POLITECNICO DI MILANO

Facoltà di Ingegneria

Dipartimento di Elettronica, Informazione e Bioingegneria

Master of Science in

Environmental and Geomatic Engineering

Using ENSO information to improve the operation of the Hoa Binh reservoir, Vietnam

Supervisor:

PROF. ANDREA CASTELLETTI

Assistant Supervisor:

DR. MATTEO GIULIANI

Master Graduation Thesis by:

HUNG VUONG PHAM
Student Id n. 813497

Academic Year 2014-2015

ACKNOWLEDGMENTS

My study in Politecnico di Milano is funded by Integrated and sustainable water Management of Red-Thai Binh Rivers System in changing climate (IMRR) project. Therefore, I would like to extend my sincere gratitude to university of Politecnico di Milano, Italy and Institute of Water Resources Planning, Vietnam for giving me this great opportunity.

First, I would like to thank my research supervisor, *Prof. Andrea Castelletti*, for his teaching, his orientation, his instruction and his examination of my thesis. I really appreciate his supports and his enthusiasm.

Second, I would like to express my gratitude to my co-supervisor, *Dr. Matteo Giuliani*, for his teaching, his instruction, his technical assistance and his correction of my thesis. Without his help I could not accomplish my thesis on time.

Third, I would like to say thanks to all Professors in Politecnico di Milano who taught and gave me knowledge in different fields. They contributed to my success study in Italy. I also would like to thank all assistants, especially *Emanuele Mason* and *Ahmad Alsahaf*, who gave me great technical supports.

I would like to acknowledge my colleagues in Institute of Water Resources Planning, Vietnam, who provided all the data I need for this thesis. Especially, I want to thank *Dr. Xuan Quach*, *Anh Truong*, and *Nam Vu* for their supports and their orientations.

Finally, I would like to thank my family and my friends who always support and encourage me. I really appreciate their supports and their contributions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ENSO El Niño Southern Oscillation

IIS Iterative Input variable Selection

IVS Input Variable Selection

MEI Multivariate ENSO Index

SISO Single Input-Single Output

SOI Southern Oscillation Index

SST Pacific Sea Surface Temperature

ISA Information Selection and Assessment

EVPI Expected Value of Perfect Information

POP Perfect Operating Policy

BOP Basic Operating Policy

IOP Improved Operating Policy

EVSI Expected Value of Sample Information

DDP Deterministic Dynamic Programming

SLP Sea Level Pressure

NOAA National Oceanic and Atmospheric Administration

ONI Ocean Niño Index

IOP Improved Operating Policy

HV Hypervolume

WL Water Level

RRB Red River Basin

DDP Deterministic Dynamic Programming

IMRR Integrated and sustainable water Management of Red-Thai Binh Rivers System in changing climate

ABSTRACT

Climate change is one of the most important factors taking responsibilities for the changes of water cycles and extreme events such as floods and droughts. Recent observational and theoretical studies demonstrated that there is a teleconnection between these phenomena and the large-scale low frequency fluctuations such as El Niño Southern Oscillation (ENSO). Therefore, the operations of many water systems is expected to benefit from a better understanding of these global phenomena and their impacts at the river basin scale.

This thesis proposes a framework for detecting ENSO teleconnection and using this information for improving water systems operations. The framework is applied to the operations of the Hoa Binh reservoir, a multi-objective reservoir in Vietnam, taking into account hydropower production, water supply, and flood mitigation. The framework is structured in three main steps. First, we determine the set of ideal operating policies for the regulation of the Hoa Binh reservoir under the assumption that we perfectly know the future. Second, we automatically select the most valuable information, which is useful for the improvement of the Hoa Binh operations, using the Input Variable Selection (IVS) techniques. Finally, we design a set of Improved Operating Policies using the set of selected information and we quantify the value of this information information by measuring the associated improvement in the system performance.

Numerical results illustrate that our framework successfully improves the operations of the Hoa Binh reservoir with respect to the three conflicting objectives. Interestingly, our results show that the use of one ENSO index related to Sea Surface Temperature in the Pacific Ocean provides a significant contribution to the Hoa Binh operations. By extending the foresight on the future hydrological conditions, which is a key information for allocating the water during the Vietnamese hydrologic year across the multiple sectors, this index produces an 3.3% increase in the quality of the designed solutions.

1 INTRODUCTION

1.1 MOTIVATION

Recently, people in the world have witnessed greater demands for reliable, high quality and inexpensive water supplies for agriculture, industry and domestic consumption due to the growth of population and industrialization [*Gerland et al.*, 2014; *Cominola et al.*, 2015]. In recent decades there have also been increasing awareness about the need of supporting healthy and diverse ecosystems, reducing and preventing floods and droughts, producing energy via hydropower or ensuring water levels adequate for ship navigation [*Giuliani and Castelletti*, 2013; *Castelletti et al.*, 2013; *Giuliani et al.*, 2014a]. Water managers are challenged to meet these multiple and often conflicting demands [*Loucks et al.*, 2005].

Additionally, managers are facing with growing uncertainties and variability of natural regimes due to changes in the underlying climatic conditions [*IPCC*, 2013]. Climate change is expected to accelerate this trend, ultimately increasing the probability of extreme events.

Medium-to-long range streamflow forecast provides a key assistance in anticipating hydro-climatic adverse events and prompting effective adaptation measures. For example, accurate medium-long range streamflow forecasts have a great potential to improve water reservoir operation by enabling more efficient allocation of water volumes in time (e.g., via hedging). Unfortunately, these forecasts often lacks of reliability and accuracy, especially when low-frequency climate forcing (e.g., ENSO) is not intense enough to improve the forecast lead time (e.g., in Europe), and might be computationally very demanding.

1.2 OBJECTIVES OF THE THESIS

Low frequency climatic fluctuations, in fact, are known to be a major factor causing extreme hydrological events and significant variations in water resources at global and regional scales [*Kahya and Dracup*, 1993]. There have been significant advances in physically-based models of the hydro-climatic systems in recent decades, but their prediction skills beyond a few days or weeks remain rather limited [*Sharma*, 2000]. Instead, statistical approaches that relate "at-site" hydrology to large-scale fluctuations such as El Niño Southern Oscillation (ENSO) could provide a basis for useful seasonal to inter-annual flow forecasts [*Sharma*, 2000]. Several works can be found in literature investi-

gating the relationship between ENSO and hydro-meteorological processes throughout the world, but there is currently no consensus on how this task should be undertaken.

The main goal of this thesis is to provide a novel procedure for: *(i)* identifying the impact of ENSO on the hydro-meteorological conditions at the river basin scale; *(ii)* Assessing the potential improvement of current operating policies using the identified ENSO information from previous step.

The methodology, which rely on the Information Selection and Assessment (ISA) [*Giuliani, Pianosi, and Castelletti*, 2015b], is composed of three main methodological steps which can be summarized as follows:

First, we compute the upper bound of the system performance by assuming perfect foresight and we assess the value of additional information as the difference between this ideal solution and current operation.

Second, using input variable selection, we then select the most relevant information, including ENSO information, to explain the release trajectory associated to the upper bound operating policy.

Finally, we derive the optimal policy conditioned upon the selected variables by Multi-Objecting Evolutionary Direct Policy Search. Additionally, an assessment of the improvement is carried out to determine how much the operating policy was improved.

In this thesis, the above approach is applied on the operation of Hoa Binh reservoir in Vietnam. This reservoir is a typical example for multi-users and multi-demands in Red River basin, Vietnam. It was built to meet three main objectives, namely mitigation of flood damages, hydropower production and water supply to civil, industrial and agriculture users. The most difficult question is how water resources systems can become more integrated and sustainable?

In the literature, there are few researches which focus on the operation of Hoa Binh reservoir. For instance, *Le Ngo et al.*, 2008 used MIKE 11 model to analyze and compare three alternative operating policies of flood control and hydropower production concerning on the flood season only. *Quach*, 2011 used different methods and tools to assess the ability for operation improvement of the Hoa Binh reservoir and to design compromise operating policies that can reduce the conflict among different operation objectives. *Vu and Truong*, 2010, *Cavalli*, 2012 and *Beltrame and Carbonin*, 2013 used different approaches and methods to model the inflow of Hoa Binh reservoir.

The results of these studies, however, leave space for further improving the management of the water resources in the Red River basin by better informing the operating policies adopted for regulation of the main reservoirs in the system. In particular, ENSO indices

can be used to extend the foresight of the managers in order to better manage particularly dry or wet conditions. Indeed, according to *Räsänen and Kummu*, 2013, there is a strong but lagged relationship between ENSO and hydrology of the Mekong, including Vietnam. The influence was found to be significantly stronger during the decay years of the ENSO events when compared to the onset years. However, the influence of ENSO was not uniform respect to spatial scale and time scale. This thesis aims to reply to these questions (e.g., which ENSO index is the most important or which time scale should be considered) by demonstrating the benefits of using ENSO in the operations of Hoa Binh reservoir.

1.3 OUTLINE OF THE THESIS

This thesis is structured in six chapters, organized as follows:

CHAPTER 2 Provides the basic concept about El Niño Southern Oscillation (ENSO) phenomena, its characteristics, its indexes and its effects on water system.

CHAPTER 3 Provides the short description of the Information Selection and Assessment (ISA) framework and its tools need to follow in order to obtain the result.

CHAPTER 4 Provides an overview of the Hoa Binh reservoir, its characteristics and its model.

CHAPTER 5 Reports the experiments carried out and discusses the results.

CHAPTER 6 Summarizes the conclusions of the thesis and provides some recommendations for future studies and researches.

2 EL NIÑO SOUTHERN OSCILLATION

2.1 INTRODUCTION

El Niño Southern Oscillation (ENSO) is a large scale coupled ocean atmosphere phenomenon centered in the equatorial Pacific. It consists of two oceanic phases - the warm *El Niño* phase and the cold *La Niña* phase - that are connected to the atmosphere through a seesaw atmospheric pressure fluctuation in the South Pacific called the *Southern Oscillation* [*Shrestha and Kostaschuk*, 2005].

Typically, ENSO events occur at irregular intervals, with a characteristic return frequency of 2-7 years (which makes ENSO a quasi-periodic phenomenon) and they usually persist for 1–2 years. No two events are completely alike: they evolve according to a consistent pattern, but they differ in timing, intensity, extent and duration [*Kahya and Dracup*, 1993].

Over the last years, many evidences suggested that ENSO is associated to climate anomalies throughout the world. Especially, ENSO-streamflow teleconnection was identified in many regions in the world. Strong and regionally consistent ENSO-streamflow teleconnections are identified in Australia and New Zealand, South and Central America, and weaker signals are identified in some parts of Africa and North America [*Chiew and McMahon*, 2002].

2.2 DEFINITION OF ENSO

The ENSO cycle is a scientific term that describes the fluctuations in temperature between the ocean and atmosphere in the east-central Equatorial Pacific. ENSO is composed of two opposite phases, namely *El Niño* and *La Niña*. La Niña is sometimes referred to as the cold phase of ENSO and El Niño as the warm phase of ENSO.

El Niño and La Niña episodes typically last nine to twelve months, but some prolonged events may last for years. While their frequencies can be quite irregular, El Niño and La Niña events occur on average every two to seven years. Typically, El Niño occurs more frequently than La Niña.

EL NIÑO El Niño episodes refer to the large-scale ocean-atmosphere climate interaction linked to a periodic warming in sea surface temperatures across the central and east-central Equatorial Pacific. El Niño was originally recognized by fishermen of the coast of South America in the 1600s, with the appearance of unusually warm water in the Pacific Ocean. The name was cho-

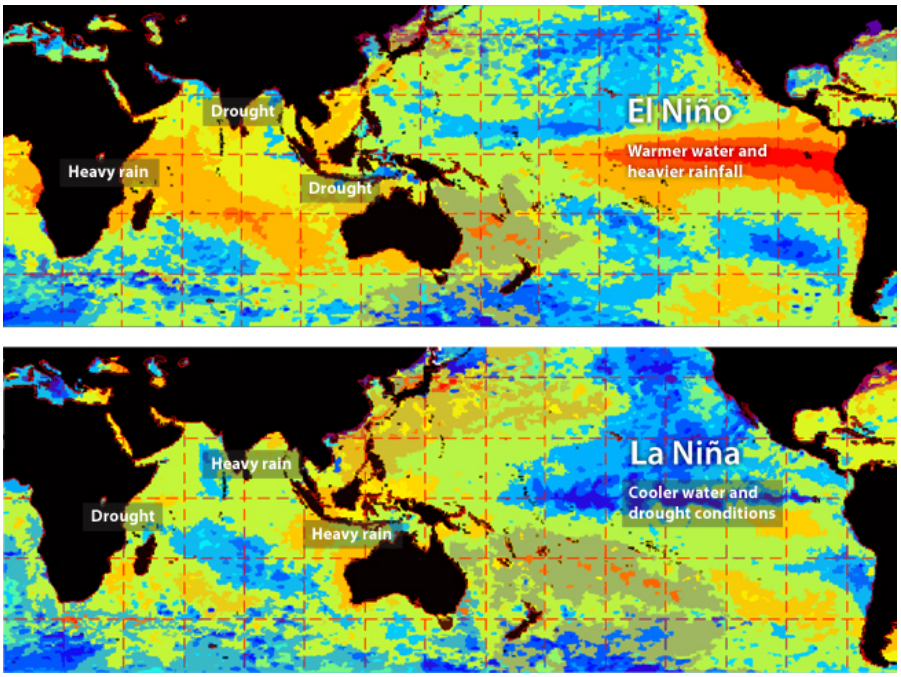


Figure 2.1: El Niño and La Niña. [1]

sen based on the time of year (around December) during which these warm water events tended to occur [NOAA].

LA NIÑA La Niña episodes represent periods of below-average sea surface temperatures across the east-central Equatorial Pacific. Global climate La Niña impacts tend to be opposite those of El Niño impacts. In the tropics, ocean temperature variations in La Niña also tend to be opposite those of El Niño [NOAA].

During normal conditions, the high pressure in the eastern Tropical Pacific Ocean and the low pressure in the western part cause the trade winds to blow westwards (i.e., normally, trade winds are driven from the eastern to the western Pacific by the differences in atmospheric pressures). The trade winds tend to transport warm water westwards along the equator, maintaining a buildup of warm surface water in the western tropical Pacific - Indonesia region. This large warm pool is the most extensive region in the world of water warmer than 28°C and is considered a major source of atmospheric heating, which drives large-scale convective circulation patterns [*Kahya and Dracup*, 1993]. On the other hand, the trade winds cause - with their circulation - an upwelling of cold water in the coastal areas of South America. This atmospheric pressure pattern creates wetter climate (heavy rainfall) in the western Pacific and drier climate in the eastern Pacific Ocean [*Allan et al.*, 1996].

The signs of the development of an El Niño event are:

1 http://mit.whoi.edu/student-research?tid=1621&cid=83611&article=53506

- Cooling of the western tropical Pacific Ocean;
- Sustained warming of the central and eastern tropical Pacific Ocean;
- Rise in surface pressure over the Indian Ocean, Indonesia and Australia (above normal pressure over Darwin);
- Fall in air pressure over the central and eastern tropical Pacific Ocean (below normal pressure over Tahiti).

As the air pressure pattern of normal condition reverses, the trade winds in the South Pacific decrease in strength: they actually either weaken or turn eastwards (i.e., they start heading east). This causes warm surface water to flow east, spreading from the west Pacific and the Indian Ocean to the east Pacific, and consequently creates drier climate to the west and wetter to the east: extensive droughts diffuse through the western Pacific (Australia) and rainfall gets to the normally dry eastern Pacific (Peru) [*Beltrame and Carbonin*, 2013].

On the contrary, La Niña events have opposite characteristics compared with those of El Niño events, which have just been described above.

2.3 THE ROLE OF ENSO IN HYDROLOGICAL REGIMES

Recently, several researches and studies have recognized the teleconnection between ENSO and global hydro-climatic variability, such as the variations in precipitation and streamflow. For instance, *Cobb et al.*, 2003 assessed future climate in relationship with ENSO; *Chiew and McMahon*, 2002 presented an overview of global ENSO-streamflow teleconnection; *Allan et al.*, 1996 showed the relationship between Southern Oscillation and climatic variability. Some other researches have been carried out in Asian countries, such as *Räsänen and Kummu*, 2013; *Shrestha and Kostaschuk*, 2005; *Ronghui et al.*, 2004; *Wang, Wu, and Fu*, 2000.

The teleconnection between ENSO and climate can be used to producing long-range weather forecasts as provided by meteorological agencies in several countries such as USA, Australia, China, India and Sri Lanka. Numerous investigations made over the last three decades have indeed demonstrated significant correlations between ENSO and rainfall anomalies in many areas [*Chiew and McMahon*, 2002].

Recent studies have also demonstrated that ENSO information becomes more and more useful for informing water resources management. ENSO events can now accurately be predicted up to one year in advance using a physical model of the coupled ocean-atmosphere system [*Simpson et al.*, 1993]. Therefore, it helps the managers and the decision makers to improve their operating policies, ultimately allow-

ing the mitigation of conflicts among stakeholders and water users and maximizing the overall system benefit.

2.4 ENSO INDICES

To detect ENSO phenomena, several indexes have been formulated, such as Southern Oscillation Index (SOI), Ocean Niño Index (ONI), Pacific Sea Surface Temperature (SST) - calculated over the different areas of the tropical Pacific Ocean known as Niño $1+2$, Niño 3, Niño 3.4 and Niño 4 - and Multivariate ENSO Index (MEI). They are the most common indexes, but it is hard to say which one is the most important on specific study sites. Different researches have used different approaches with different indexes or a combination of some indexes. For example, *Chiew and McMahon*, 2002 used SOI and MEI for the research in 581 catchments worldwide; *Kiem and Franks*, 2001 used SOI, SST and MEI for the research in Australia.

