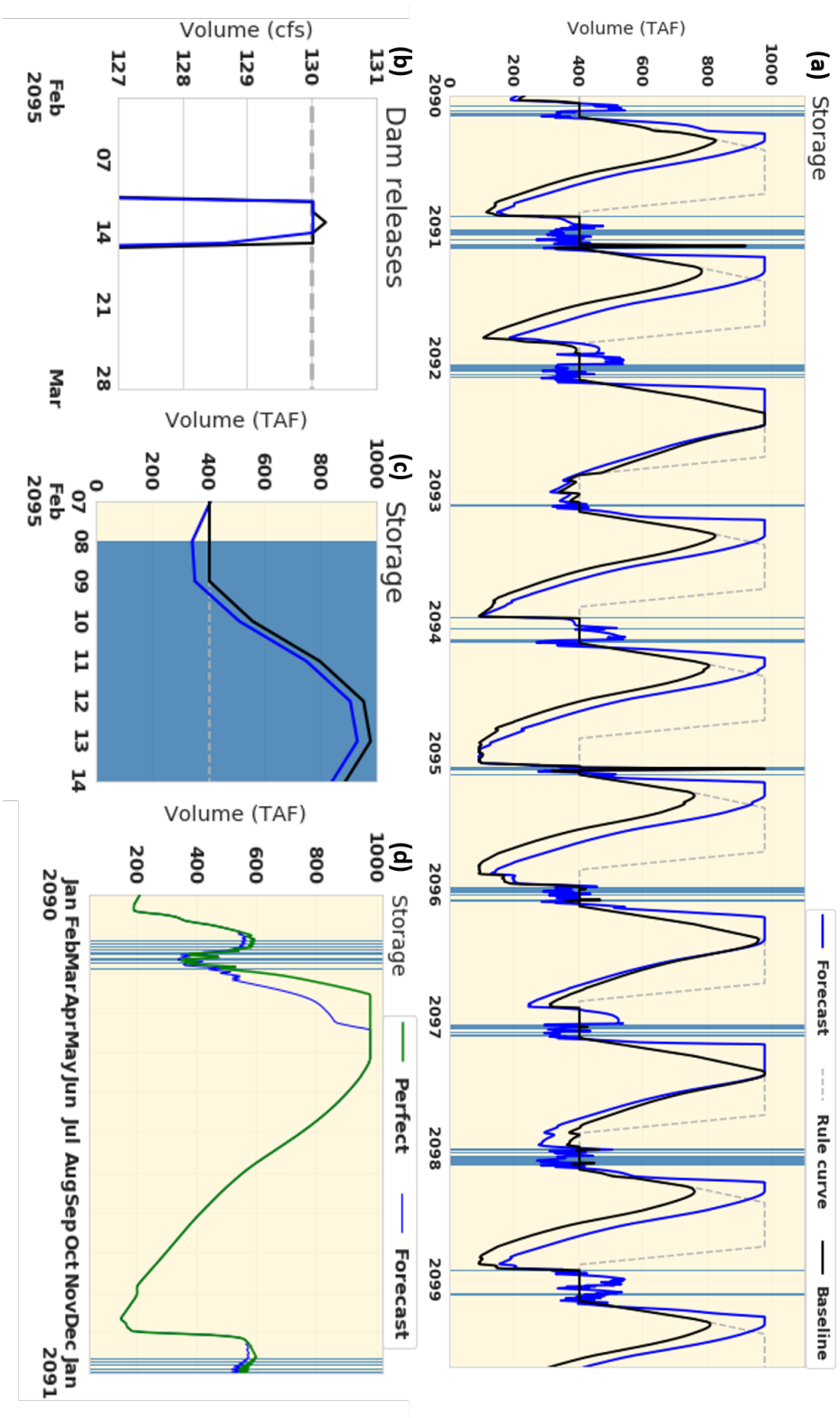

Figure 5.27: *(a) Storage levels in a wet scenario managed by the baseline operation, black line, and the forecast-based operation, blue line. (b) and (c) show the moment when a flood happens. (d) Comparison between the synthetic forecast operation (blue line) and the perfect forecast operation (green line). The background colors represent the type of action taken by the policy: blue for releasing excess and light yellow for releasing the demand.*

During the dry season, the system dynamics are slightly different. The water demand is often higher than the water available and the risk of flooding is minimal. The main difference between baseline and forecast-based solution relies in the more effective hedging strategy when forecasts are used. Forecast indeed allows storing more water during the wet season (Figure 5.28a) and thus providing a more continuous water supply than using the baseline solution (Figure 5.28b). When taking into account the perfect forecast (figure 5.28c), it is seen that not much could be changed in comparison with the actual forecast, since during the dry season there is limited inflow and the accuracy of the forecast is of less importance.

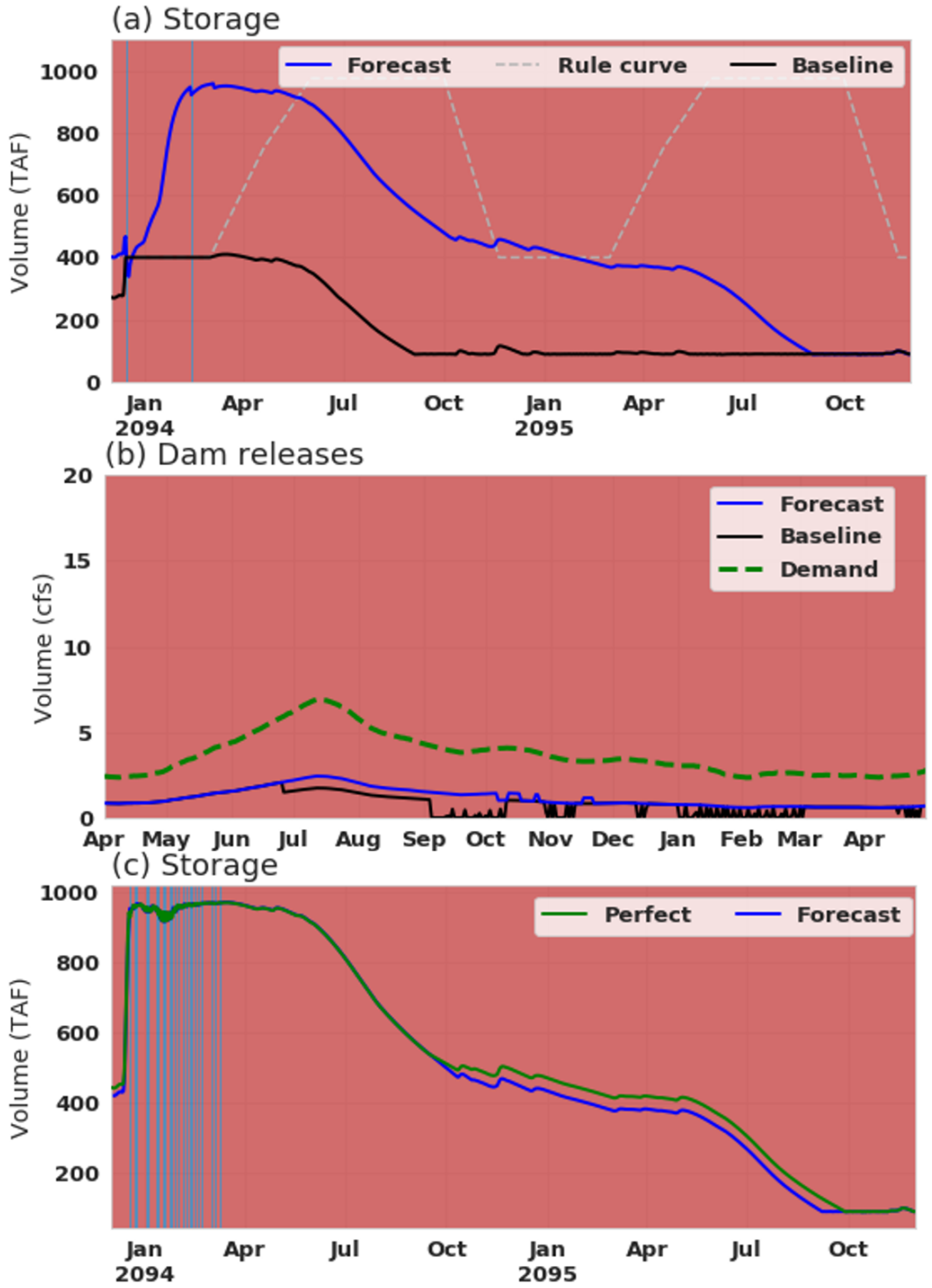

Figure 5.28: *(a) Storage levels during a dry season in a dry scenario for both baseline (black line) and forecast-based (blue) operations, with the latter storing more water than the former. (b) Corresponding period of time for release of water by the dam in both operations and the demand curve of water for the same period (dashed line). (c) Comparison between the storage management for both operations using an actual forecast (blue line) and a perfect forecast (green line). The background colors represent the type of action taken by the policy: red for hedging, blue for releasing excess and light yellow for releasing the demand.*

Similar results are obtained for the dry season of the wet scenario. Higher levels of storage for forecast-based operations during the wet season (figure 5.29a) allow a more effective hedging that reduces the water supply costs. In this case, the flexibility of the forecast-based solution that is not constrained by the flood control pool plays an important role in allowing to store large volumes of water. Furthermore, as figure 5.29 (b) indicates, after the month of September, the baseline operation starts to decrease the volume of water released until it reaches zero cfs for some days during the months of November and December. Conversely, the forecast-based operation keeps supplying a constant volume of water. Similarly to the previous case, the perfect operation looks similar to the one informed by the synthetic forecasts (Figure 5.29c), confirming that the extremely limited water availability becomes the limiting factor while the accuracy of the forecasts play a minor role.