2.4.1 *Southern Oscillation Index*

The Southern Oscillation Index (SOI) is defined as the difference in Sea Level Pressure (SLP) between Papeete (Tahiti) and Darwin (Australia). More specifically, the SOI is calculated as the difference in monthly averages of standardized mean sea level pressure at each station. Formally, the SOI index is formulated as follows:

$$\mathrm{SOI} = \frac{\mathrm{Standardized\ Tahiti} - \mathrm{Standardized\ Darwin}}{\mathrm{MSD}} \tag{2.1}$$

Where

$$\mathrm{Standardized\ value} = \frac{\mathrm{Actual\ SLP} - \mathrm{Mean\ SLP}}{\mathrm{Standard\ Deviation}} \tag{2.2}$$

$$\mathrm{Standard\ Deviation} = \sqrt{\frac{\sum_{k=1}^{N} \left(\mathrm{Actual\,SLP} - \mathrm{Mean\ SLP}\right)^2}{N}} \tag{2.3}$$

$$\mathrm{MSD} = \sqrt{\frac{\sum_{k=1}^{N} \left(\mathrm{Standardized\,Tahiti} - \mathrm{Standardized\,Darwin}\right)^2}{N}} \tag{2.4}$$

where N is number of months; MSD is Monthly Standard Deviation; and SLP is Sea Level Pressure.

The geographical areas used for its computations are shown in figure 2.2. The SOI index is available on the website of National Oceanic and Atmospheric Administration (NOAA). A representative trajectory of this index is reported in figure 2.3.

2 http://www.climate.gov/news-features/understanding-climate/climate-variability-southern-oscillation-index

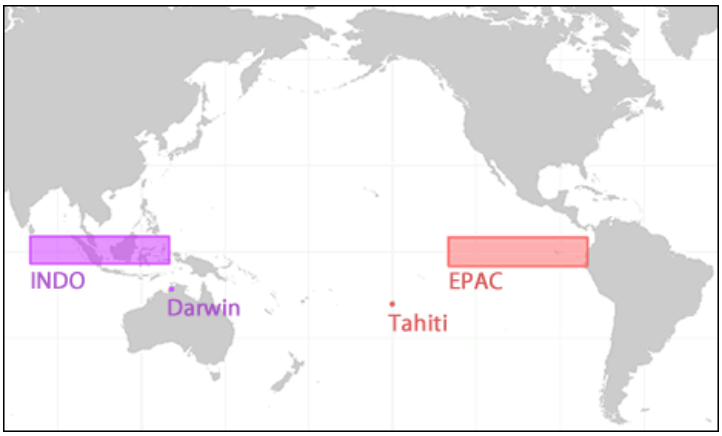


Figure 2.2: Location of Tahiti and Darwin, and of the Eastern Pacific (EPAC) and Indonesia (INDO) regions. [2]

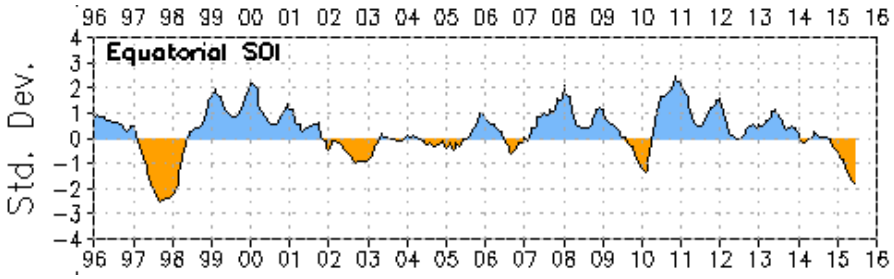


Figure 2.3: Southern Oscillation Index. [3]

Typically, prolonged periods of negative SOI values associate with El Niño episodes while prolonged periods of positive SOI values associate with El Niña episodes [NOAA].

2.4.2 *Sea Surface Temperature*

The Pacific Sea Surface Temperature (SST) is defined as the skin temperature of the ocean surface water. Recently, the NOAA developed some products, e.g., CoastWatch SST, SST Anomaly Charts, SST Contour Charts, SST Imagines, SST Monthly Mean and SST Nighttime Field Imagine Charts, which allowed to measure SST and SST anomaly in specific position as well as in grid. Numerous SST indexes have been derived in different areas of the equatorial Pacific Ocean.

- Niño1+2: 80°W – 90°W, 10°S – Equator;
- Niño3: 90°W – 150°W, 5°S – 5°N;
- Niño3.4: 120°W – 170°W, 5°S – 5°N;
- Niño4: 150°W – 180°W, 5°S – 5°N.

The location of SST regions are shown in figure 2.4. A representative

3 http://www.cpc.ncep.noaa.gov/products/CDB/Tropics/figt2.gif

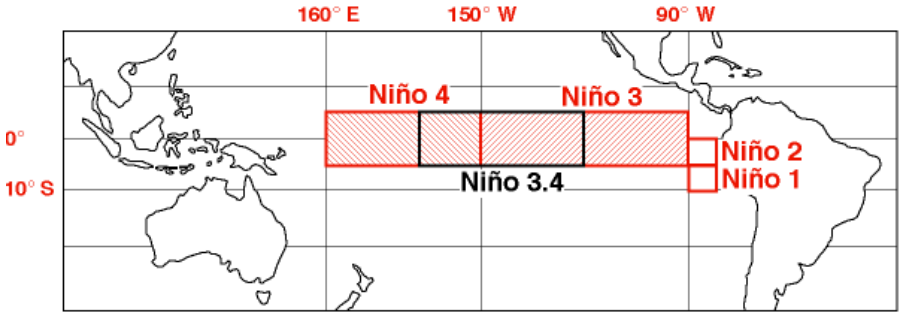


Figure 2.4: Equatorial Pacific SST regions. [4]

trajectory of this index is reported in figure 2.5.

The SSTs characterizing the identified regions are averaged for each corresponding ENSO index [*Chandimala and Zubair*, 2007]. For instance, the Niño4 index is defined as the seasonal SST averaged over the Niño4 region, which is located in the central/eastern Pacific Ocean between 5°S and 5°N in latitude and 150°W and 180°W in longitude.

Typically, El Niño is associated with warmer than usual Sea Surface Temperatures, while La Niña events are characterized by lower temperatures [NOAA].

2.4.3 *Oceanic Niño Index*

The anomaly of Pacific Sea Surface Temperature from the 30-year average of the three-month average is known as the Ocean Niño Index (ONI) (see figure 2.6). NOAA has determined the average monthly sea surface temperature for a particular swath of the tropical Pacific Ocean by averaging measurements collected there over the past 30 years. Scientists refer to that swath as the Niño 3.4 region. To filter out month-to-month variability, average sea surface temperature in the Niño 3.4 region is calculated for each month, and then averaged with values from the previous month and following one. This running three-month average value is compared with average sea surface temperature for the same three months during 30-year. The World Meteorological Organization suggests that the 30-year normal be based on a 30-year base period that starts at the beginning of each decade (e.g., 1951 – 1980, 1961 – 1990, etc.) [*Smith and Reynolds*, 2003]. To be up-to-date, the period 1981-2010 can be used today as 30-year normal.

Typically, El Niño episodes are characterized by a positive ONI greater than or equal to +0.5°C while La Nina episodes are characterized by a negative ONI less than or equal to -0.5°C [NOAA].

4 http://www.ucar.edu/communications/newsreleases/1998/ninatip.html
5 http://www.cpc.ncep.noaa.gov/products/CDB/Tropics/figt5.gif
6 http://ggweather.com/enso/oni.htm

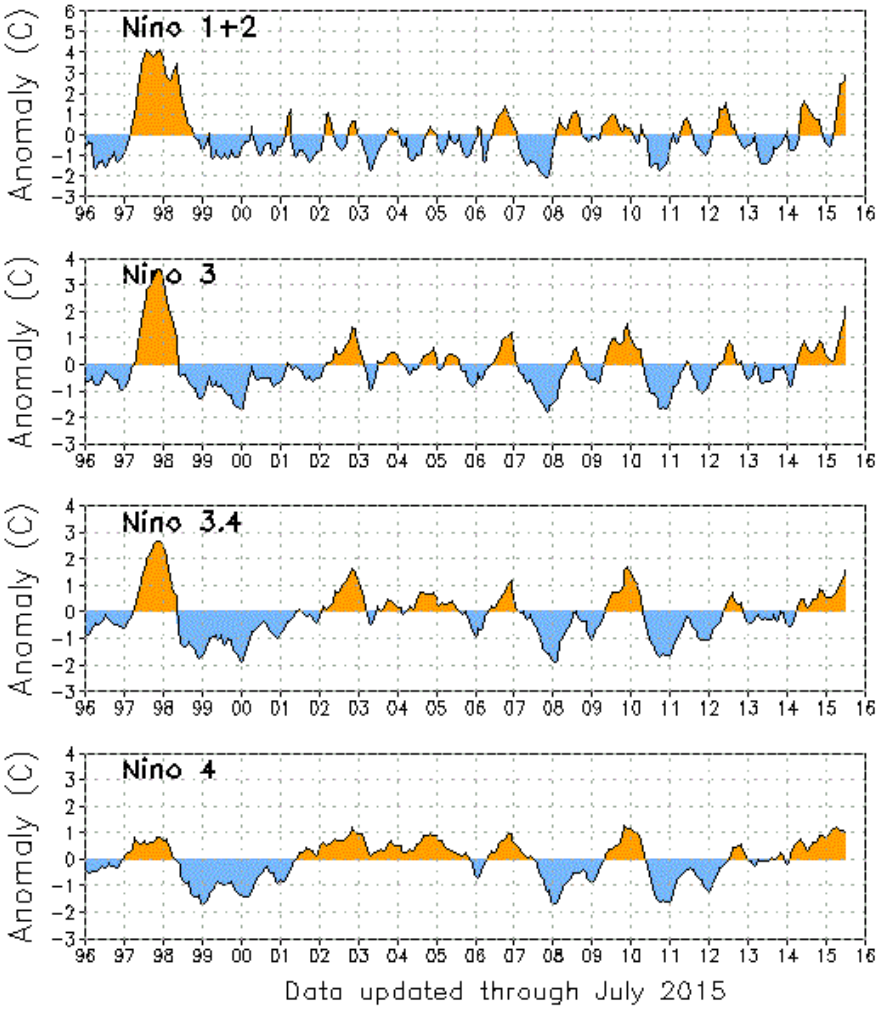


Figure 2.5: Pacific Sea Surface Temperature anomaly indexes. [5]

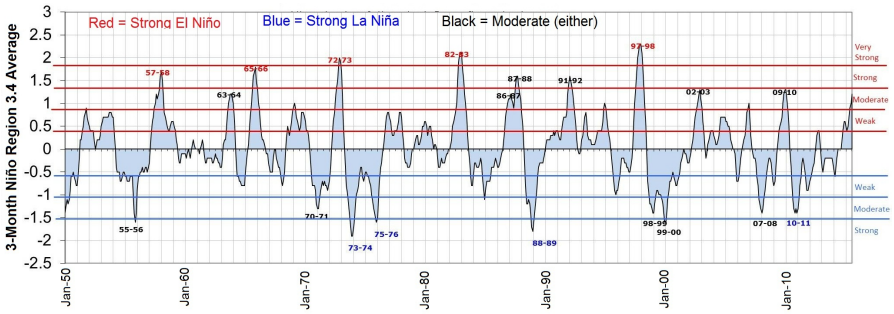


Figure 2.6: Ocean Niño Index. [6]

2.4.4 *Multivariate ENSO Index*

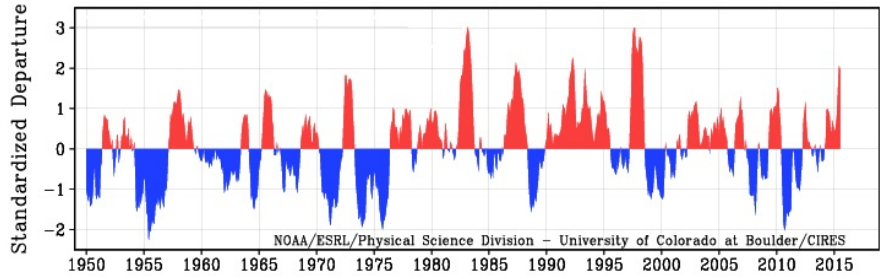


Figure 2.7: Multivariate ENSO index. [7]

The Multivariate ENSO Index (MEI) is a complex index which is based on six observed variables over the tropical Pacific Ocean, including SST, SLP, zonal and meridional components of the surface wind, Surface Air Temperature and total cloudiness fraction of the sky. Because of its integrating information in comparison with SOI and SST-based indexes, which are each based on a single variable, the MEI would result less vulnerable to non-ENSO related variability [*Kiem and Franks*, 2001]. Commonly, El Niño conditions are associated with large positive MEI values, while La Niña events are characterized by negative MEI values [*Chiew and McMahon*, 2002] (see figure 2.7).

7 http://www.esrl.noaa.gov/psd/enso/mei/

3

ISA FRAMEWORK

The Information Selection and Assessment (ISA) framework is a methodology which allows to automatically select the most valuable information for informing water systems operations and to provide metrics for quantitatively and economically assessing the value of this information [*Giuliani, Pianosi, and Castelletti*, 2015b].

The ISA framework is composed by three main interlaced steps as shown in figure 5.6.

STEP 1 The Expected Value of Perfect Information (EVPI) is quantified as the benefit obtained by completely eliminating uncertainty from dynamic model process, i.e., assuming that perfect information on the future system's conditions is available.

STEP 2 Selecting the information that more valuably contribute to the system operations among the observed exogenous information using Input Variable Selection (IVS) algorithm.

STEP 3 The operations of the system is re-designed to cater for the selected information from step 2. In addition, the Expected Value of Sample Information (EVSI) is estimated, in order to obtain strategic insights on the potential for using the selected information.

3.1 PROBLEM FORMULATION

We consider a general water reservoir management problem [*Castelletti, Pianosi, and Soncini-Sessa*, 2008] of the following form:

$$\min_{\mathbf{u}_{[0,H-1]}} J(\mathbf{x}_{[0,H]}, \mathbf{u}_{[0,H-1]}, \varepsilon_{[1,H]}) \tag{3.1a}$$

subject to

$$\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \varepsilon_{t+1}) \qquad t = 0, ..., H-1 \tag{3.1b}$$

$$(\mathbf{x}_0, \varepsilon_{[1,H]}) \text{ given} \tag{3.1c}$$

where

- $\mathbf{u}_t$ is the vector of release decisions at each time step $t = 0, \ldots, H-1$ over the evaluation horizon $[0, H]$.
- $J(\cdot)$ is the objective function, which measures the overall cost (to be minimized) encountered over the evaluation horizon. It depends on the sequence of decisions $\mathbf{u}_{[0,H-1]}$, the trajectory of

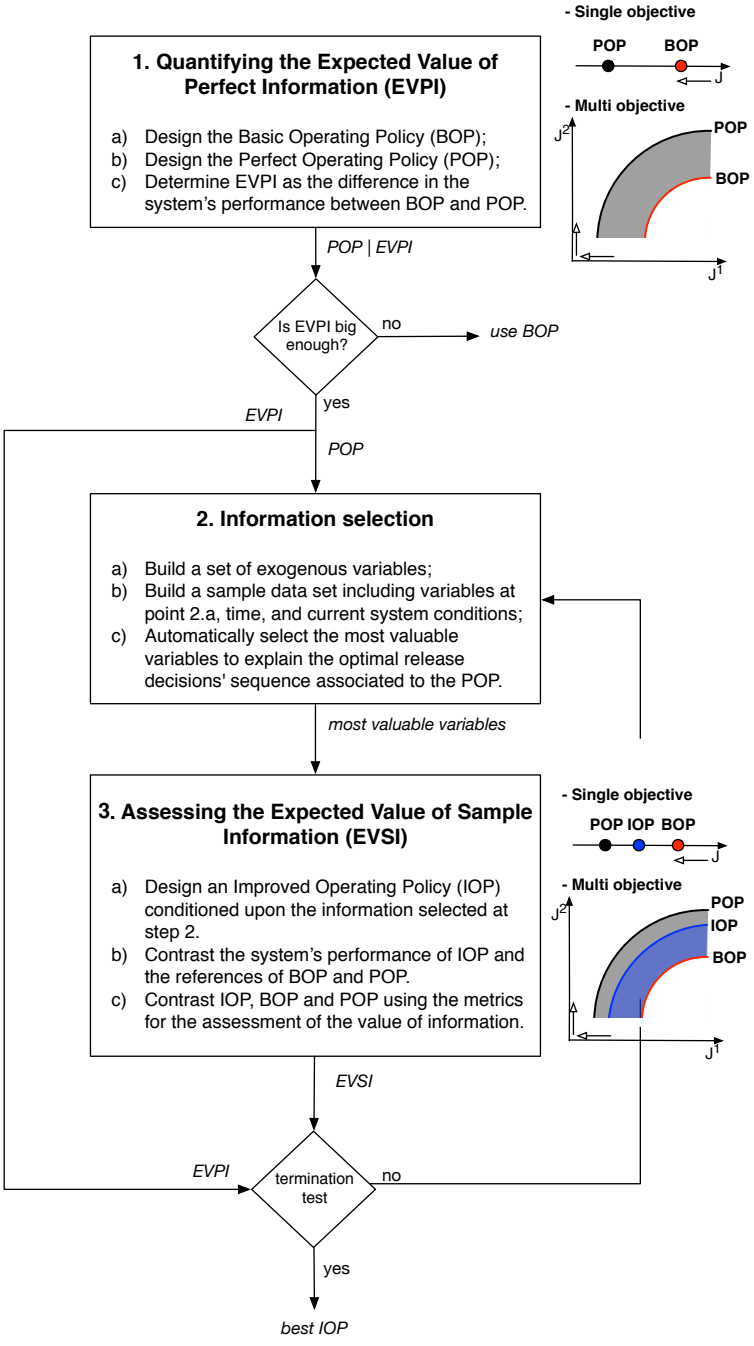


Figure 3.1: Schematization of the Information Selection and Assessment framework [*Giuliani, Pianosi, and Castelletti*, 2015b].

the external drivers $\varepsilon_{[1,H]}$ (e.g., reservoir inflows), and the resulting trajectory of the state vector $\mathbf{x}_{[0,H]}$ (e.g., reservoir storage).

- $f(\cdot)$ is the transition function of the system (e.g., reservoirs' water balance and flow-routing), whose recursive application allows the dynamic simulation of the system's evolution in time. In the adopted notation, the time subscript of a variable indicates the instant when its value is deterministically known. The reservoir storage is measured at time t and thus is denoted as x_t, while inflow in the interval $[t, t+1)$ is denoted as q_{t+1} because it can be known only at the end of the time interval.