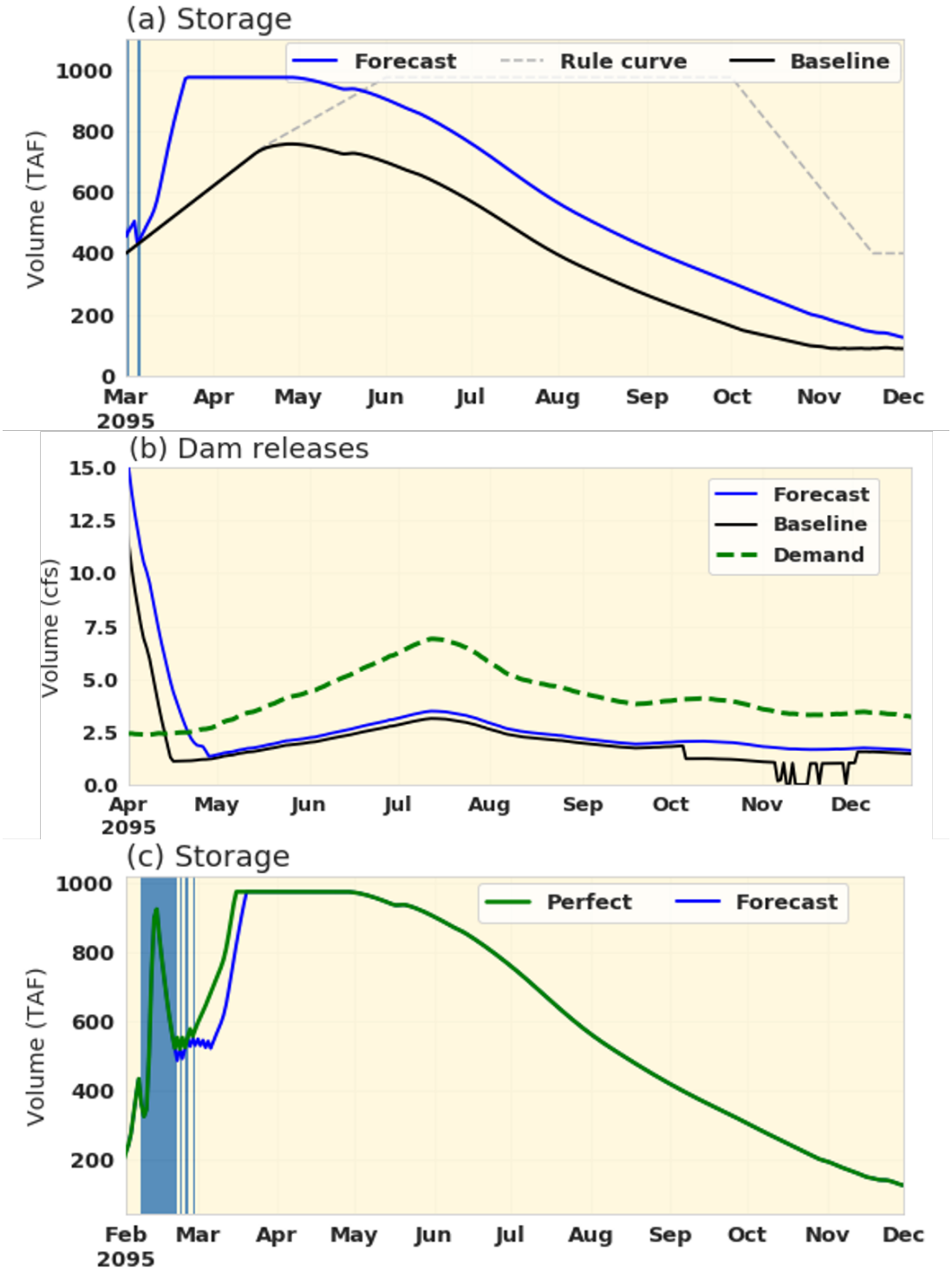

Figure 5.29: *(a) Storage levels in a dry season of a wet scenario for both baseline (black line) and forecast-based (blue) operations, with the latter storing more water than the former. (b) Corresponding period of time for release of water by the dam in both operations and the demand curve of water for the same period (dashed line). (c) Comparison between the storage management for both operations using an actual forecast (blue line) and a perfect forecast (green line). The background colors represent the type of action taken by the policy: red for hedging, blue for releasing excess and light yellow for releasing the demand.*

Summarizing, the forecast proves to be the most suitable for wet conditions and this is valid for both the wet scenarios and wet seasons, as the larger the amount of water inflow, the higher the potential of hedging strategies for reducing the water supply costs. During really dry conditions, such as a long dry season in a dry scenario, the value of forecast is of little use, given the complete lack of water incoming. At the same time, the forecast proved useful in avoiding flood events, because it has the capacity of anticipating the inflow and releasing water beforehand.

5.3.3 Robustness and adaptation

The entire analysis conducted so far designs a set of basic, perfect, and forecast-based policies over a historical and a future horizon, assuming the adaptation of the operating policy to each climate condition is a straightfoward adaptation option. However, given the future uncertainty of the climate scenarios, there is a genuine interest in understanding if the adaption of past policies into the future is possible and if a policy can account for more than one future scenario. Figure 5.30 illustrates these cases, by showing the change in the operating policies between a historical operation (a) and a future operation (b) for the same scenario and the diverse operating strategies for different scenarios (c) and (d) over the same future period. Therefore, in this section two additional potential benefits of forecast information are discussed, looking at its contribution in designing adaptive and robust operating policies, respectively.

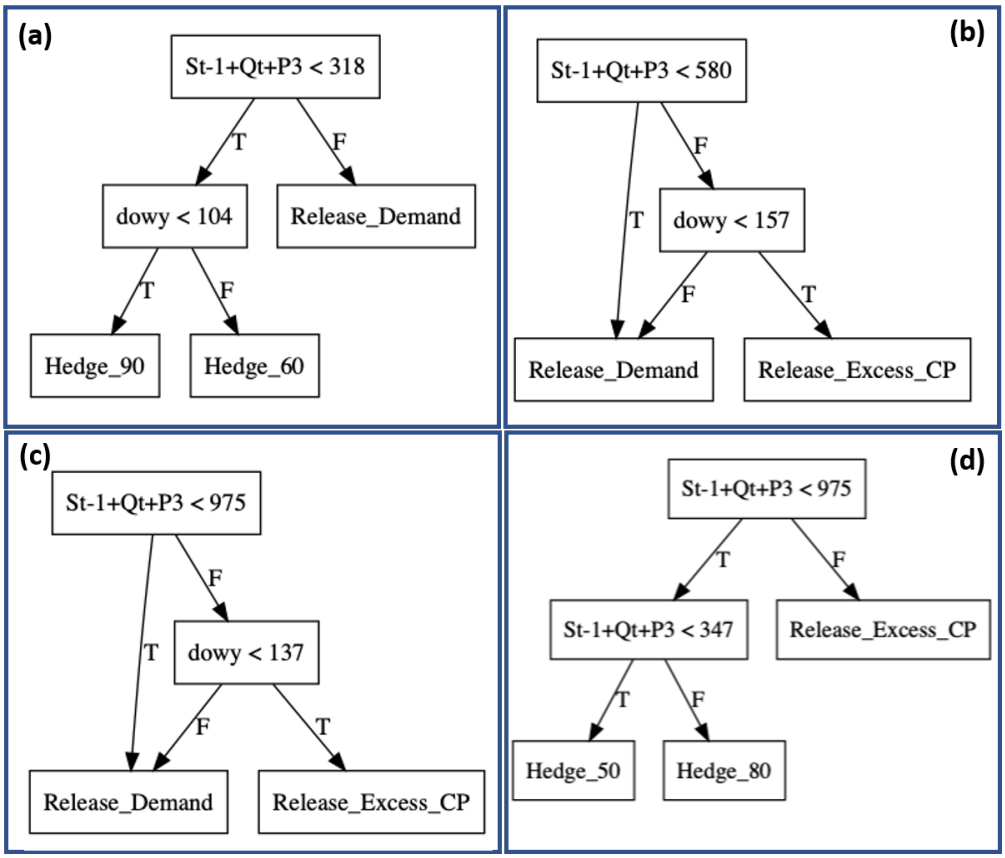

Figure 5.30: *Comparison between the resulting policy for the wet scenario in the historical time period (a) and the policy for the same wet scenario but now in the future period (b); Example of how two scenarios can generate different actions for a similar state, (c) is the policy for a wet scenario, while (d) is the one for a dry scenario.*

The first experiment aims at assessing if a policy optimized over the historical period is capable of adapting to the future period only because it is informed by the forecast. This analysis hence requires quantifying the performance over a future period of a baseline policy and a forecast-based solution, both designed over historical conditions. Results in Table 5.7 show that adding forecast information is not producing consistently more adaptive solutions than the baseline. In more details, scenarios 17 and 25 present flood events for both types of policy, meaning a proper future optimization is required for flood prevention. Then, scenarios 2, 32, 79 and 80 show a better performance under the baseline optimization than under the forecast optimization due to the fact that the latter suffers from flooding while the former does not. Lastly, the rest of the scenarios do not show any flood event and in these cases, the forecast-based operations offer a batter performance than the baseline. Therefore, there is a suggestion that the policies optimized for a forecast-based operation are more conservative in terms of hedging for water supply than the baseline operations at the

cost of increasing the flood risk.

Scenario	Past forecast policy	Past baseline
2	4.84E+10	1.80
5	1.17	1.26
17	1.76E+12	3.89E+08
20	1.15	1.55
24	0.11	0.79
25	5.19E+11	7.20E+10
32	2.73E+11	4.15
40	1.62	5.15
70	1.67	2.82
79	7.71E+10	3.69
80	1.31E+11	6.61
84	3.24	5.29
93	0.05	0.31

Table 5.7: *Comparison of the total operating costs between the forecast policy optimized in the past and the past baseline for simulations during the future time period.*

The second experiment aims at exploring the potential for designing a robust solution with respect to the uncertainty of different climate projections. Table 5.8 presents the results obtained when a policy optimized over a specific scenario (on the column) is simulated over a different scenario (on the row). The row/column "Critical flood scenario" indicates that the wet scenario can only avoid flood events when operated by its own policy. Results suggest that the only policy that is able to avoid floods for all scenarios is the one trained for the wet policy. Conversely, all the other policies suffer from flood events under the wet scenario. This means that the wet policy is the only one that could be possibly considered robust from a flood-prevention perspective.

	Optimized Policies				
Scenario	Wet	Intermediate	Dry	Driest	Critical flood scenario
Wet	0.023	3.32E+11	1.82E+11	4.65E+10	Flood
Intermediate	2.844	1.719	1.813	2.257	No flood
Dry	3.404	2.501	2.156	2.476	No flood
Driest	6.620	4.574	4.107	3.920	No flood
Critical flood scenario	No flood	Flood	Flood	Flood	

Table 5.8: *Analysis of robustness from a flood prevention perspective. The columns describe the optimize policy and the rows describe the simulated scenario. The Verdict scenario indicates if the given scenario suffered from flood events under any policy and the Verdict policy indicates if the corresponding policy results in flood events for any type of scenario.*

However, it is worth mentioning that the performance of this robust solution is outperformed by the adaptation of the baseline solution over the future (see Table 5.9) in all scenarios except the wet one. Moreover, the performance of the robust solution is comparable with the one of the baseline operations optimized over the historical conditions. This result hence suggests that although forecasts can contribute in increasing the flexibility of the solutions in adapting to the future climate (at least in terms of lower water supply costs), they do not produce significant improvement in terms of robustness of the forecast-based operations.