3.2 QUANTIFYING THE EXPECTED VALUE OF PERFECT INFORMATION

The EVPI is defined as the performance improvement that could be achieved under the assumption that we perfectly know information of the future and we use it to optimize the operations of the water system. As the matter of fact that it cannot be evaluated for the future, it can be reconstructed for the past, provided that time series of the external drivers are available over a sufficiently long historical horizon and at reasonable accuracy. These time series are used to design the optimal release decisions' sequence by solving backward the problem. We call Perfect Operating Policy (POP) such a reconstructed optimal operations. Formally, the decision vector $\mathbf{u}_t \cong \mathrm{POP}(t, \mathbf{x}_t, \varepsilon_{[t+1,H]})$ can be determined at each time t as a function of the current state vector $\mathbf{x}_t$ and the sequence of future external drivers $\varepsilon_{[1,H]}$ assumed to be perfectly known.

The performance of this Perfect Operating Policy (J^{POP}) is an absolute measure of system performance, which depends also on the characteristics of the system under study (e.g., reservoir capacity/inflow ratio). Thus, the EVPI has to be estimated by contrasting POP performance (i.e., J^{POP}) with the value of the objective function that could be obtained, over the same evaluation horizon, by a poorly informed operating policy (i.e., Basic Operating Policy - BOP) relying on a basic set of information. For instance, we can consider the release plan, which means an operating policy that depends only on the day of the year (i.e., $\mathbf{u}_t = \mathrm{BOP}(t)$).

Since the BOP relies on the basic information, we expect that $J^{POP} < J^{BOP}$ and the difference $J^{POP} - J^{BOP}$ represents the EVPI, meaning the performance improvement generated by the use of perfect information. A positive and sufficiently big EVPI indicates a potential benefit in using more information to improve the operations of the system, whereas a small EVPI suggests that is not worth trying to improve the Basic Operating Policy.

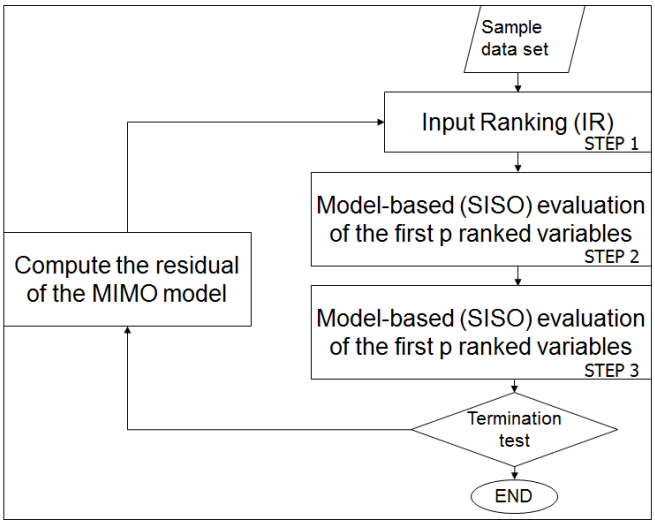


Figure 3.2: Flowchart of the IIS algorithm. [*Galelli and Castelletti*, 2013]

3.3 INFORMATION SELECTION

In order to close the gap between POP and BOP (i.e., the EVPI is big), we need to identify a set of sample information, observable at time t, that can act as an effective surrogate of the sequence of future external drivers and characterizes as much accurate as possible the optimal release decisions' sequence associated to the POP [*Giuliani, Pianosi, and Castelletti*, 2015b]. The final goal of this step is the selection of a set of potential explanatory variables, which explain well the variable of interest (i.e., the optimal sequence of release decisions $\mathbf{u}^{POP}_{[0,H-1]}$), among a large number of candidates. There are several methods can be applied to solve such information selection problem, such as cross-correlation or mutual information analysis, or more complex Input Variable Selection (IVS) techniques [*Galelli et al.*, 2014]. In this thesis, we used the hybrid model-based/model-free Iterative Input variable Selection (IIS) algorithm [*Galelli and Castelletti*, 2013] combined with extremely randomized trees.

The tree-based IIS is a novel hybrid approach that incorporates some of the features of model-based IVS techniques into a fast model-free method, which is able to handle very large candidate input sets. The optimal subset is incrementally built combining the information content of the data estimated through a ranking-based procedure followed by a model-based forward selection process [*Galelli and Castelletti*, 2013]. The IIS algorithm is composed of three main steps as shown in the figure 3.2.

STEP 1 The IIS algorithm runs an Input Ranking algorithm to sort the n candidate inputs according to a non-linear statistical measure of significance. The score of each candidate input estimates its contribution to the underlying model of the output variable y (i.e., the optimal sequence of release decisions $u^{POP}_{[0,H-1]}$. In prin-

ciple, the first variable in the ranking should be the most significant in explaining the output. However, the potential redundancy, collinearity, and nonlinear relationships between inputs and outputs can bias the initial ranking, with the most relevant input variables not be listed in the very top positions. To reduce the risk for misselection, the first p variables in the ranking are individually evaluated in the following step.

STEP 2 The relative significance of the first p−ranked variables obtained from the previous step is assessed against the observed output y through a model-based procedure. To this end, p Single Input-Single Output (SISO) models are identified with an appropriate model building algorithm and compared in terms of a suitable distance metric (e.g., mean-squared error) between the output y and each SISO model prediction. The best performing input among the p considered is added to the set of the variables selected to explain y.

STEP 3 A model building algorithm is then run to identify a Multi Input-Single Output model mapping the variables selected into the output y. The IIS algorithm is then iterated until either the best variable returned is already in the set, or the performance of the underlying model does not significantly improve. At each iteration, the process is repeated using the residuals as the new output variable in steps 1 and 2. The reevaluation of the ranking on the model residuals every time a candidate variable is selected ensures that all the candidates that are highly correlated with the selected variable, and thus may become useless, are discarded. This strategy reinforces the SISO model-based evaluation in step 2 against the selection of redundant variables and is independent of the model building and Input Ranking algorithms adopted [*Galelli and Castelletti*, 2013].

The IIS algorithm relies on Extremely Randomized Trees, a nonparametric tree-based regression method already experimented within a wide range of applications, such as image classification, bioinformatics, environmental modeling, and water reservoirs operation [*Galelli and Castelletti*, 2013]. Tree-based regressors are all based on the idea of decision trees, that are tree-like structures, composed of decision nodes, branches, and leaves, which form a cascade of rules leading to numerical values. The tree is obtained by first partitioning at the top decision node, with a proper splitting criterion, the set of the input variables into two subsets, thus creating the former two branches. The splitting process is then repeated in a recursive way on each derived subset, until some termination criterion is met (e.g., the numerical values belonging to a subset vary just slightly or only few elements remain). When this process is completed, the tree branches represent the hierarchical structure of the subset partitions, while the leaves are

the smallest subsets associated to the terminal branches. Each leaf is finally labeled with a numerical value.

The particular structure of Extra-Trees can be exploited to rank the importance of the n input variables in explaining the selected output behavior on the basis of a score associated to the variance reduction of each input variable [*Galelli and Castelletti*, 2013]. This procedure has two main advantages: *(i)* it can be applied to any kind of dataset since it does not require any assumption on the statistical properties of the input dataset; *(ii)* generally, it is faster and more efficient since it does not rely on computationally intensive methods to estimate the information contained in the data.

3.4 ASSESSING THE VALUE OF SAMPLE INFORMATION

Once the best surrogate of the future external drivers $\mathcal{I}_t$ is obtained, the design of the Improved Operating Policy (IOP) will be carried out, with the decision vector determined as $\mathbf{u}_t = \text{IOP}(t, \mathbf{x}_t, \mathcal{I}_t)$. By properly defining $\mathcal{I}_t$, the IOP should have the potential to fill the gap between the upper and lower bound solutions (i.e., $J^{POP} < J^{IOP} < J^{BOP}$) and, possibly, having J^{IOP} as close as possible to J^{POP}.

In contrast to the identification of the POP performed under the assumption of perfect information of future conditions, in the IOP design problem the trajectory $\varepsilon_{[1,H]}$ is not deterministically known and the external drivers are modeled as external disturbances. The problem formulation is more complex and requires adopting proper optimization tools capable of optimizing close-loop operating policies, possibly conditioned on a large set of information, which is not modeled but directly embedded into the operating policies.

In order to reduce the limiting effects of both the curse of modeling [*Tsitsiklis and Van Roy*, 1996] and curse of dimensionality [*Bellman*, 1956], we need an approximate dynamic programming method that (*i*) is scalable with respect to the state-decision space, thus overcoming the curse of dimensionality; (*ii*) is able to deal with a potentially large number of variables, possibly with different ranges of variability, and (*iii*) allows conditioning the operating policy on any exogenous information in a completely model-free fashion. In this thesis, a policy search method is adopted. Policy search methods rely on a simulation-based optimization approach that first parameterizes the operating policy within a given family of functions and, then, optimizes the policy parameters (i.e., the decision variables of the problem) with respect to the operating objectives of the problem. In addition, this approach can be effectively combined with multi-objective evolutionary algorithms when the problem is characterized by high-dimensional decision spaces, noisy and multi-modal objective functions, and multiple, competing operating objectives [*Giuliani et al.*, 2015a].

It is worth noting that the procedure underlying the ISA framework should not be considered as a sequence of independent steps, but should be applied in an iterative fashion. In particular, it is not guaranteed that the contribution of each variable in explaining the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$ is equivalent to the contribution in the associated operating policy performance. The marginal increment of each selected variable should be therefore verified by simulating the corresponding IOP. At first, we consider only the candidate variable with the highest potential to improve reservoir operations, namely the one associated to the highest ability in explaining the optimal sequence $\mathbf{u}^{POP}_{[0,H-1]}$. The associated Improved Operating Policy is designed and evaluated with respect to the references represented by the performance of the POP and the BOP. The procedure is then iterated by incrementally selecting information, designing the associated Improved Operating Policy, and evaluating the corresponding EVSI. When either the attained performance is satisfactory or the marginal improvement in the EVSI between two consecutive iterations is negligible, the procedure ends.

3.5 METRICS TO ASSESS THE VALUE OF INFORMATION

The quantification of the value of information for the analysis of both the EVPI and the EVSI is straightforward in single-objective problems, where it can be easily measured as the difference in the values of the objective function considered. However, the majority of water resources management problems involve multiple competing objectives [*Kasprzyk et al.*, 2009; *Giuliani, Galelli, and Soncini-Sessa*, 2014b]. For instance, multi-objective reservoir includes hydropower production, flood prevention and water supply. Therefore, the solution of a multi-objective problem is indeed not unique, but rather a set of Pareto optimal (or approximate) solutions. According to *Maier et al.*, 2014, assessing the effectiveness of multi-objective problems' solutions requires to evaluate:

- The distance of the final solutions from the true Pareto front;
- The coverage of the non-dominated space;
- The extent of the non-dominated front.

In this thesis, the Hypervolume (HV) was used in order to capture both convergence and diversity [*Maier et al.*, 2014]. The HV measures the volume of objective space dominated by an approximation set, with HV calculated as the difference in hypervolume between the best known Pareto optimal front (i.e., the set of POPs) and the considered approximation set (i.e., BOPs or IOPs), as shown in figure 3.3. This metric allows set-to-set evaluations, where the Pareto front with the higher HV is deemed the better.

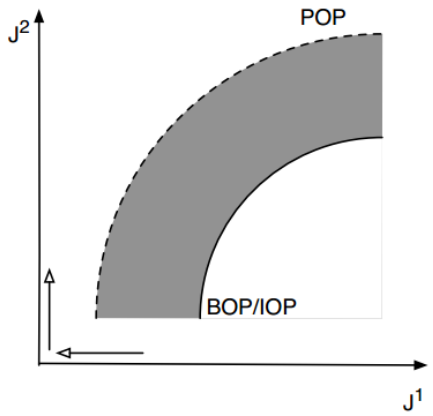


Figure 3.3: Hypervolume indicator for assessing the value of information in multi-objective problems.

4

CASE STUDY

4.1 THE RED RIVER BASIN

The Red River Basin is the second largest basin of Vietnam. It is located between 20°00N and 25°30N, and 100°00E and 107°10E. The total area of the basin is approximately 169,000 km^2, of which 81,240 km^2 (48%) in China's territory, 86,600 km^2 (51.35%) in Vietnam, and the rest in Laos [*Quach*, 2011]. The Red River, the main river downstream from Viet Tri, is composed of three upstream tributaries, originated in China, namely Da, Thao and Lo Rivers (see figure 4.1). Their contributions to the total flow of Red River are 42%, 19% and 25.4%, respectively.

The whole basin is characterized by two distinguished seasons: rainy season from May to October and dry season from November to April of the following year. Annual rainfall varies from 1,200 to 4,800 mm/year in the Vietnamese part of the catchment, and about 80% of rainfall occurs in the rainy season. The flood (high flow) season is from June to October, and low flow season is longer, from November to the next May. Because of uneven rainfall, flows through the basin are unevenly distributed in time, causing floods and water-logging in the rainy season and water shortages in the dry season [*Quach*, 2011].

Among water users, irrigation is the largest one, accounting for 75.5% of total used water. The agricultural land occupies approximately 1,874,100 ha and forestry land occupies 2,570,775 ha. The other users are industry, domestic, livestock and environment (see table 4.1). In order to supply these rapidly increasing water demands, sev-

Table 4.1: Water demand for all sectors. Source: IMRR project.

Water user	Unit	Absolute value	Relative value
Cultivation	$10^6 m^3$	8,271.7	67.35%
Aquaculture	$10^6 m^3$	1,000.5	8.15%
Urban, industries	$10^6 m^3$	1,284.4	10.46%
Rural and domestic	$10^6 m^3$	532.4	4.33%
Water for livestock	$10^6 m^3$	67.9	0.55%
Water for environment	$10^6 m^3$	1,125	9.16%

eral reservoirs have been built. The main ones are Thac Ba reservoir located in Chay River, Tuyen Quang reservoir located in Gam River, Son La and Hoa Binh reservoir located in Da River. Some other reser-

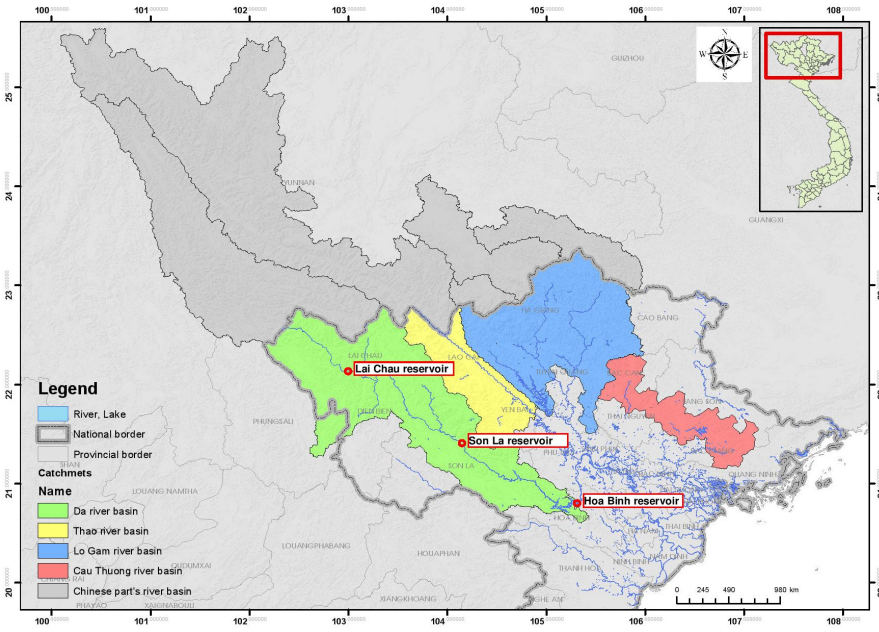


Figure 4.1: The Red River Basin and its main tributaries. Source IMRR project.

voirs have been recently constructed, such as Lai Chau located in Da River, Huoi Quang and Ban Chat located in Nam Mu River.

4.2 EL NIÑO SOUTHERN OSCILLATION IN VIETNAM

There are numerous studies that have identified the ENSO as a significant source of inter-annual hydrometeorological variability in many parts of the world, including Mekong River basin. *Ronghui et al.*, 2004 have revealed that the ENSO cycle strongly influences the strength of the East Asian summer and winter monsoons, and this is well reflected in both the inter-annual variations of summer rainfall anomalies and the summer and winter monsoon circulation anomalies in East Asia. According to *Hapuarachchi*, 2005, El Niño events in Mekong River basin are typically associated with lower than average precipitation and discharge during the Northern Hemisphere summer and autumn (and sometimes winter). ENSO has minimal impact on spring (March to May) precipitation, and discharge that occurs as a result of spring precipitation, because this period usually represents a transitional stage in the ENSO cycle. More recently, *Räsänen and Kummu*, 2013, found that there is strong but lagged relationship between ENSO and the hydrology in the Mekong River basin. The precipitation and discharge decrease during El Niño events and increase during La Niña events. Nevertheless, this correlation was not uniform throughout the whole basin and varied in the period of research. These works suggest a potential for developing a prediction method for ENSO induced hydrological anomalies, as ENSO index values from December-

February months explained approximately 50% of the inter-anual variation of the Mekong's discharge. Consequently, it may significantly improve the possibility of prediction and preparing for extreme hydrological events, such as floods and droughts, and thus reduce the negative impacts on infrastructure and people's livelihoods by improving water resources management.

4.3 MAIN STAKEHOLDERS

4.3.1 *Flood*

Ha Noi and its unusually overdeveloped River Urban Area are protected by a system of two series of dykes for a total length of 2,700 km. There are 21 main dike systems with the total length of 533.9 km (see figure 4.2), which are classified into 5 levels: special level (37.7 km), 1^{st} level (249.187 km), 2^{nd} level (45.006 km), 3^{rd} level (72.165 km) and 4^{th} level (129.786 km).