Scenario	Robust forecast policy	Baseline past	Baseline future
Wet	0.023	1.75544E+12	3.89E+08
Intermediate	2.844	2.82	2.55
Dry	3.404	3.69	2.69
Driest	6.620	6.61	5.49

Table 5.9: *Analysis of robustness from a water supply perspective. The columns describe the robust policy obtained in table 5.8 and the the baseline policies, in historical and future periods, while the rows describe the simulated scenarios.*

6

Conclusions and future research

Many projects suggest that water stress is expected to increase globally driven mainly by population growth and climate change, contributing to increasing the water scarcity around the world. Climate change is also likely to change water patterns at a local scale, culminating in intense short-term storms that might cause floods impacting millions of people. Water infrastructure, such as reservoirs, can help mitigate the problems related to water supply whilst also buffering incoming storms and ultimately reducing flood damages. One of the main ways of enhancing the management of a water reservoir is to add a forecast system for informing its operating policy. Forecasts can anticipate both short-term events like floods and long-term phenomena, such as droughts, and there are many studies indicating the advantage of using forecast systems in water operations. However, the contribution of a forecast under a climate change scenario characterized by more frequent and intense extreme events is still unknown.

This thesis contributes a novel procedure to quantify the future value of forecast information in a future under varying conditions generated by climate change. The proposed methodology is structured in three main blocks and is demonstrated in the case study of Folsom reservoir (California), whose operations is balancing water supply and flood control. The first block is responsible for selecting a subset of diverse climate scenarios among a set of climate large ensemble of climate projections. The second block consists in developing a synthetic forecast model to generate streamflow forecast for the future horizon

that project the skill of the existing forecast system. By matching the synthetic-forecast generator with the future climate projections, various ensembles of synthetic forecasts are obtained for the multiple climate change scenarios selected in the first step. Finally, a recently developed policy optimization framework is used to design optimal operating policies for all scenarios with and without the use of forecast, over the historical and future time periods. This allows quantifying the forecast value as the improvement in system performance generated by the forecast information, as well as analyzing its evolution over time.

Results show that water operations in the future, with and without forecast, are expected to substantially increase in cost for all scenarios tested, meaning the performance of the reservoir is likely to decrease with time. Moreover, it has also been demonstrated that the absolute value of forecast is expected to increase in all scenarios, ranging from 0.1 $(TAF/day)^2$ to 1.7 $(TAF/day)^2$ for non-flood scenario, and, up to billions of $(TAF/day)^2$ for scenarios where a flood is prevented. Nonetheless, when considering the relative gain, a different interpretation was obtained, since not every single scenario indicates increasing forecast value with time. Some scenarios suggest that in the future the relative gain of the forecasts might deteriorate up to 60% and the reason to that lies in the characteristics of the future projections. Results show that scenarios deemed as dry usually suffer from a decrease in the relative value of the forecast, ranging from -60% to +5%, whilst wet scenarios tend to present positive relative gains, from 0% to 100%. Further experiments confirmed this correlation between the forecast value and the level of wetness of a projection. The justification for this result is linked to the characteristics of the forecast system used as reference, since its range of prediction is of three days in advance. This type of short-term forecast performs better at preventing floods than at predicting long-term droughts, which explains why wetter scenarios where flood events are more likely have better overall performance than dry scenarios in which dry spells last up to two years. Finally, the adaptability and robustness of forecast-based operations were also investigated. Results show that even though forecast-based solutions designed over the historical time period are more susceptible to the enhanced floods of the future as seen in some scenarios, they are able of reducing the water supply costs of 36% when compared to the baseline. When simulating policies to multiple scenarios, it was found that there are policies that can prevent floods from taking place for the majority of the scenarios tested and also improve the water supply performance with re-

spect to the historical baseline operations optimized to each scenario, with an average increase in performance of 3%. Yet, this robust solution is 18% less efficient when compared with future baseline operations.

The results of this work suggest that while forecasts are useful in general and can contribute at preventing floods and reducing the impacts of droughts even under different hydrological regimes due to climate change, the forecast value depends on the perspective of the analysis and on the stakeholders involved. The absolute forecast value is projected to increase in the future, while the relative one might increase or decrease depending on the type of future scenario that will occur. This stalemate has the potential to be conflicting because, for instance, a forecast provider might be interested in using the absolute gain to determine its product is a long-lasting solution, whereas a reservoir operator might be more interested in the relative performance of the forecast and might be concerned of changing its current strategy to embed forecast if their value will actually decrease with time.

These conclusions suggest a number of directions for further investigation. Among the future projections, at least three showed a flood occurrence even for the perfect operation and had to be excluded, given the limitation of this work in dealing only with improving the management of the reservoir. A possible continuation would be to develop an adaptation strategy combining both management and structural improvements, in order to create a policy capable of dealing even with the worst possible outcome. A crucial assumption of this work is considering the residual propagation of the synthetic forecast stationary over time. Because the impacts of climate change are still far from being fully understood, this assumption might be questionable and would require to run a sensitivity analysis on the results with respect to potential increase or decrease in forecast skills. Moreover, the diverging results obtained analyzing absolute and relative forecast values suggest the need of more thorough and careful analysis on the mutual relationship and dependency on the underlying climate conditions, but also the possibility of trying different forecast models, such as long-term ones, more suited for long dry spells. The robustness experiment tested in this work is a posteriori, which means we tried to find the policies that could satisfy other scenarios, but naturally all of them were optimized to a specific scenario. Running a real robust optimization, conducted over multiple scenarios may allow designing solutions that better cope with the uncertainty of the future climate. Finally, we considered the Folsom reservoir as a single entity, but it is actually part of a larger reservoir network; studying

the coordinated operations of the multireservoir systems using forecast information might provide novel perspectives on the value of forecast and the ways it can contribute to water operations.

Bibliography

Adamowski, J. F. (2008), Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis, pp. 247–266, doi: 10.1016/j.jhydrol.2008.02.013.

Anghileri, D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen, and D. P. Lettenmaier (2016), Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, *Water Resources Research*, doi: 10.1002/2015WR017864.

Bierkens, M. F. P., and L. P. H. van Beek (2009), Seasonal Predictability of European Discharge: NAO and Hydrological Response Time, *Journal of Hydrometeorology*, doi: 10.1175/2009jhm1034.1.

Bing Maps (2019).

Block, P., and B. Rajagopalan (2007), Interannual Variability and Ensemble Forecast of Upper Blue Nile Basin Kiremt Season Precipitation , *Journal of Hydrometeorology*, *8*(3), 327–343, doi: 10.1175/jhm580.1.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2008), Water reservoir control under economic, social and environmental constraints, *Automatica*, *44*(6), 1595–1607, doi: 10.1016/j.automatica.2008.03.003.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *Journal of Hydrology*, *375*(3-4), 613–626, doi: 10.1016/j.jhydrol.2009.06.005.

Culley, S., S. Noble, A. Yates, M. Timbs, S. Westra, H. R. Maier, M. Giuliani, and A. Castelletti (2016), Maximum operational adaptive capacity, *Water Resources Research*, pp. 6751–6768, doi: 10.1002/2015WR018253.Received.

Damania, R., S. Desbureaux, M. Hyland, A. Islam, S. Moore, A.-S. Rodella, J. Russ, and E. Zaveri (2017), *Uncharted Waters: The New Economics of Water Scarcity and Variability*, doi: 10.1596/978-1-4648-1179-1.

Denaro, S., D. Anghileri, M. Giuliani, and A. Castelletti (2017), Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data, *Advances in Water Resources*, *103*, 51–63, doi: 10.1016/j.advwatres.2017.02.012.

Ehsani, N., C. J. Vörösmarty, B. M. Fekete, and E. Z. Stakhiv (2017), Reservoir operations under climate change: Storage capacity options to mitigate risk, *Journal of Hydrology*, *555*, 435–446, doi: 10.1016/j.jhydrol.2017.09.008.

Elsafi, S. H. (2014), Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan, doi: 10.1016/j.aej.2014.06.010.

Escriva-Bou, A., H. McCann, E. Hanak, J. Lund, and B. Gray (2016), Accounting for California Water, *California Journal of Politics and Policy*, doi: 10.5070/P2CJPP8331936.

Faber, B. A., and J. R. Stedinger (2001), Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts, *Journal of Hydrology*, *249*(1-4), 113–133, doi: 10.1016/S0022-1694(01)00419-X.

FAO (2017), *The future of food and agriculture: trends and challenges*, 180 pp., doi: 10.4161/chan.4.6.12871.

Fletcher, S. M., M. Lickley, and K. Strzepek (2019), Learning about climate change uncertainty enables flexible water infrastructure planning, *Nature Communications*, *In Press*(2019), 1–11, doi: 10.31223/OSF.IO/2TM7X.

Georgakakos, A. P., H. Yao, M. Kistenmacher, K. P. Georgakakos, N. E. Graham, F. Y. Cheng, C. Spencer, and E. Shamir (2012), Value of adaptive water resources management in Northern California under climatic variability and change: Reservoir management, *Journal of Hydrology*, *412-413*, 34–46, doi: 10.1016/j.jhydrol.2011.04.038.

Giuliani, M., F. Pianosi, and A. Castelletti (2015), Making the most of data: An information selection and assessment framework to improve water systems operations, *Water Resources Research*, doi: 10.1002/2015WR017044.

Giuliani, M., A. Castelletti, R. Fedorov, and P. Fraternali (2016), Using crowdsourced web content for informing water systems operations in snow-dominated catchments, *Hydrology and Earth System Sciences*, *20*(12), 5049–5062, doi: 10.5194/hess-20-5049-2016.

Hamlet, A., and Lettenmaier (1999), Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals, *Journal of Water Resources Planning and Management*, *125*(6), 333–341, doi: 10.1061/(ASCE)0733-9496(1999)125:6(333).

Hanak, E., J. Lund, A. Dinar, B. Gray, R. Howitt, J. Mount, P. Moyle, and B. Thomposn (2011), *Managing California's Water: From Conflict to Reconciliation*.