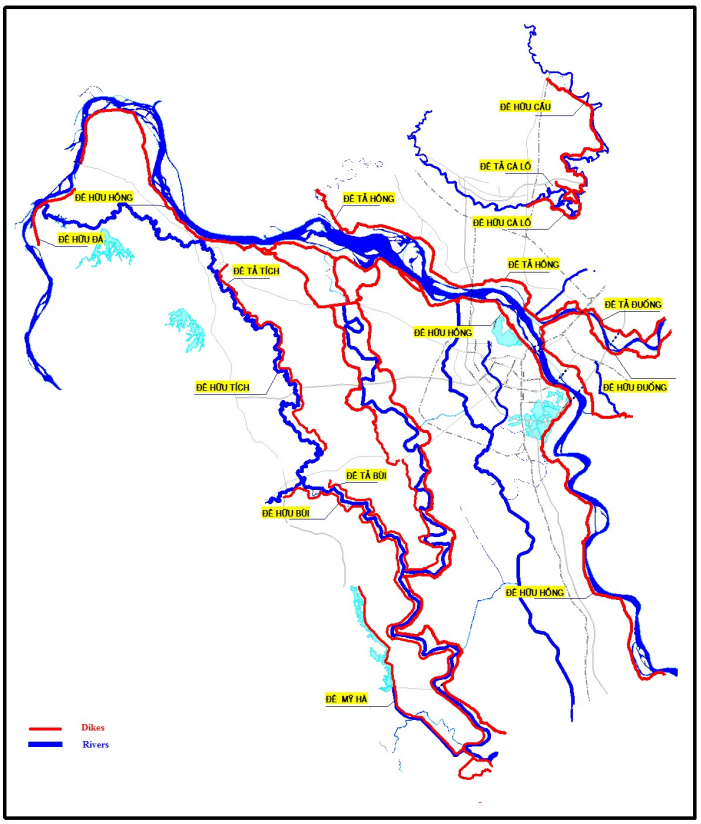


Figure 4.2: The dike system in Ha Noi.

Even with the protection of this dyke system, floods still happened in Ha Noi. Some historical flood events were recorded in Red River basin from 1913 to 2010, e.g., those happened in 1913, 1915, 1945, 1968, 1971, 1978, 1986, 1996, 2002, 2008 and 2010. For instance, in 1971,

Table 4.2: Historical flood events at some stations in the Red River Basin.

No.	Station	River	Average flow m^3/s	Peak flow m^3/s	Date
1	Son Tay	Red	16,785	37,800	Aug-1971
2	Lai Chau	Da	7,242	14,200	Aug-1932
3	Ta Bu	Da	9,919	22,700	Aug-1996
4	Hoa Binh	Da	9,618	22,700	Aug-1996
5	Yen Bai	Thao	5,143	10,350	Aug-1971
6	Ham Yen	Lo	2,897	5,700	Jul-1986
7	Chiem Hoa	Gam	3,188	6,220	Aug-1971
8	Tuyen Quang	Lo	5,156	11,700	Aug-1971
9	Vu Quang	Lo	5,467	14,000	Aug-1971

from August 12^{th} to 21^{st}, due to combined effects of the tropical converged range, Western low grooves combined with Pacific high pressure, very heavy rain happened in the whole basin. The average rainfall was 255 mm in the Red River basin and was 200 mm in the Red River Delta. On the Da River, at Hoa Binh station, the flow was 14,800 m^3/s. On the Thao River, the flow at Yen Bai was 10,530 m^3/s, and on the Lo River at Vu Quang, the flow reached 14,000 m^3/s. The flood of the Lo River was the biggest one in history. The flood peak in Ha Noi was 14.13 m which excess 1.63 m and 2.63 m above the third and the second alarming levels, respectively [*Quach*, 2011]. According to the price of 1971, total property damage to the State sectors under management of the Central Government was more than VND 44 billion. In addition, 100,000 deaths and flood effects such as epidemics and interrupted production could not be counted out.

Another major flood record is the one that happened in 1996 despite the Hoa Binh reservoir was already in operations for several years. From 9^{th} to 20^{th} August, 1996, it kept raining in the Red River basin. During this period, the average rainfall of Red River basin was 432 mm; that figure of the Da, Thao and Lo River basin was 380 mm, 317 mm, and 349 mm, respectively. The flood peak measured in Ha Noi was 12.43 m (equivalent to 13.30 m if converted [1]) at 21 o'clock 19^{th}, 0.93 m above the third alarming level, and lasted for 6 days. The peak flow at Ha Noi was 14,800 m^3/s and at Son Tay was 27,400 m^3/s. There were four people died, 61 missed and 161 injured. Houses, schools, clinics and hospitals were collapsed for a couple of days.

1 'converted' here means that if dikes had not broken, the water level could have reached 14.8 m

4.3.2 *Water shortages*

In the Red River basin, water shortages often happen in the dry season (from November to the next May), and they become more serious in recent years. The minimum water level in the Red River at Ha Noi station has been getting lower and lower. It was 2 m, 1.94 m, and 1.46 m in 1999, 2004 and 2006, respectively. Especially, it reached 0.66 m in the end of March, and 0.4 m in the beginning of April in 2010, the latter being the historical event in the last 100 years. The minimum discharge at Son Tay in March 2002 was only 380 m^3/s, while total water demand this time was about 500 m^3/s, leading to a lack of more than 10 million m^3/day.

The most important reason is the uneven distribution of the rainfall. The total rainfall in the dry season (7 months) accounts for approximate 24% of the total rainfall in the year. The lowest flow occurs in February or March when the irrigation demands increase dramatically due to the rice cultivation period. The second reason is due to the abnormality of the weather, specifically the *La Niña* phenomenon happening with the cycle of 2 to 7 years. The third reason is erosion of the rivers due to the operation of Hoa Binh reservoir (see section 4.3.5). Forest overexploited and deforestation is the man-made reason that reduces the ground water table.

The last but not least is the low inflow from China. According to the National Center for meteorology and hydrology, the inflows from China in all three tributaries are at the lowest levels in history. The minimum flows in November 2009 and March 2010 on the Da River at Lai Chau station are 23 m^3/s and 56 m^3/s, respectively, while the corresponding lowest levels in the past were 332 m^3/s and 103 m^3/s. These figures in the Thao River at Lao Cai station were 141 m^3/s and 130 m^3/s in comparison with the lowest ones of 161 m^3/s and 97 m^3/s in the past [*Quach*, 2011]. This effect is due to the construction and operation of new dams in China. China has more than 5,600 large dams in operation or under construction that are 30 meter or higher. According to the First National Census for Water, there were about 38,000 dams higher than 15 meter in China at the end of 2012. There were 144 dams 30 meter or higher under construction in 2012, of which 107 would be more than 60 meter high [*Hydropower & Dam*, 2014].

4.3.3 *Electricity*

Hydropower is the main energy source of Vietnam; it contributed from 35 to 40% of total energy consumed. Hoa Binh hydropower plant was the biggest one that annually supplied about 8 billion kWh since 1994, accounting for 35% energy of the northern part, and 15% of the country. In 20 years of operation, the yearly energy production

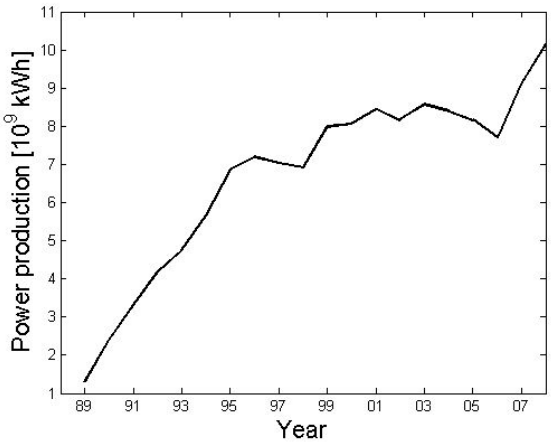


Figure 4.3: Energy production of Hoa Binh reservoir from 1989 to 2008.

has increased from 1.3 TWh in 1989 to 10.136 billion kWh in 2008 (see figure 4.3).

The electric demand of Vietnam has been increased about 15 to 20 % per year. Vietnam has started to import electricity from China since 2005. In 2011, it is estimated that the electricity deficit is about 4.67 TWh, and it is also forecasted that hydropower electricity continues contributing about 33% of total energy production of Vietnam [*Pham, Nguyen, and Nguyen*, 2011]. Therefore, Hoa Binh hydropower plant will keep playing the most important role in supplying hydropower to the electric system of the country. However, balancing the energy production with other conflicting interests will be a great challenge for the decision makers in designing the reservoir operations.

4.3.4 *Navigation and water pollution*

The water level required for local navigation on the Red River is 2.15 m at Ha Noi station. Water level data from 1956 to 2004 shows that on average the annual number of days in which the water level was below 2.15 m is 10 days. Particularly, this situation lasted for 62 days and 23 days in 1963 and 2004, respectively. These interruptions caused a considerable cost to the transportation sector and its users.

Water pollution in the river network is another issue related to the decline of water level. The observation of flow rate shows that less water have been diverted to Nhue River leading to environmental damages in Nhue River basin because of lack of water for dilution. Due to drought events as well as the lack of control of management, the Minimum Environmental Flow have not been guaranteed.

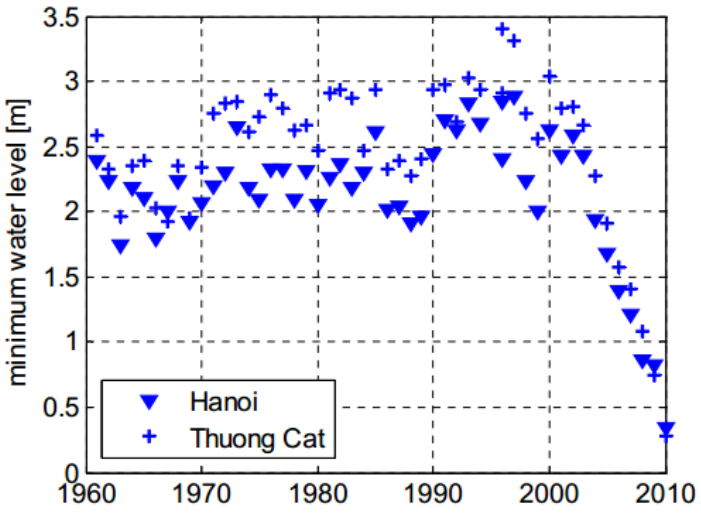


Figure 4.4: Minimum water levels recorded at Ha Noi station and Thuong Cat station.

Table 4.3: Water level vs discharge at Son Tay cross-section over 3 decades.

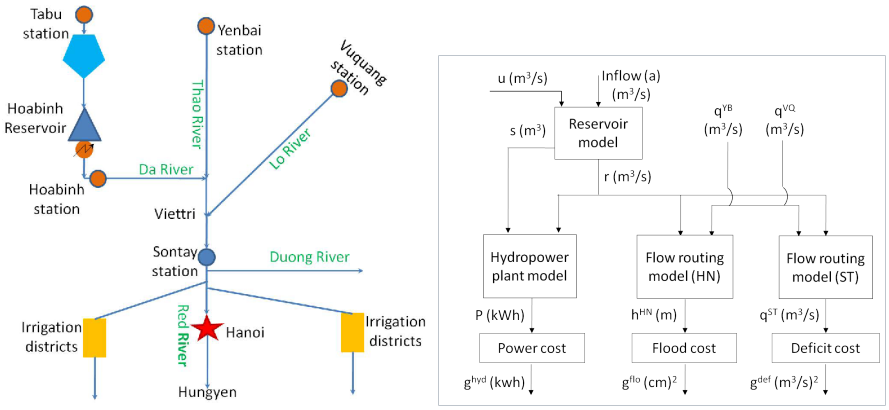
Period	Discharge (m^3/s)	Level (m)	Discharge (m^3/s)	Level (m)
1971-1979	1,000	5.67	2,000	7.07
1981-1989	1,000	5.69	2,000	7.05
1991-1998	1,000	5.25	2,000	6.65

4.3.5 *River bed erosion*

Before 1989, the cross-sections of downstream river were getting smaller due to sedimentation. With the same discharge, the water level increased from 20 to 30 cm in comparison with the previous decade. However, due to the construction and operation of Hoa Binh reservoir, downstream river bed has been eroding. The water level in the downstream has been lower in the dry season even with higher release regulated by Hoa Binh. With the same discharge of 1,000 m^3/s, water level at Son Tay has decreased 0.44 m from 5.69 m to 5.25 m (see table 4.3). The statistical observation illustrates that the minimum water levels recorded at Ha Noi station and Thuong Cat station are decreasing (see figure 4.4). The noticeable acceleration in river bed degradation is a danger for agriculture water supply as well as for infrastructure stability.

4.4 MODELLING OF THE SYSTEM

This thesis is focused on studying the operations of the Hoa Binh reservoir, one of the largest reservoirs in the RRB which is in operations for 20 years. The upstream catchment, the reservoir, including hydropower, and the downstream river network are the three main



(a) Scheme **(b)** Model

Figure 4.5: The scheme of the Hoa Binh water system and of its model

components of the Hoa Binh water system, which are described in details in the following sections.

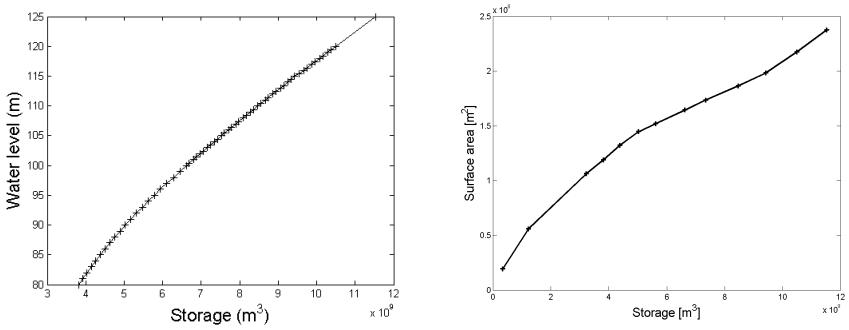
4.4.1 *Upstream catchments*

The three main upstream catchments of the Red River Basin are the Da, Thao and Lo catchments.

The Da River originates from the mountainous Nguy Son region in the South-Western part of Yunnan province in China at an altitude of 2,440 m above the mean sea level. From its source it travels a distance of about 925 km before joining the Thao and Lo River at Viet Tri, about 60 km downstream of the Hoa Binh Dam. The main tributaries of Da River in Vietnam area are Nam Na and Nam Mu on the left and the Nam Po and Nam Muc rivers on the right. The basin area is 52,900 km^2, the largest tributary of the Red River, accounting for 37% of the Red River basin area at Son Tay, of which the area inside the territory of Vietnam is 26,800 km^2.

The Thao River enters Vietnam at Lao Cai in Lao Cai province. The total catchment area of the Thao River is 51,800 km^2, of which 12,000 km^2 belongs to Vietnam. The key location in the Thao basin upstream of the confluences with the Da and Lo Rivers is the Yen Bai station, covering 93% of the flow of the total sub-basin.

The Lo river basin area is 39,000 km^2, of which 22,000 km^2 located in Vietnam. The river basin consists of three major tributaries: the Lo River, the main stream; the Gam river, entering Lo River some 10 km to the north of Tuyen Quang; and the Chay River, joining Lo River at 25 km downstream of Tuyen Quang. The key location in the Lo-Gam-Chay sub-basin is Vu Quang, covering 95% of the flow of the total sub-basin.



(a) WL vs. storage **(b)** surface area vs. storage

Figure 4.6: Characteristic curves of Hoa Binh reservoir.

4.4.2 *The Hoa Binh reservoir*

Hoa Binh reservoir was constructed on the Da River, in Hoa Binh province, in 1979 and was completed in April 1994. This is a multi-purpose reservoir, including hydropower production, flood mitigation and water supply for Red River delta. The technical designed parameters of the dam as well as flood discharge structures are listed it in table 4.4.

During the flood season the management of the dam must comply with the guideline reported in the *Decision No.198/2011/QD-TTg* issued on February 10^{st}, 2011 by the Prime Minister of Vietnam.

4.4.2.1 *Characteristic curves of the Hoa Binh reservoir*

Relationship between the water level and the storage

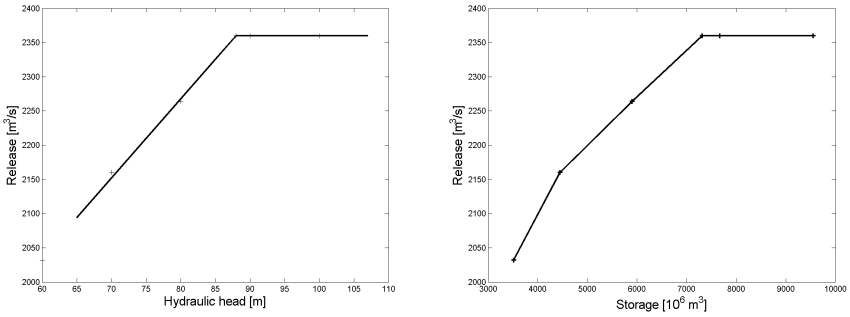
The water level of the Hoa Binh reservoir ranges from the dead level of 80 m to the maximum level of 122 m, and the corresponding storage varies from 3.8 billion m^3 to about 9.862 billion m^3. From available bathymetry data (pairs of water levels and corresponding storage values), two functions have been built. The first one allows to compute the water level of the reservoir (h^{up}) given the storage (s). The second is the inversion of the first, allowing to estimate the storage (s) given the water level (h^{up}). The relationship between the storage and the water level is showed in figure 4.6a.

Relationship between the surface area and the storage

Analogously, the *surface area* (S) of the reservoir increases from 100 km^2 to more than 200 km^2 in the range of the storage variation. A piecewise linear function was identified based on the available data to estimate the surface (S) given the value of the storage (s) (figure 4.6).

Table 4.4: Designed paramters of Hoa Binh reservoir.