Herman, J. D., and M. Giuliani (2018), Policy tree optimization for threshold-based water resources management over multiple timescales, *Environmental Modelling and Software*, *99*, 39–51, doi: 10.1016/j.envsoft.2017.09.016.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (2003), An overview of the north atlantic oscillation, in *Geophysical Monograph Series*, doi: 10.1029/134GM01.

IPCC (2014), Climate Change 2014, *Tech. rep.*, doi: 10.1073/pnas.1116437108.

IPCC (2018), *Summary for Policymakers*, World Meteorological Organization, Geneva, Switzerland.

IRENA (2018), *Renewable energy statistics 2018*, International Renewable Energy Agency, Abu Dhabi.

Jajarmizadeh, M., S. Harun, and M. Salarpour (2012), A review on theoretical consideration and types of models in hydrology, *Journal of Environmental Science and Technology*.

Jenkins, M. W., J. Lund, and A. Draper (2001), Improving California water management: optimizing value and flexibility.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resources Research*, doi: 10.1029/95WR02966.

Li, Y., M. Giuliani, and A. Castelletti (2017), A coupled human-natural system to assess the operational value of weather and climate services for agriculture, *Hydrology and Earth System Sciences*, *21*(9), 4693–4709, doi: 10.5194/hess-21-4693-2017.

Lynch, P. (2008), The origins of computer weather prediction and climate modeling, *Journal of Computational Physics*, *227*(7), 3431–3444, doi: 10.1016/j.jcp.2007.02.034.

Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle (2011), Soil Moisture, Snow, and Seasonal Streamflow Forecasts in the United States, *Journal of Hydrometeorology*, doi: 10.1175/jhm-d-11-046.1.

Maher, K. (2008), Potential Use of Real-time Information for Flood Operation Rules for Folsom Reservoir.

Nayak, M. A., J. D. Herman, and S. Steinschneider (2018), Balancing Flood Risk and Water Supply in California: Policy Search Integrating Short-Term Forecast Ensembles With Conjunctive Use, *Water Resources Research*, *54*(10), 7557–7576, doi: 10.1029/2018WR023177.

Nigam, S., and S. Baxter (2014), General Circulation of the Atmosphere: Teleconnections, in *Encyclopedia of Atmospheric Sciences: Second Edition*, doi: 10.1016/B978-0-12-382225-3.00400-X.

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo (2008), New dimensions in early flood warning across the globe using grand-ensemble weather predictions, *Geophysical Research Letters*, *35*(10), 1–7, doi: 10.1029/2008GL033837.

Ramos, M. H., S. J. Van Andel, and F. Pappenberger (2013), Do probabilistic forecasts lead to better decisions?, *Hydrology and Earth System Sciences*, *17*(6), 2219–2232, doi: 10.5194/hess-17-2219-2013.

Soncini-Sessa, R. (2007), Integrated and participatory water resources management: theory.

Toth, E., A. Brath, and A. Montanari (2000), Comparison of short-term rainfall prediction models for real-time flood forecasting, *Journal of Hydrology*, doi: 10.1016/S0022-1694(00)00344-9.

Turner, S. W., J. C. Bennett, D. E. Robertson, and S. Galelli (2017), Complex relationship between seasonal streamflow forecast skill and value in reservoir operations, *Hydrology and Earth System Sciences*, *21*(9), 4841–4859, doi: 10.5194/hess-21-4841-2017.

United Nations (2017), World Population Prospects - Population Division - United Nations.

U.S. Army Corp of Engineers and U.S. Bureau of Reclamation and California Department of Water Resources and H. I. Consulting (2017), Folsom Dam modification project water control manual update.

U.S. Bureau of Reclamation (2008), Central Valley Project and State Water Project, operations criteria and plan, biological assessment.

Yuan, X., E. F. Wood, and Z. Ma (2015), A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdisciplinary Reviews: Water*, *2*(5), 523–536, doi: 10.1002/wat2.1088.

Zarfl, C., A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner (2014), A global boom in hydropower dam construction, *Aquatic Sciences*, *77*(1), 161–170, doi: 10.1007/s00027-014-0377-0.

Zhang, X., and F. Yang (2004).

Zhao, T., D. Yang, X. Cai, J. Zhao, and H. Wang (2012), Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast, *Water Resources Research*, *48*(1), 1–15, doi: 10.1029/2011WR010623.

**

Appendix A

Validation of policy tree results

The simulation of the policy trees for the selected scenarios in both historical and future time periods allows for the generation of different types of results. However, it is necessary, as a very first step, to verify that the results are feasible. So that can be accomplished, the initial step is to compare the resulting operation costs obtained by the policy tree simulation in its many forms with the costs obtained by the dynamic programming, demonstrated in subsection 5.1.3.

After generating an ensemble of 30 traces in section 5.2, a final prediction value is calculated based on the ensemble. In this work, two types of ensemble were initially considered for the policy tree, one where the predicted value corresponds to the 90th percentile of the ensemble and the other to the maximum value of the ensemble. The first validation test is to assess if the results generated by the policy tree surpass the upper boundary, represented by the DDP. As table 1 indicates, between the two variations of the ensemble and the DDP, the best operational cost is still obtained by the DDP operation, which is expected given the perfect information it possess and this supports the same pattern found in the work by *Herman and Giuliani* (2018). Moreover, given that only one variation of the ensemble is needed, one of the two needs to be selected as the actual ensemble. Again taking table 1 as reference, it is noticeable how similarly the 90th percentile of the forecast ensemble and the maximum value of the ensemble produce results. Since the idea is to establish which variation of the ensemble is most suitable and given that the usual operator profile expected in this type of operation consists in a more conservative approach, the "Max" policy is perceived as the best candidate and is thus selected for further

use.

Time period	Scenario	DDP	Max	90th percentile
Historical	Dry	0	0	0
	Intermediate	0	0	0
	Wet	0.0138	0.14	0.16
	Absolute	0.01	0.01	0.01
	Relative	0.1505	0.37	0.39
Future	Dry	1.5169	2.16	2.13
	Intermediate	1.2521	1.72	1.69
	Wet	0.0113	0.02	0.02
	Absolute	3.6999	3.92	3.91
	Relative	1.06E+08	4.13E+11	4.14E+11

Table 1: *Table illustrating the costs of each scenario for different types of policy. The columns show the type of model used and the rows define the type of scenario for each time period.*

Next step consists in the comparison between the maximum value of the ensemble with both the single trace policy and the perfect policy. The objective of this test is to understand how the costs are affected by the use of the ensemble and if there is any chance of it bypassing the synthetic residual behaviour, since the residual is stationary. Therefore, it is necessary to include a policy that uses a single trace, thus not relying on statistical operations when predicting the inflow for the next three days, instead of having the whole ensemble of forecasts. In addition, it is also required to have the perfect policy, which instead of having a synthetic forecast of the next three days, subjected to forecasting errors, has the perfect knowledge of the inflow of the next three days and is expected to provide better forecasting. The difference between these three policies can be better understood with table 2.

Time period	Scenario	Max	One trace	Perfect
Historical	Dry	0	0	0
	Intermediate	0	0	0
	Wet	0.14	0.16	0.13
	Absolute	0.01	0.01	0.01
	Relative	0.37	0.37	0.37
Future	Dry	2.16	2.13	2.05
	Intermediate	1.72	1.77	1.59
	Wet	0.02	0.02	0.02
	Absolute	3.92	3.89	3.85
	Relative	4.13E+11	4.27E+11	3.96E+11

Table 2: *Table highlighting the difference in values between the ensemble of forecasts, a single trace and the perfect policy. The rows define the type of scenario for each time period.*

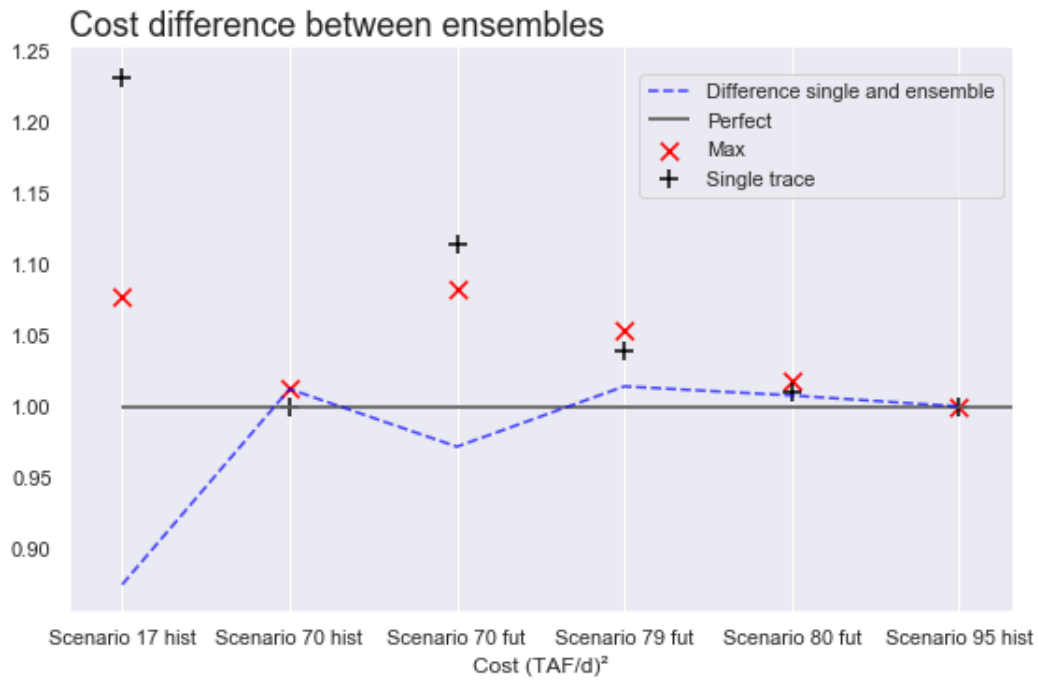

Figure 1: *Estimation of costs by the perfect forecast, the ensemble and single trace. The grey line stands for the standardized perfect-operation cost, while the markers represent how much more expensive the operations are for each case. The traced blue line illustrates the difference in cost between the ensemble and a single trace policy.*

There is a confirmation that the perfect policy is the one presenting the best results along all the scenarios, guaranteeing the feasibility of the policy results. Moreover, there is no clear domination between the max and one trace variations and figure 1 illustrates the case, where the difference between the operational costs among the ensembles and the single policy are showed. Since there is the absence of any evidence suggesting a statistical prevalence or relation between the values, it is safe to attest the maximum value of the ensemble is still preserving the forecast error with respect to the observed inflow and thus duly replicating an actual ensemble of forecasts and not overperforming the predictions.