No	Parameter	Unit	Value
1	**The parameter of the reservoir**		
1.1	The area of the catchment	km^2	51,700
1.2	The annual flow	m^3	1,781
1.3	The designed flood peak flow rate correspoding probaibility P = 0.01%	m^3	49,000
1.4	The checked flood peak flow rate	m^3/s	63,000
1.5	The normal Water Level (WL)	m	117
1.6	The dead WL	m	80
1.7	The WL before flood	m	98.8
1.8	The excess WL	m	122
1.9	The WL correspoding checked flood	m	122
1.10	The total storage	10^6 m^3	9,862
1.11	The active storage	10^6 m^3	6,062
1.12	The dead storage	10^6 m^3	3,800
1.13	The storage for flood prevention	10^6 m^3	3,000
1.14	The surface area corresponding to normal WL	km^2	208
2	**The parameter of the main dam**		
2.1	Type of dam		Soil and rock
2.2	The maximum height	m	128
2.3	The length of the dam	m	660
2.4	The level of crest	m	123
3	**The flood discharge structure**		
3.1	The maximum flood discharge	m^3/s	41,240
3.2	The level of spillway	m	102
3.3	The size of spillway	n(BxH)	6(15x13)
3.4	The level of bottom gate	m	56
3.5	The size of bottom gate	n(BxH)	12(6x10)
4	**The hydropower plant**		
4.1	The number of uint	unit	8
4.2	The maximum flow rate pass through the turbines	m^3/s	2,400
4.3	The designed hydraulic head	m	88
4.4	The maximum hydraulic head	m	105
4.5	The minimum hydraulic head	m	60.5
4.6	The total designed power	MW	1,920
4.7	The guaranteed power	MW	707 (+123)
4.8	Type of turbine		Francis
4.9	The average annual power	10^6 kWh/year	9,132(+1,267)



(a) Release through penstocks vs. hydraulic head

(b) Release through penstocks vs. storage

Figure 4.7: The release curves through penstocks.

4.4.2.2 *Rating curves of the Hoa Binh reservoir*

Penstocks

Turbines are fed by penstocks which are located below the dead water level of 80 m. The release through penstocks is controlled individually by a valve that can be opened at any aperture. The release varies from zero to a maximum of 295 m^3/s (equivalent to $q_t^{t,max}$ = 2,360 m^3/s for eight turbines) based on the water level of the reservoir.

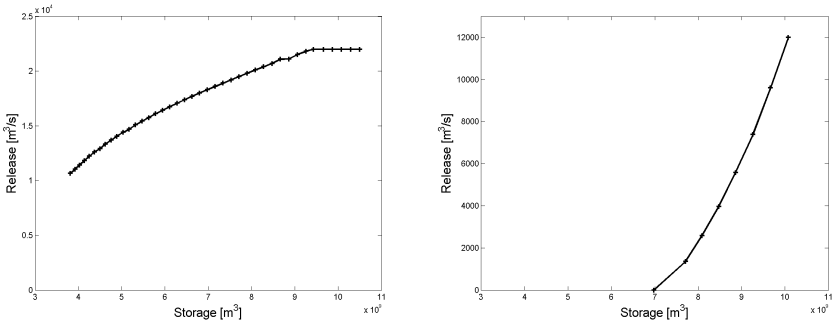
The maximum release through the penstocks is thus given by

$$r_t^{max} = \begin{cases} 0.12538s + 1603.1 & \text{if } s \leqslant 4441.4 \\ 0.06959s + 1851.6 & \text{if } 4441.4 < s \leqslant 7306 \\ 2360 & \text{otherwise} \end{cases} \tag{4.1}$$

where s is the storage (10^6 m^3).

Bottom gates

In order to control the water level, the Operator may activate 12 *bottom gates* and 6 *spillways* to increase the outflow. While turbines allow continuous control (i.e., each release between 0 and 2,360 m^3/s is allowed), bottom gates and spillways can only be completely opened or closed except for one bottom gate which can be opened with three apertures. The only rule to respect strictly is that the first six bottom gates that are opened or the last six that are closed have to be operated with 6 hours of gap between each operation, the so called *6 hour-rule*. The purpose of this rule is to avoid too fast variation of downstream flow, considering that discharge through two bottom gates (1,000 - 1,833 m^3/s each) is of the same magnitude of the discharge through the eight turbines (2,360 m^3/s). However, once the first six bottom gates are opened, any operation of further bottom



(a) bottom gate **(b)** spillways

Figure 4.8: Rating curve of Hoa Binh reservoir.

gates and spillways is allowed. The rating curve of the bottom gates (maximum release as a function of the storage) is shown in figure 4.8*a*.

$$r_b^{max} = 0.2831 \times 10^{-8} s^3 - 0.6996 \times 10^{-4} s^2 + 0.7012 s - 920.44 \quad (4.2)$$

where s is storage ($10^6 m^3$).

Spillways

The spillways are located at elevation of 102 m to the normal water level of the reservoir. The release of one spillway varies from zero to 2,350 m^3/s depending on the hydraulic head, for example, the water level. Therefore, the total possible maximum release through spillways reaches 14,100 m^3/s. The release function is represented as equation (4.3), and the data points are shown as figure 4.8*b*.

$$r_s^{max} = -0.0745 \times 10^{-7} s^3 + 0.3182 \times 10^{-3} s^2 - 3.1398 s + 8940.6 \quad (4.3)$$

where s is storage ($10^6 m^3$).

4.4.2.3 *Mass balance equation of the reservoir*

The water volume of the reservoir at time t is a function of the inflow, the amount of evaporation and the volume delivered through regulation structures and spillways in the time interval $[t, t+1)$. The *inflow* a_{t+1} of the reservoir is a function of the cumulative water volume from its tributaries, the distributed runoff along its banks and the direct precipitation on the water surface. The *evaporated volume* E_{t+1} is proportional to the *area* S_t *of the water surface* and the *specific evaporation* volume e_{t+1}. The first depends on the storage, and the second on a complex multitude of factors (water temperature, air temperature, relative humidity, atmospheric pressure, wind velocity and relative insolation). The released volume, simply called *release* and denoted

with r_{t+1}, is the entire volume released through penstocks, bottom gates and spillways. Instead, the control variable is the *release decision* u_t (simply called *control*) that coincides with r_{t+1} when the spillways or the bottom gates are not operating. Therefore, the mass balance equation of the reservoir is:

$$s_{t+1} = s_t + (a_{t+1} - r_{t+1})\,\Delta - E_{t+1} \tag{4.4}$$
$$E_{t+1} = e_{t+1} S(s_t) \tag{4.5}$$
$$r_{t+1} = R(s_t, u_t, a_{t+1}, e_{t+1}) \tag{4.6}$$

where $\Delta = 60 \times 60 \times 24$ s/day. The evaporation function (4.5) and the release function (4.6) are discussed in the following paragraphs.

4.4.2.4 *Model of the hydropower plant*

The daily energy production P_{t+1} (kWh) of the Hoa Binh hydropower plant is given by the product of the release through the turbines r^t_{t+1}, the hydraulic head H_{t+1}, and the turbine efficiency η, i.e.,

$$P_{t+1} = \varphi\ g\ \gamma\ \eta_{t+1}\ r^t_{t+1}\ H_{t+1} \tag{4.7}$$

where φ is a coefficient of dimensional conversion, g is gravitational acceleration (equal to 9.81 m/s^2) and γ is water density (equal to 1,000 kg/m^3).

The hydraulic head H_{t+1} is the difference between the water level upstream (h^{up}_t) and downstream (h^{do}_{t+1}) of the reservoir. The minimum release through turbines recorded in history is 38 m^3/s. Therefore, the release through turbines r^t_{t+1} will be zero if total release r_{t+1} of the reservoir is smaller than 38 m^3/s, will be equal to r_{t+1} if $r_{t+1} \leqslant$ 2,360 m^3/s, and $q^{t,max}$ (2,360 m^3/s) otherwise, i.e.,

$$r^t_{t+1} = \begin{cases} 0 & \text{if } r_{t+1} < 38 \\ \min(r_{t+1}, q^{t,max}) & \text{otherwise} \end{cases} \tag{4.8}$$

4.4.3 *Downstream river network*

The model of the downstream river network is approximated by two data-driven functions as follows:

$$q^{ST}_{t+1} = f_1(r_{t-\tau+1}, q^{YB}_{t-\tau+1}, q^{VQ}_{t-\tau+1})$$
$$h^{HN}_{t+1} = f_2(r_{t-\tau+1}, q^{YB}_{t-\tau+1}, q^{VQ}_{t-\tau+1})$$

where r_t is release of Hoa Binh reservoir, q^{YB}_t and q^{VQ}_t are flow at Yen Bai and Vu Quang respectively; τ is the regression order.

Conceptually, the regression order (τ) can be chosen between zero and infinite. Practically, the value of τ should be small enough to keep at minimum the state of the system. It is seen that it takes more or less

one day (one step) for water to travel from Hoa Binh reservoir, Yen Bai, and Vu Quang to Son Tay and as well as to Ha Noi, i.e., τ can be set in between 0 and 1. If $\tau = 0$, the downstream models have no state variables. If $\tau = 1$, three lagged flow variables are added among the network inputs and consequently the global model has three more state variables [*Quach*, 2011]. In order to find a balance between the need for model accuracy and the reduction of model complexity, the first option ($\tau = 0$) was chosen in this study.

The downstream river network models are identified using Artificial Neuron Networks. The architecture of the neural networks comprises two levels: a hidden layer, constituted of n sigmoid neurons, and an output layer, composed of one neuron. The activation functions associated to the neurons in the first hidden layer are hyperbolic tangents, while in the output layer the activation function is linear. The downstream models thus take the form:

$$q_{t+1}^{ST} = a_1 + \sum_{i=1}^{n} b_{1i} \, tansig(c_{1i} r_{t+1} + d_{1i} q_{t+1}^{YB} + e_{1i} q_{t+1}^{VQ} + f_{1i}) \tag{4.9a}$$

$$h_{t+1}^{HN} = a_2 + \sum_{i=1}^{n} b_{2i} \, tansig(c_{2i} r_{t+1} + d_{2i} q_{t+1}^{YB} + e_{2i} q_{t+1}^{VQ} + f_{2i}) \tag{4.9b}$$

4.4.4 *Operating objectives*

4.4.4.1 *Hydropower production*

The Vietnamese electricity market is regulated by the Government, and the energy sold at a fixed rate is decided on the basis of the average energy production cost and the current economic development strategy. The price changes depend on the type of energy users (industrial or domestic use) and the total energy consumed but not within the day or the week. It means that maximizing the energy production is equivalent to maximizing the associated revenue. The energy production strongly depends on the availability of the water in the reservoir, for example, there is an increasing frequency of power shortages in the last three months of the dry season, from April to June. The immediate cost, g_{t+1}^{hyd}, is defined as the daily energy production P_{t+1} [kWh] as the following formula:

$$g_{t+1}^{hyd} = -P_{t+1} \tag{4.10}$$

It is worth noting that the production in (4.10) changes in sign because the optimization problem is formulated as a minimization problem (see equation (3.1)). Finally, the objective function for hydropower

production, J^{hyd} [kWh/day], is defined as the daily average of the immediate cost g^{hyd}_{t+1} ,i.e.,

$$J^{hyd} = \frac{1}{H}\sum_{t=0}^{H-1} g^{hyd}_{t+1} \tag{4.11}$$

4.4.4.2 *Flood mitigation*

Floods mainly occur in July and August and inundations produce enormous damages every time dykes break, as regularly happened nearly once per decade in the last century [*Quach*, 2011]. In principle, an accurate modelling of flood inundations and the associated damages require to combine a 2D model of the floodplain to estimate the flooded surface area and a record of past flood recovery costs and associated river flow rates to interpolate the corresponding damages. Practically, the size and shape of the floodplains are quickly changing due to the fast development of River Urban Area, thus making totally incomparable damages registered in different years. Therefore, damages can not be included as an indicator in the decision model.

Nevertheless, it comes without saying that high and persisting flood water levels in Ha Noi correspond to the high risk of dike break, and consequently high potential damages. Therefore, an indirect way of accounting for flood damages is to penalize operating policies that produce river water levels higher than some appropriately selected threshold. Once again, economic relevance and risk perception are implicitly accounted for using some empirical coefficients: the higher potential damage of floods in August on the summer-autumn crop is given a higher weight, while the increased stakeholders' risk aversion to extreme flood is modelled by using a power law [*Quach*, 2011]. The immediate cost has the following form:

$$g^{flo}_{t+1} = \begin{cases} 0 & \text{if } h^{HN}_{t+1} \leqslant \bar{h} \\ (h^{HN}_{t+1} - \bar{h})^m & \text{otherwise} \end{cases} \tag{4.12}$$

where h^{HN}_{t+1} is the water level [cm] at Ha Noi station, $\bar{h}$ (950 cm) is the 1^{st} alarm flood level [*Hansson and Ekenberg*, 2002], and m is the coefficient reflecting the risk aversion here assumed to be equal to 2. Choosing m to be equal to 2 will emphasize the higher risk aversion and the higher damages corresponding to the higher water level excess.

Finally, the indicator for flood mitigation sector, J^{flo} [cm^2/day], is defined as the average of the immediate cost g^{flo}_{t+1} ,i.e.,

$$J^{flo} = \frac{1}{H}\sum_{t=0}^{H-1} g^{flo}_{t+1} \tag{4.13}$$

4.4.4.3 *Water supply*

Wet-rice agriculture is key to national food security but also the most important segment of the Vietnamese economy [*Quach*, 2011]. Red River Basin (RRB) is the second largest rice production area, which is composed of 31 irrigation schemes with total area of 850,000 ha, in the country after the Mekong Delta. In this study, the average water deficit is used as a proxy of the annual crop yield. This is a provably reasonable hypothesis under the assumption that the considered operating policies will not move too much away from the current average water supply [*Soncini-Sessa, Weber, and Castelletti*, 2007]. Farmers are not sensitive to the magnitude of the daily deficit, for example, several small deficits might be more acceptable than one single severe shortage that might strongly affects the crop production. Therefore, a behavioral coefficient n is used to characterize farmers' risk aversion. The value $n = 1$ means no risk aversion, while the value $n \to \infty$ means maximum aversion. Correspondingly, the immediate cost for the water supply is formulated as a power function:

$$g^{irr}_{t+1} = \begin{cases} 0 & \text{if } q^{ST}_{t+1} > w_t \\ (w_t - q^{ST}_{t+1})^n & \text{otherwise} \end{cases} \tag{4.14}$$

where w_t and q^{ST}_{t+1} [m^3/s] are the daily water demand (given as shown in figure 4.9) and supply at Son Tay. The power n is fixed equal to 2, which ensure that vulnerability is a minimum [*Hashimoto, Stedinger, and Loucks*, 1982].

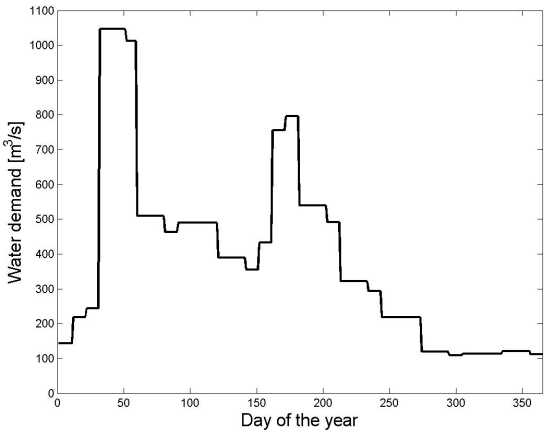


Figure 4.9: Yearly pattern of water demand at Son Tay.

The computation of water demand takes into account all water use sectors, including industrial, domestic use, irrigation, livestock, aquaculture, navigation and environment (see table 4.1). The water demand for industry was estimated by assuming a specific water demand of 1,000 m^3/1,000 USD for food industry, 400 m^3/1,000 USD for

light industry, and 200 $m^3/1{,}000$ USD for heavy industry. The water demand for domestic use is about 150 liters per person per day for the urban area, and 60 liters per person per day for the rural area. The water demand for livestock is computed based on the the water criteria per head per day, depending on the size of the animal. The water requirements for aquaculture is 20,000 m^3 per ha per month. The water demand for irrigation is computed using the CROPWAT software. There is no official criteria for environmental water requirements. In the project, it is set from 10 to 30 % of total other demands. And the water requirement for navigation is transformed from the water level required at some certain points along the Red River.

Finally, the indicator for water supply sector, J^{irr} $[(m^3/s)^2/day]$, is defined as the average of the immediate cost g^{irr}_{t+1} ,i.e.,

$$J^{irr} = \frac{1}{H} \sum_{t=0}^{H-1} g^{irr}_{t+1} \tag{4.15}$$

5

NUMERICAL ANALYSIS AND RESULTS

5.1 EXPERIMENTAL SETTING

The aim of this thesis is to apply the ISA framework (chapter 3) to demonstrate the potential for using the ENSO information (chapter 2) in the operation of the Hoa Binh reservoir (chapter 4). The main steps of the framework are summarized in figure 5.1.

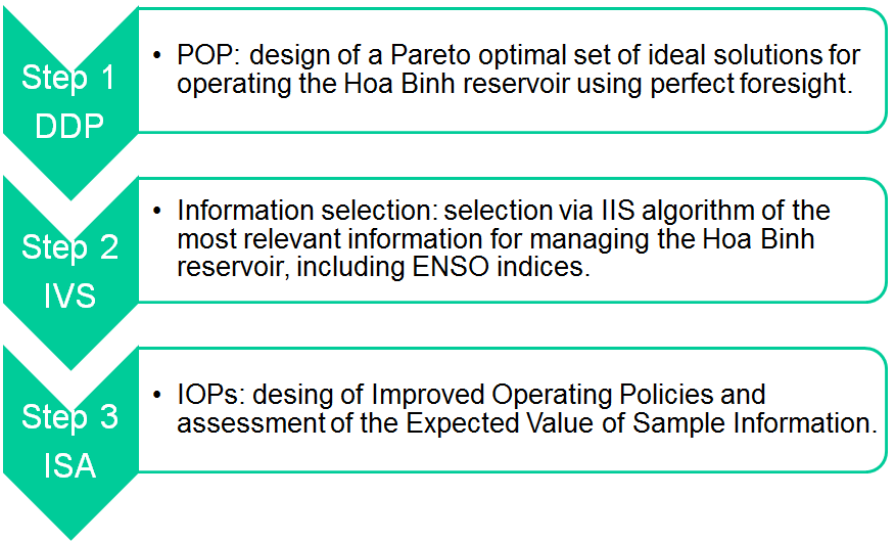
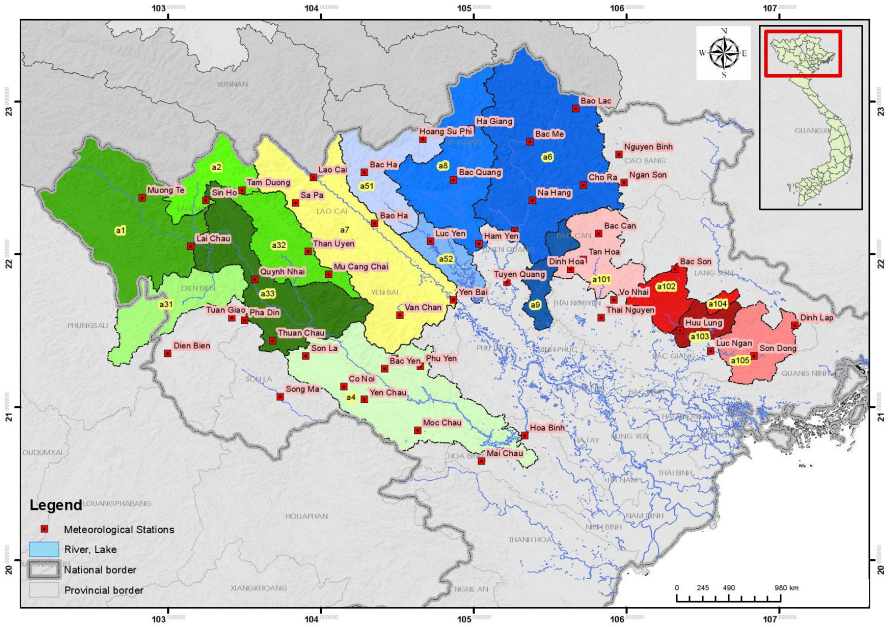


Figure 5.1: Summary of the ISA framework used in this work.

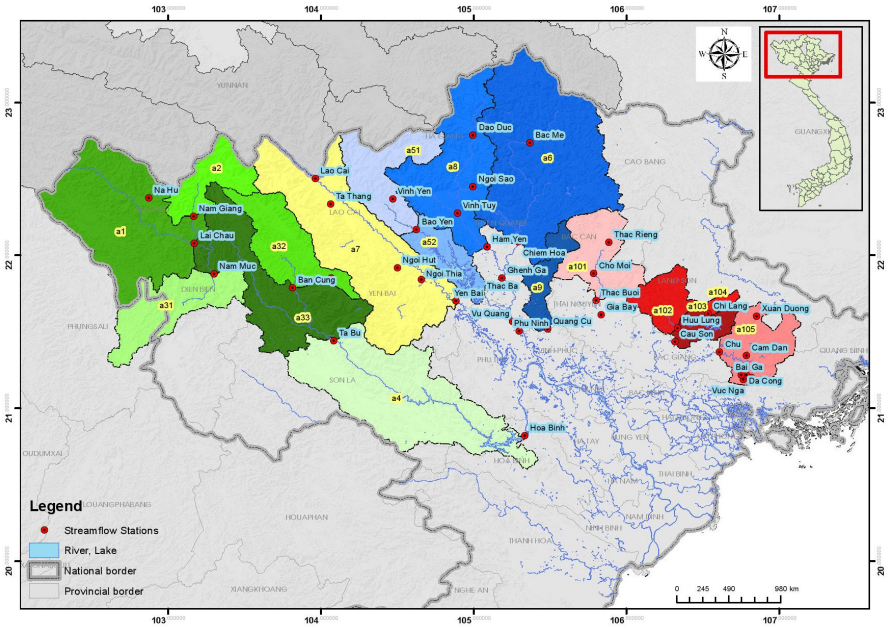
5.1.1 *Hydro-meteorological data*

In Red River basin, there is a distributed hydro-meteorological monitoring system which observes and measures a variety of hydrological and meteorological variables. This system provides an accurate and reliable dataset, including meteorological observations (e.g., air temperature [°C], relative humidity [%], wind [m/s], precipitation [mm]) and hydrological observations (e.g., water lever [m], flow [m^3/s]). To cover as much as possible the entire Red River basin, in this work we used the meteorological and hydrological stations shown as red squares in figure 5.2*a* and red circles in figure 5.2*b*, respectively.

The definition of the set of candidate exogenous variables relies on the outcome of some previous studies such as IMRR project and *Beltrame and Carbonin*, 2013. , which identified the variables that mostly affect the hydrology in the system. These variables are reported in table 5.1 and table 5.2.



(a) Meteorological stations



(b) Hydrological stations

Figure 5.2: Hydro-meteorological stations in Red River basin. Source: IMRR project.

Table 5.1: Selected meteorological stations and selected variables.

Location	Station	Variable
Da river basin	Muong Te	Precipitation
Da river basin	Tam Duong	Precipitation
Lo Gam river basin	Bao Lac	Precipitation
Lo Gam river basin	Bac Me	Precipitation
Lo Gam river basin	Ha Giang	Precipitation
Da river basin	Spatial/average	Precipitation

Table 5.2: Selected hydrological stations and selected variables.

Location	Station	Variable
Da river basin	Hoa Binh	Flow
Da river basin	Ta Bu	Flow
Da river basin	Lai Chau	Flow
Da river basin	Nam Giang	Flow
Thao river basin	Yen Bai	Flow
Lo Gam river basin	Vu Quang	Flow

5.1.2 *ENSO data*

In the absence of a previous study on the ENSO effects in Red River basin, there is no consensus on the classification of El Niño and La Niña for Vietnam as well as which ENSO indices mostly affect the hydrological system in Vietnam . *Räsänen and Kummu*, 2013 in a study conducted over the Mekong River suggested to use the ONI index while *Ronghui et al.*, 2004 suggested to use SST anomalies. However, ONI is an elaboration on SST3.4 indices (see section 2.4.3). For this reason, in this work we decided to use the SST3.4 anomaly. The data, which are available from 1950 to present, have been obtained from NOAA's website.

5.1.3 *Perfect Operating Policy design*

TIME PERIOD: The construction of Hoa Binh reservoir was completed in 1994 and the full operation of reservoir was started in 1995 when it was completely filled up. Therefore, to construct the POP solution, the period from January 1st, 1995 to December 31st, 2006 was chosen. This time horizon is used for both the optimization and evaluation of the policy performance, thus en-

suring that the POPs represent the upper bound in terms of system performance.

DISCRETIZATION: The application of Deterministic Dynamic Programming (DDP) requires that at every moment in time, the state (storage) and the control (release decision) belong to a finite set [*Soncini-Sessa, Weber, and Castelletti*, 2007]. Our original system does not comply with the condition, so a discretized model must be derived by discretization. The denser the discretization is, the better the optimization results will be attained, but the longer the computing time is. It is thus necessary to accept a trade-off between accuracy and computational complexity. The storage of Hoa Binh reservoir ranges from 3.8×10^9 m^3 to 9.862×10^9 m^3, corresponding to the dead water level (80 m) and the maximum water level (122 m). This physical interval of storage is discretized into 67 values with an increment of 10^8 m^3. To avoid water level falling below the dead water level, one more point is added to the left of the minimum value of the storage vector, which is associated with a maximum possible release of zero. So that, by construction, the storage cannot go below this value. For the discretization of release decision, we set an uniform grid of 28 equally spaced value of the decision, with an increment of 500 m^3/s, ranging from 0 to 13,000 m^3/s.

COMBINATION OF THE WEIGHTS: As the discussion in section 4.4.2, Hoa Binh reservoir is a multi-purpose reservoir. As a result, solving the policy design problem via DDP requires to aggregate the 3 objectives via the weighting method. The higher value of the weight is, the higher concern about this objective is. For example, the value of zero means that we neglect this objective whereas the value of one means that we concentrate on this objective and neglect all others. In particular, we used twenty-five combinations of weights reported in table 5.3 that ensure a good coverage of the tradeoffs surface.

WATER DEMAND: As the discussion on section 4.1, there is an increase in the water demand due to the growing population in the Red River delta. Moreover, due to the erosion of the river bed, as the discussion on section 4.3.5, the real water demand is much higher than the one measured and calculated at Son Tay station. To take into account this phenomenon, the water demand used in this analysis was increased (see figure 5.3).

5.1.4 *Input Variable Selection experiments*

INPUT VARIABLES: The input variables include the day of the year (t), the storage obtained from DDP experiment (s_t), hydro-meteorological variables (see chapter 5.1.1) and ENSO variables (see chapter

Table 5.3: The combination of weights.

Policy	J^{hyd}	J^{irr}	J^{flo}	Policy	J^{hyd}	J^{irr}	J^{flo}
1	0.95	0.03	0.02	14	0.3	0.19	0.51
2	0.9	0.04	0.06	15	0.25	0.2	0.55
3	0.85	0.05	0.1	16	0.2	0.22	0.58
4	0.8	0.07	0.13	17	0.15	0.23	0.62
5	0.75	0.08	0.17	18	0.1	0.24	0.66
6	0.7	0.09	0.21	19	0.05	0.25	0.7
7	0.65	0.1	0.25	20	0.03	0.27	0.7
8	0.6	0.12	0.28	21	0	0.28	0.72
9	0.55	0.13	0.32	22	0.5	0.14	0.36
10	0.5	0.14	0.36	23	1	0	0
11	0.45	0.15	0.4	24	0	1	0
12	0.4	0.17	0.43	25	0	0	1
13	0.35	0.18	0.47				

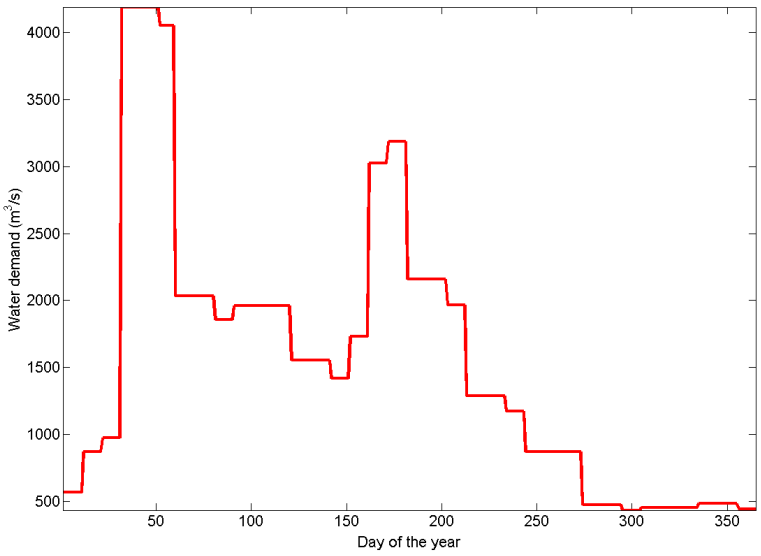


Figure 5.3: Increased water demand.

5.1.2). The sample dataset χ_t comprises the variables is listed in table 5.4, and the detail is reported in the appendix (see table A.1).

Table 5.4: Input variables for the information selection step.

Code	Variable	Code	Variable
1	t	14	A_Niño3.4_lag1M
2	a_{t+1}	15	A_Niño3.4_lagM2
3	q_t^{Tabu}	16	A_Niño3.4_lag3M
4	$q_t^{Laichau}$	17	A_Niño3.4_lag4M
5	$q_t^{Namgiang}$	18	A_Niño3.4_lag5M
6	q_t^{Yenbai}	19	A_Niño3.4_lag6M
7	$q_t^{Vuquang}$	20	A_Niño3.4_lag7M
8	$P_t^{Muongte}$	21	A_Niño3.4_lag8M
9	$P_t^{Tamduong}$	22	A_Niño3.4_lag9M
10	$P_t^{DaSpatialAverage}$	23	A_Niño3.4_lag10M
11	P_t^{Baolac}	24	A_Niño3.4_lag11M
12	$P_t^{Bachme/Bacme}$	25	A_Niño3.4_lag12M
13	$P_t^{Hagiang}$	26	s_t

PARAMETER n_{min}: The parameter n_{min} is the minimum number of observations needed to partition a node in the construction of the extra trees used by the IIS algorithm. This parameter adjusts the degree of noise attenuation of the output variable, so often called parameter smoothing. Theoretically, n_{min} can take values between 1 and N, where N is the number total data of training dataset. Large values of n_{min} produce little branched trees, with the effect of increasing the bias and reducing variance. On the contrary, small values of n_{min} produce very detailed information which well explains the calibration data but often generates overfitting over the calibration dataset and poor performance in validation. In this study, the results of preliminary analysis suggested to use n_{min} equal 15.

PARAMETER K: The parameter K, known as alternative inputs or cut-directions, controls the level of randomness in the tree building process. K can be chosen in the interval $[1, \ldots, n]$, with n being the number of input variables. The smaller K is, the stronger the randomization of the trees is and the weaker the dependence of their structure on the values of the output variable in the training dataset is [*Galelli and Castelletti*, 2013]. According to *Geurts, Ernst, and Wehenkel*, 2006, the optimal value of K for regression

problems is equal to the number n of inputs. Therefore, K was set equal to 26 in our experiments.

PARAMETER M: The parameter M, known as number of trees in the ensemble, controls the reduction of variance in the extra trees forest. There is no default value, the bigger the value of M is, the better accuracy of estimation is. It is worth noting that increasing value of M will lead to increased computational time. Practically, different values of M were tested and the value of 500 was chosen for the final experiments.

PARAMETER VLAG: The parameter $Vlag$ defines the method of validation to be used ,e.g., the classical K-fold cross-validation and the repeated random subsampling one. Results of preliminary analysis suggested to use the repeated random subsampling technique as it produced higher model performance (e.g., higher values of R^2).

5.1.5 *Improved Operating Policy design*

The set of IOPs was designed via Evolutionary Multi-Objective Direct Policy Search [*Giuliani et al.*, 2015a], with the operating policies parameterized as gaussian Radial Basis Functions as they are capable of representing functions for a large class and have been demonstrated to be more effective than other universal approximators in solving water resources operations problems (e.g., *Giuliani et al.*, 2014a; *Giuliani et al.*, 2015a). The number of basis of the Radial Basis Functions is set equal to $U = V + 1$, where V is the arity of the policy input vector $\mathcal{I}_t = (t, \mathbf{x}_t, \mathbf{I}_t)$. The corresponding number of policy parameters to be optimized is $n_\theta = U(2V + 1)$.

To perform the optimization, we use the self-adaptive Borg MOEA [*Hadka and Reed*, 2013], which has been shown to be highly robust across a diverse suite of challenging multi-objective problems, where it met or exceeded the performance of other state-of-the-art MOEAs [*Hadka and Reed*, 2012]. Since the Borg MOEA has been demonstrated to be relatively insensitive to the tuning parameters, we use the default algorithm parameterization suggested by *Hadka and Reed*, 2013, overcoming the limitations of tuning the algorithm parameters to the specific problem. Each optimization was run for 250,000 function evaluations, with the simulation of the system performed over the same horizon of DDP (i.e., 1995-2006). To improve solution diversity and avoid dependence on randomness, the solution set from each formulation is the result of 20 random optimization trials. The final set of Pareto optimal policies for each experiment is defined as the set of non-dominated solutions from the results of all the optimization trials.

5.2 RESULTS

5.2.1 *Quantifying Expected Value of Perfect Information*

The result of DDP experiments are shown in figure 5.4. The top panel illustrates a 3D visualization of the policies performance, where J^{irr} is represented on the *x-axis* (to be minimized), J^{flo} on the *y-axis* (to be minimized), and J^{hyd} on the *z-axis* (to be maximized). The red arrows show the direction of increasing preference, with the best solution that is the one closest to origin and having the height as big as possible. For instance, solution *A* is the closest point to the origin, which is the best for water supply and flood prevention, but has lowest values of hydropower production. On the contrary, solution *B* is the best for hydropower production, but it is the worst for both water supply and flood prevention. Therefore, the compromise solutions are the ones balancing the trade-off between the three conflicting objectives (e.g., the points inside the red rectangle in figure 5.4).

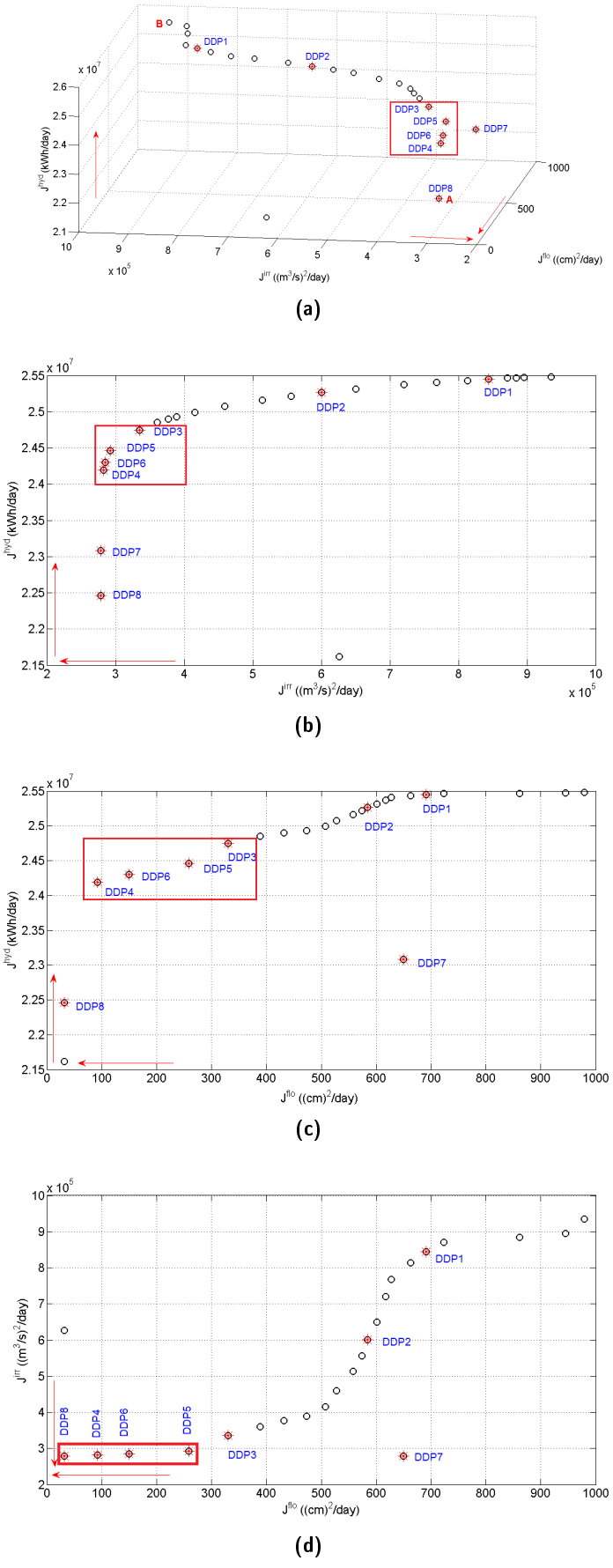


Figure 5.4: Performances of Perfect Operating Policies. The red crosses are the selected POPs for the IVS experiments

In order to explore the conflicts among the three objectives, the DDP solutions are projected into 2D visualizations (see last three panels in figure 5.4). It can be seen clearly from panel *(b)* and panel *(c)* that there is a big conflict between hydropower production and water supply sector or flood prevention. Specifically, we want to maximize J^{hyd} on the *y-axis* and to minimize the J^{irr} and J^{flo} on the *x-axis*. Therefore, the solutions on the top-left corner are the best solutions.

Finally, figure 5.4*d* illustrates that there is no conflict between water supply and flood prevention because minimizing J^{irr} leads to minimizing J^{flo}. Particularly, J^{irr} requires to release water from February to March, thus reducing the levels in the reservoirs and, hence, making space for buffering floods. Interestingly, in the bottom-left part of the panel (the red rectangle), J^{flo} is more sensitive than to J^{irr} (i.e., small change in J^{irr} leads to big change in J^{flo}).

5.2.2 *Information selection*

The IVS experiments were carried out in two steps. In the first step, preliminary tests were performed in order to identify the optimal setting for running the IIS algorithm. Specifically, we considered 8 different POPs as reference points, namely *DDP1* to *DDP8* (the red crosses in figure 5.4). For the preliminary tests, the number of repetitions required to filter the randomness of the extra trees used by the IIS algorithm is equal to 5. In the second step, the final experiments were carried out with the optimal setting value determined in the previous step. The number of repetitions is increased to 50 to get more robust and significant results.

The results of preliminary tests are reported in appendix (see table A.2 and table A.3). It is interesting to observe that the selected variables change depending on the solution considered. For instance, in the case of *DDP1*, which is the best solution regarding the hydropower production, the IIS algorithm selected the *time (t)* and the *reservoir inflow* (a_{t+1}). This information indeed allows turbining as much as possible without varying the level in order to maximize hydropower production. On the contrary, in case of *DDP8*, which is the best solution for water supply objective, the *time (t)*, the Hoa Binh *storage* (s_t) and the *precipitation* in Muong Te station ($P_t^{Muongte}$) were selected in order to balance the water released from Hoa Binh with the lateral contributions of Thao and Lo rivers.

Interestingly, some solutions in the region of the maximum curvature (e.g., *DDP4* and *DDP6*) selected ENSO information for balancing high J^{hyd} and low J^{irr} due to the extended foresight provided by the ENSO indices. Specifically, the ENSO information, namely A_Niño3.4_lag1M, was mostly selected in *DDP6*, which was considered as reference solution for the next experiments.

Table 5.5: Performance of IVS: Selected times of variables

Code	Variable	Selected time	Code	Variable	Selected time
1	t	46	14	A_Niño3.4_lag1M	35
2	a_{t+1}	12	15	A_Niño3.4_lag2M	0
3	q_t^{Tabu}	35	16	A_Niño3.4_lag3M	0
4	$q_t^{Laichau}$	2	17	A_Niño3.4_lag4M	0
5	$q_t^{Namgiang}$	0	18	A_Niño3.4_lag5M	0
6	q_t^{Yenbai}	5	19	A_Niño3.4_lag6M	0
7	$q_t^{Vuquang}$	1	20	A_Niño3.4_lag7M	0
8	$P_t^{Muongte}$	0	21	A_Niño3.4_lag8M	0
9	$P_t^{Tamduong}$	0	22	A_Niño3.4_lag9M	0
10	$P_t^{DaSpatialAverage}$	6	23	A_Niño3.4_lag10M	0
11	P_t^{Baolac}	20	24	A_Niño3.4_lag11M	0
12	$P_t^{Bachme/Bacme}$	10	25	A_Niño3.4_lag12M	0
13	$P_t^{Hagiang}$	0	26	s_t	50

From the results of the preliminary tests, we decided to further analyze solution *DDP6* by running the IIS algorithm 50 times. The results are summarized in the table 5.5 and the full results are reported in appendix (see table A.4 and table A.5). It can be seen that the most important variables are the *time* (t), the Hoa Binh *storage* (s_t), the *flow* in Ta Bu station (q_t^{Tabu}) and the *anomaly* of ElNiño3.4 with lag time of one. Figure 5.5 and table 5.6 show the accumulated performance R^2 of the underlying model, which obtain $R^2 = 71.05\%$, as well as the contribution of each selected variable ΔR^2, evaluated as the variation of the coefficient of determination R^2 at each iteration. The results show that

Table 5.6: Results of the IIS algorithm.

Variable	Frequency	ΔR^2	R^2
t	46	36.55	36.55
s_t	50	31.29	67.85
q_t^{Tabu}	35	0.28	68.13
A_Niño3.4_lag1M	35	2.92	71.05

the *time (t)* and the Hoa Binh *storage* (s_t) have the largest contribution, i.e., 36.55% and 31.29%, respectively. Interestingly, ENSO information, i.e., A_Niño3.4_lag1M, contributed 2.92% to the performance of the model. It means that the anomaly of Niño3.4 information is expected to improve the operations of the Hoa Binh reservoir.

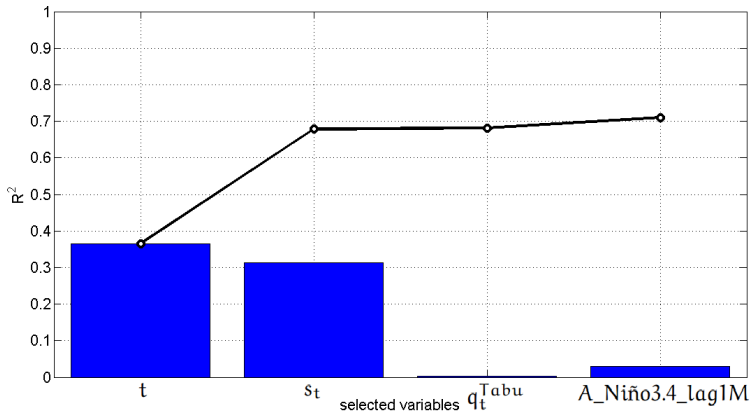


Figure 5.5: Performance improvement by each selected variable. The continuous line is the accumulated performance of the underlying model.

5.2.3 *Improved Operating Policy and assessment of the Expected Value of Sample Information*

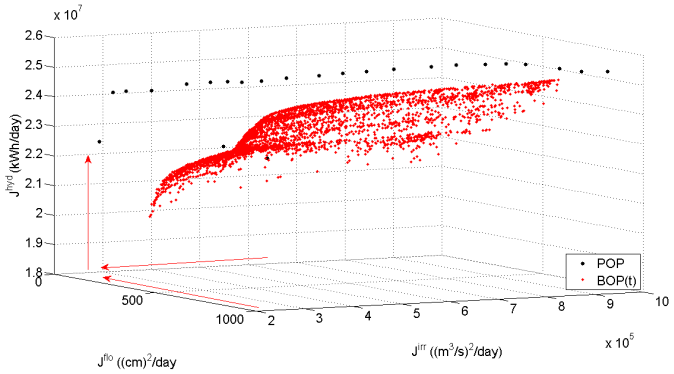
Figure 5.6 and table 5.7 report the results of the experiments when we incrementally add the variables selected by the IIS algorithm to the set of policy input and we iteratively design and analyze the resulting IOP both in terms of their performances with respect to the three operating objectives (i.e., J^{hyd}, J^{flo} and J^{irr}) and to the associated EVSI, measured in terms of hypervolume indicator.

Table 5.7: Hypervolume indicator.

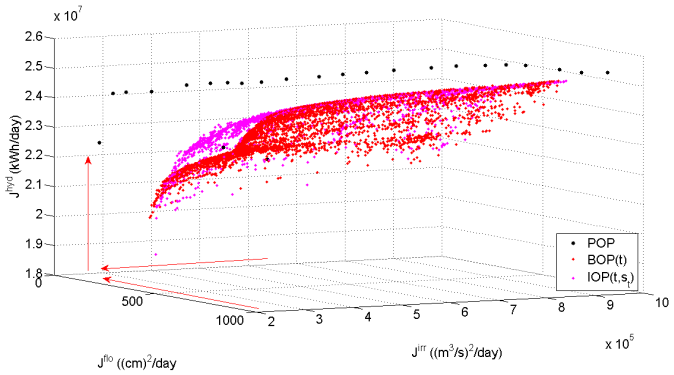
Policy	Hypervolume
$BOP(t)$	0.3034
$IOP(t, s_t)$	0.3903
$IOP(t, s_t, q_t^{Tabu})$	0.3830
$IOP(t, s_t, q_t^{Tabu}, A\_Niño3.4\_lag1M)$	0.4106

It is not surprising that the day of the year (t) was selected as the first variable due to the strong influence of the seasonality and of the monsoon period on the system dynamics. The set of designed policies based on day of the year is called Basic Operating Policy, represented by the red points in figure 5.6*a*. The Hypervolume (HV), a quantitative indicator of the value of this information, equals 30.34% confirming that there is a big gap between BOP and POP.

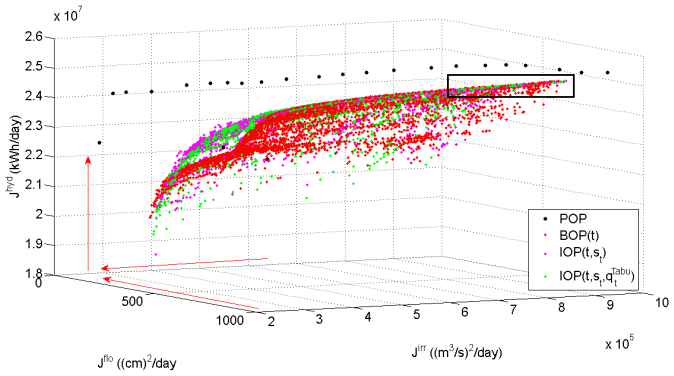
Then, the second selected variable is the Hoa Binh *storage* (s_t). The associated policies $IOP(t, s_t)$ are represented by magenta points in figure 5.6*b*. The comparison of the performance of BOP(t) and $IOP(t, s_t)$ shows a large contribution associated with the Hoa Binh



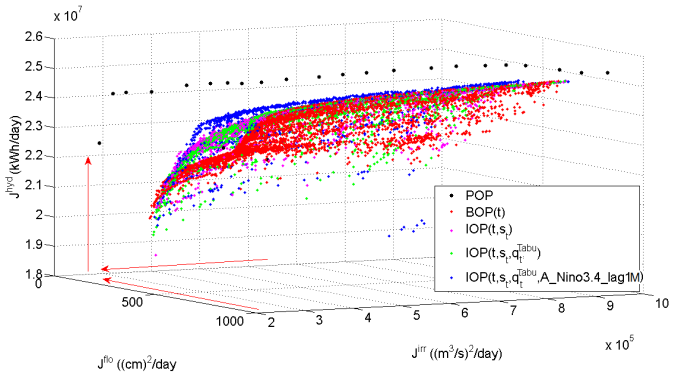
(a) POP, BOP(t)



(b) POP, BOP(t), IOP(t, s_t)



(c) POP, BOP(t), IOP(t, s_t), IOP(t, s_t, q_t^{Tabu})



(d) POP, BOP(t), IOP(t, s_t), IOP(t, s_t, q_t^{Tabu}), IOP(t, s_t, q_t^{Tabu}, A_Niño3.4_lag1M)

Figure 5.6: Comparison of Perfect, Basic, and Improved Operating Policies performance in the objective space.

storage. Specifically, the HV increases from 30.34% to 39.03% (i.e., +8.69%) when moving from BOP(t) to IOP(t, s_t). It means that conditioning the Hoa Binh policy on the reservoir storage allows improving the three objectives. However, there is still space for further improvements if we compare this set of solutions with the performance of the POP. This reason motivates further improvement of the operating policy by introducing additional exogenous information.

The third variable selected is the *flow* at Ta Bu station (q_t^{Tabu}). The associated IOP(t, s_t, q_t^{Tabu}) are represented by green points in figure 5.6*c*. It can be seen that there are some improvements on the right corner of the figure (see the black rectangle), whereas there are some declines on the left corner with respect to IOP(t, s_t). This means that the flow at Ta Bu station will be useful only for some particular objectives. This is confirmed by the values of HV which indeed slightly degrade when moving from IOP(t, s_t) to IOP(t, s_t, q_t^{Tabu}).

Finally, the last selected variable is the ENSO information (i.e., the anomalies of El Niño with lag time of one month). The performance of IOP(t, s_t, q_t^{Tabu}, A_Niño3.4_lag1M) is shown in figure 5.6*d* by the blue points. This set of IOPs appears to be the closest to POPs. The HV indicator confirms again this conclusion since it increases from 38.03% to 41.60%, corresponding to a 3.3% of improvement.

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

This thesis provides a novel procedure for identifying the impact of El Niño Southern Oscillation (ENSO) on the water system at the basin scale and using this information for improving the management of the system. The Information Selection and Assessment (ISA) framework is used for selecting the most valuable information for improving water systems operations and quantifying the value of the selected information. The proposed procedure is applied to the operations of the Hoa Binh reservoir, a multi-objective reservoir, in Vietnam.

The results show that ISA framework is capable of selecting the most important variables which allow to better understand the hydrometeorological processes, and to improve the operating policies of Hoa Binh reservoir. Besides traditional variables such as the day of the year and the reservoir storage, the use of ENSO information (i.e., the anomaly of SST3.4) is demonstrated to contribute in informing the design of compromise operating policies that balance the three conflicting objectives associated to the operations of the Hoa Binh reservoir.

Although there is a large space between the Perfect Operating Policies, designed under the assumption of perfect foresight on the future, and the Basic Operating Policies relying only on the time information, the use of additional information allows improving the system performance. Indeed, by incrementally adding the variable selected by the Iterative Input variable Selection algorithm to the reservoir operating policies, the corresponding Improved Operating Policy (IOP) moves towards Perfect Operating Policy (POP). Interestingly, ENSO information, namely *A_Niño3.4_lag1M*, is demonstrated to improve the operations of the Hoa Binh reservoir , with the corresponding Pareto front that attains a +3.3% in terms of hypervolume indicator with respect to the set of solutions not conditioned on the ENSO index.

6.2 FUTURE RESEARCH DIRECTIONS

The results presented in this thesis rely on several simplifications and assumptions that would lead to the further discussions and improvements. First, the analysis focuses on three main objectives, namely hydropower production, water supply and flood control, while neglecting other issues such as river bed erosion, navigation and ecosys-

tem conservation. Taking into account these objectives, the model becomes more complex with the involvement of other stakeholders.

Second, some indicators are constructed based on the current context, which may not be suitable for the future. For instance, instead of maximizing hydropower production, we can import electricity from China due to the fact that a lot of dams have been built in China, leading to the lower cost of electricity. Therefore, for example, new scenarios can be built focusing on other stakeholders such as water supply and flood control.

Finally, this thesis focuses only on the operation of Hoa Binh reservoir in Da River basin. In fact, several reservoirs have been constructed and operated in Da River basin as well as in Red River basin. Therefore, further researches should take into account the coordinate management of this multi-reservoir network.

BIBLIOGRAPHY

[1] R. Allan, J. Lindesay, D. Parker, and D. Parker. "El Nino: Southern oscillation and climatic variability." In: (1996) (cit. on pp. 5, 6).

[2] R. Bellman. "Dynamic programming and Lagrange multipliers." In: *Proceedings of the National Academy of Sciences of the United States of America* 42.10 (1956), p. 767 (cit. on p. 17).

[3] L. Beltrame and D. Carbonin. "ENSO teleconnection patterns on large scale water resources systems." MA thesis. Politecnico di Milano, Italy, 2013 (cit. on pp. 2, 6, 37).

[4] A. Castelletti, F. Pianosi, and R. Soncini-Sessa. "Water reservoir control under economic, social and environmental constraints." In: *Automatica* 44.6 (2008), pp. 1595–1607 (cit. on p. 12).

[5] A. Castelletti, H. Yajima, M. Giuliani, R. Soncini-Sessa, and E. Weber. "Planning the optimal operation of a multioutlet water reservoir with water quality and quantity targets." In: *Journal of Water Resources Planning and Management* 140.4 (2013), pp. 496–510 (cit. on p. 1).

[6] A. Cavalli. "Direct use of hydroclimatic information in water reservoir operation." MA thesis. Politecnico di Milano, Italy, 2012 (cit. on p. 2).

[7] J. Chandimala and L. Zubair. "Predictability of stream flow and rainfall based on ENSO for water resources management in Sri Lanka." In: *Journal of Hydrology* 335.3 (2007), pp. 303–312 (cit. on p. 9).

[8] F. H. S. Chiew and T. A. McMahon. "Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability." In: *Hydrological Sciences Journal* 47.3 (2002), pp. 505–522 (cit. on p. 4).

[9] F.H.S. Chiew and T.A. McMahon. "Global ENSO- streamflow teleconnection, streamflow forecasting and interannual variability." In: *Hydrological Sciences Journal* 47:3.18 (2002), pp. 505–522 (cit. on pp. 6, 7, 11).

[10] K.M. Cobb, C.D. Charles, H. Cheng, and R.L. Edwards. "El Nino/Southern Oscillation and tropical Pacific climate during the last millennium." In: *Nature* 424.6946 (2003), pp. 271–276 (cit. on p. 6).

[11] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A.E. Rizzoli. "Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review." In: *Environmental Modelling & Software* 72 (2015), pp. 198–214 (cit. on p. 1).

[12] S. Galelli and A. Castelletti. "Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling." In: *Hydrology and Earth System Sciences* 17.7 (2013), pp. 2669–2684 (cit. on pp. 15, 16, 42).

[13] S. Galelli and A. Castelletti. "Tree-based iterative input variable selection for hydrological modeling." In: *Water Resources Research* 49.7 (2013), pp. 4295–4310 (cit. on pp. 15–17).

[14] S. Galelli, Greer B. Humphrey, Holger R. Maier, A. Castelletti, Graeme C. Dandy, and Matthew S. Gibbs. "An evaluation framework for input variable selection algorithms for environmental data-driven models." In: *Environmental Modelling & Software* 62 (2014), pp. 33–51 (cit. on p. 15).

[15] P. Gerland, Adrian E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, Bailey K. Fosdick, J. Chunn, N. Lalic, et al. "World population stabilization unlikely this century." In: *Science* 346.6206 (2014), pp. 234–237 (cit. on p. 1).

[16] P. Geurts, D. Ernst, and L. Wehenkel. "Extremely randomized trees." In: *Machine learning* 63.1 (2006), pp. 3–42 (cit. on p. 42).

[17] M. Giuliani and A. Castelletti. "Assessing the value of cooperation and information exchange in large water resources systems by agent-based optimization." In: *Water Resources Research* 49.7 (2013), pp. 3912–3926 (cit. on p. 1).

[18] M. Giuliani, S. Galelli, and R. Soncini-Sessa. "A dimensionality reduction approach for many-objective Markov Decision Processes: Application to a water reservoir operation problem." In: *Environmental Modelling & Software* 57 (2014b), pp. 101–114 (cit. on p. 18).

[19] M. Giuliani, F. Pianosi, and A. Castelletti. "Making the most of data: an information selection and assessment framework to improve water systems operation." In: *Water Resources Research* 17.12 (2015b). (under review), pp. 667–673 (cit. on pp. 2, 12, 13, 15).

[20] M. Giuliani, A. Castelletti, F. Pianosi, E. Mason, and Patrick M. Reed. "Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations." In: *Journal of Water Resources Planning and Management* (2015a), p. 04015050 (cit. on pp. 17, 43).

[21] M. Giuliani, J.D. Herman, A. Castelletti, and P. Reed. "Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management." In: *Water Resources Research* 50.4 (2014a), pp. 3355–3377 (cit. on pp. 1, 43).

[22] D. Hadka and P. Reed. "Borg: An auto-adaptive many-objective evolutionary computing framework." In: *Evolutionary Computation* 21.2 (2013), pp. 231–259 (cit. on p. 43).

[23] D. Hadka and P. Reed. "Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization." In: *Evolutionary Computation* 20.3 (2012), pp. 423–452 (cit. on p. 43).

[24] K. Hansson and L. Ekenberg. "Flood mitigation strategies for the Red River Delta." In: *Proceeding of the 2002 Joint CSCE/EWRI of ASCE International Conference on Environmental Engineering, An International Perspective on Environmental Engineering, Niagara Falls, Ont., Canada, July*. 2002, pp. 21–24 (cit. on p. 34).

[25] H.P. Hapuarachchi. "Relationship between ENSO and snow covered area in the Mekong and Yellow River basins." In: *Regional Hydrological Impacts of Climatic Change: Hydroclimatic variability* 2 (2005), p. 255 (cit. on p. 21).

[26] T. Hashimoto, Jery R. Stedinger, and Daniel P. Loucks. "Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation." In: *Water resources research* 18.1 (1982), pp. 14–20 (cit. on p. 35).

[27] The international journal on Hydropower & Dam. *World Atlas Industry Guide, 2014*. Buying online at `http://www.hydropower-dams.com/world-atlas-industry-guide.php?c_id=159`. 2014 (cit. on p. 24).

[28] IPCC. "The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change." In: *Cambridge, United Kingdom and New York, USA* (2013) (cit. on p. 1).

[29] E. Kahya and J.A. Dracup. "US Strearnflow Patterns in Relation to the El Nifio/Southern Oscillation." In: *Water Resources Research* 29.8 (1993), 2491–Z503 (cit. on pp. 1, 4, 5).

[30] J.R. Kasprzyk, P.M. Reed, B.R. Kirsch, and G.W. Characklis. "Managing population and drought risks using many-objective water portfolio planning under uncertainty." In: *Water Resources Research* 45.12 (2009) (cit. on p. 18).

[31] A.S. Kiem and S.W. Franks. "On the identification of ENSO-induced rainfall and runoff variability: a comparison of methods and indices." In: *Hydrological sciences journal* 46.5 (2001), pp. 715–727 (cit. on pp. 7, 11).

[32] L. Le Ngo, H. Madsen, D. Rosbjerg, and C.B. Pedersen. "Implementation and comparison of reservoir operation strategies for the Hoa Binh reservoir, Vietnam using the MIKE 11 model." In: *Water Resources Management* 22.4 (2008), pp. 457–472 (cit. on p. 2).

[33] D.P. Loucks, E. Van Beek, J.R. Stedinger, J.P.M. Dijkman, and M.T. Villars. *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: UNESCO, 2005 (cit. on p. 1).

[34] H.R. Maier, Z. Kapelan, J. Kasprzyk, J. Kollat, L.S. Matott, M.C. Cunha, G.C. Dandy, M.S. Gibbs, E. Keedwell, A. Marchi, et al. "Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions." In: *Environmental Modelling & Software* 62 (2014), pp. 271–299 (cit. on p. 18).

[35] K.T. Pham, M.B. Nguyen, and H.D. Nguyen. "Energy supply, demand, and policy in Vietnam, with future projections." In: *Energy Policy* 39.11 (2011), pp. 6814–6826 (cit. on p. 25).

[36] X. Quach. "Assessing and optimizing the operation of the Hoa Binh reservoir, Vietnam, by multi-objective optimal control techniques." PhD thesis. Politecnico di Milano, 2011 (cit. on pp. 2, 20, 23, 24, 33–35).

[37] T.A. Räsänen and M. Kummu. "Spatiotemporal influences of ENSO on precipitation and flood pulse in the Mekong River Basin." In: *Journal of Hydrology* 476 (2013), pp. 154–168 (cit. on pp. 3, 6, 21, 39).

[38] H. Ronghui, C. Wen, Y. Bangliang, and Z. Renhe. "Recent advances in studies of the interaction between the East Asian winter and summer monsoons and ENSO cycle." In: *Advances in Atmospheric Sciences* 21.3 (2004), pp. 407–424 (cit. on pp. 6, 21, 39).

[39] A. Sharma. "Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification." In: *Journal of Hydrology* 239.1 (2000), pp. 232–239 (cit. on p. 1).

[40] A. Shrestha and R. Kostaschuk. "El Nino/Southern Oscillation (ENSO)-related variablity in mean-monthly streamflow in Nepal." In: *Journal of Hydrology* 308.1 (2005), pp. 33–49 (cit. on pp. 4, 6).

[41] H.J. Simpson, M.A. Cane, A.L. Herczeg, S.E. Zebiak, and J.H. Simpson. "Annual River Discharge in Southeastern Australia Related to." In: *Water Resources Research* 29.11 (1993), pp. 3671–3680 (cit. on p. 6).

[42] T.M. Smith and R.W. Reynolds. "Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997)." In: *Journal of Climate* 16.10 (2003), pp. 1495–1510 (cit. on p. 9).

[43] R. Soncini-Sessa, E. Weber, and A. Castelletti. *Integrated and participatory water resources management-theory*. Vol. 1. Elsevier, 2007 (cit. on pp. 35, 40).

[44] J.N. Tsitsiklis and B. Van Roy. "Feature-based methods for large scale dynamic programming." In: *Machine Learning* 22.1-3 (1996), pp. 59–94 (cit. on p. 17).

[45] P.N. Vu and V.A. Truong. "Assessing the impact of climate change on the inflow to Hoa Binh reservoir, Vietnam." MA thesis. Politecnico di Milano, Italy, 2010 (cit. on p. 2).

[46] B. Wang, R. Wu, and X. Fu. "Pacific-East Asian teleconnection: how does ENSO affect East Asian climate?" In: *Journal of Climate* 13.9 (2000), pp. 1517–1536 (cit. on p. 6).

APPENDIX

Table A.1: List of input variables of IVS experiments.

Code	Notation	Variable
1	t	Time
2	a_{t+1}	Flow of Hoa Binh reservoir
3	q_t^{Tabu}	Flow at Ta Bu station
4	$q_t^{Laichau}$	Flow at Lai Chau station
5	$q_t^{Namgiang}$	Flow at Nam Giang station
6	q_t^{Yenbai}	Flow at Yen Bai station
7	$q_t^{Vuquang}$	Flow at Vu Quang station
8	$P_t^{Muongte}$	Precipitation at Muong Te station
9	$P_t^{Tamduong}$	Precipitation at Tam Duong station
10	$P_t^{DaSpatialAverage}$	Average precipitation in Da Basin
11	P_t^{Baolac}	Precipiation at Bao Lac station
12	$P_t^{Bachme/Bacme}$	Precipitation at Bach Me station
13	$P_t^{Hagiang}$	Precipitation at Ha Giang staion
14	A_Niño3.4_lag1M	Anomaly of Niño3.4 with time lag of 1 month
15	A_Niño3.4_lag2M	Anomaly of Niño3.4 with time lag of 2 months
16	A_Niño3.4_lag3M	Anomaly of Niño3.4 with time lag of 3 months
17	A_Niño3.4_lag4M	Anomaly of Niño3.4 with time lag of 4 months
18	A_Niño3.4_lag5M	Anomaly of Niño3.4 with time lag of 5 months
19	A_Niño3.4_lag6M	Anomaly of Niño3.4 with time lag of 6 months
20	A_Niño3.4_lag7M	Anomaly of Niño3.4 with time lag of 7 months
21	A_Niño3.4_lag8M	Anomaly of Niño3.4 with time lag of 8 months
22	A_Niño3.4_lag9M	Anomaly of Niño3.4 with time lag of 9 months
23	A_Niño3.4_lag10M	Anomaly of Niño3.4 with time lag of 10 months
24	A_Niño3.4_lag11M	Anomaly of Niño3.4 with time lag of 11 months
25	A_Niño3.4_lag12M	Anomaly of Niño3.4 with time lag of 12 months
26	s_t	Storage of Hoa Binh reservoir

Table A.2: Preliminary performances of IVS (Vlag 1)

	Selected variable (Vlag = 1)						R2 (Vlag = 1)					
DDP1	1	2	4	7	-	-	0.3568	0.5207	0.5232	0.5697	-	-
	1	2	26	-	-	-	0.3648	0.5230	0.6120	-	-	-
	1	2	4	26	7	-	0.3806	0.5154	0.5197	0.5994	0.6576	-
	1	2	7	6	-	-	0.3792	0.5349	0.5760	0.5714	-	-
	1	2	4	26	6	10	0.3787	0.5071	0.5281	0.6045	0.6403	0.6387
DDP2	1	26	2	10	-	-	0.3831	0.5715	0.5988	0.6035	-	-
	1	26	10	11	-	-	0.3919	0.5597	0.5815	0.5762	-	-
	1	26	10	2	-	-	0.3610	0.5702	0.5809	0.6147	-	-
	1	26	2	6	4	-	0.4029	0.5815	0.6149	0.6389	0.6307	-
	1	26	12	11	-	-	0.3937	0.5773	0.5820	0.5773	-	-
DDP3	4	26	11	1	10	-	0.3471	0.5047	0.5426	0.6230	0.6163	-
	3	26	10	11	1	2	0.3446	0.5389	0.5587	0.5619	0.6248	0.6111
	1	26	12	-	-	-	0.3402	0.6191	0.6053	-	-	-
	1	26	12	-	-	-	0.3451	0.6268	0.6095	-	-	-
	3	26	10	11	1	6	0.3464	0.5323	0.5534	0.5629	0.6252	0.6267
DDP4	1	26	3	6	4	-	0.3807	0.6732	0.6961	0.7049	0.6967	-
	1	26	3	5	-	-	0.3878	0.6734	0.6987	0.6909	-	-
	1	26	3	7	5	-	0.3776	0.6610	0.6939	0.7098	0.7096	-
	1	26	3	-	-	-	0.3640	0.6671	0.6990	-	-	-
	1	26	5	3	7	**14**	0.3782	0.6721	0.6854	0.6860	0.7085	**0.7137**
DDP5	1	26	12	-	-	-	0.3640	0.6591	0.6396	-	-	-
	1	26	12	-	-	-	0.3582	0.6643	0.6408	-	-	-
	1	26	12	-	-	-	0.3657	0.6596	0.6432	-	-	-
	1	26	11	-	-	-	0.3553	0.6590	0.6419	-	-	-
	1	26	12	-	-	-	0.3658	0.6706	0.6500	-	-	-
DDP6	3	**14**	26	12	-	-	0.3356	**0.5163**	0.6071	0.5919	-	-
	1	26	3	11	-	-	0.3680	0.6634	0.6900	0.6785	-	-
	1	26	3	**14**	11	-	0.3579	0.6404	0.6807	**0.6915**	0.6841	-
	1	26	3	11	-	-	0.3604	0.6534	0.6765	0.6638	-	-
	1	26	**14**	6	-	-	0.3725	0.6547	**0.7053**	0.6977	-	-
DDP7	1	26	7	-	-	-	0.8196	0.8890	0.8881	-	-	-
	1	26	2	12	-	-	0.8187	0.8877	0.8988	0.8866	-	-
	1	26	2	9	-	-	0.8188	0.8897	0.8995	0.8872	-	-
	1	26	2	10	-	-	0.8143	0.8891	0.8980	0.8837	-	-
	1	26	6	-	-	-	0.8225	0.8899	0.8942	-	-	-
DDP8	1	22	8	-	-	-	0.4797	0.7910	0.6820	-	-	-
	1	23	10	-	-	-	0.4923	0.7968	0.6823	-	-	-
	1	26	7	8	-	-	0.5003	0.7010	0.7339	0.7186	-	-
	1	26	6	8	-	-	0.4793	0.7074	0.7421	0.7223	-	-
	1	26	7	8	-	-	0.5042	0.7051	0.7370	0.7224	-	-

Table A.3: Preliminary performances of IVS (Vlag 2)

	Selected variable (Vlag = 2)						R2 (Vlag = 2)					
DDP1	1	2	4	-	-	-	0.3948	0.4960	0.5381	-	-	-
	1	2	3	-	-	-	0.4148	0.5298	0.5235	-	-	-
	3	26	7	2	10	-	0.3309	0.5497	0.6380	0.7013	0.6308	-
	1	2	3	-	-	-	0.4124	0.5504	0.5409	-	-	-
	1	2	7	26	-	-	0.3703	0.5232	0.5790	0.6935	-	-
DDP2	1	26	10	4	-	-	0.4117	0.6002	0.6120	0.5783	-	-
	1	26	4	-	-	-	0.3918	0.5982	0.5755	-	-	-
	1	26	12	2	4	-	0.4154	0.5645	0.5919	0.6208	0.6101	-
	1	26	2	4	-	-	0.4058	0.5899	0.6043	0.5727	-	-
	1	26	4	-	-	-	0.4347	0.5654	0.5794	-	-	-
DDP3	1	26	2	10	-	-	0.3276	0.6305	0.6377	0.5977	-	-
	5	26	2	11	10	-	0.3158	0.4908	0.5633	0.5899	0.5697	-
	1	4	26	11	2	3	0.3421	0.5056	0.5845	0.5940	0.6214	0.6167
	2	26	11	10	-	-	0.3436	0.5394	0.5721	0.5366	-	-
	4	26	1	2	-	-	0.3212	0.5050	0.6386	0.6354	-	-
DDP4	1	26	7	3	5	***14***	0.3971	0.6496	0.6833	0.7153	0.7320	***0.7181***
	1	26	3	6	7	***14***	0.3735	0.6556	0.6780	0.7017	0.7069	***0.7145***
	1	26	3	7	-	-	0.3690	0.6597	0.7140	0.7045	-	-
	1	26	2	7	-	-	0.3951	0.6621	0.7280	0.6962	-	-
	1	26	3	7	-	-	0.3588	0.6672	0.6706	0.6892	-	-
DDP5	1	26	11	-	-	-	0.3709	0.6725	0.6357	-	-	-
	1	26	11	-	-	-	0.3725	0.6611	0.6529	-	-	-
	1	26	11	-	-	-	0.3647	0.6618	0.6401	-	-	-
	1	26	11	6	-	-	0.3607	0.6411	0.6513	0.6463	-	-
	1	26	11	-	-	-	0.3478	0.6652	0.6280	-	-	-
DDP6	1	26	3	***14***	12	-	0.3534	0.6591	0.6788	***0.7198***	0.6787	-
	1	26	3	***14***	11	-	0.3621	0.6450	0.6903	***0.7014***	0.6702	-
	1	26	***14***	12	-	-	0.3824	0.6691	***0.7072***	0.6698	-	-
	1	26	2	11	-	-	0.3235	0.6274	0.6973	0.6875	-	-
	1	26	2	***14***	11	-	0.3415	0.6642	0.6813	***0.6902***	0.6974	-
DDP7	1	26	2	12	-	-	0.8158	0.8725	0.8930	0.8822	-	-
	1	26	7	-	-	-	0.8101	0.8831	0.8958	-	-	-
	1	26	2	-	-	-	0.8325	0.8841	0.8985	-	-	-
	1	26	7	-	-	-	0.8051	0.8788	0.8891	-	-	-
	1	26	2	-	-	-	0.8271	0.8951	0.9010	-	-	-
DDP8	1	3	26	7	-	-	0.4800	0.6259	0.7461	0.7416	-	-
	1	26	8	-	-	-	0.5104	0.7018	0.6804	-	-	-
	1	22	8	-	-	-	0.4983	0.7908	0.6657	-	-	-
	1	3	26	7	8	-	0.4819	0.5951	0.7083	0.7414	0.7197	-
	1	2	26	8	7	12	0.4621	0.6095	0.7256	0.7424	0.7575	0.7410

Table A.4: Performance of Input Variable Selection.

Selected variable						R2					
1	26	3	14	-	-	0.3563	0.6683	0.6949	0.6820	-	-
1	26	4	14	3	-	0.3779	0.6522	0.6751	0.6765	0.6739	-
1	26	14	12	-	-	0.3870	0.6671	0.6807	0.6596	-	-
1	26	3	11	-	-	0.3654	0.6552	0.6943	0.6801	-	-
1	26	3	14	6	10	0.3508	0.6447	0.6675	0.6993	0.7335	0.6726
1	26	3	14	6	11	0.3578	0.6705	0.6731	0.7090	0.7128	0.6860
1	26	2	-	-	-	0.3619	0.6936	0.6930	-	-	-
3	14	26	10	-	-	0.3223	0.5061	0.5830	0.5815	-	-
1	26	2	-	-	-	0.3783	0.6663	0.6622	-	-	-
1	26	3	14	11	-	0.3749	0.6676	0.6962	0.7085	0.6815	-
1	26	2	-	-	-	0.3544	0.6677	0.6624	-	-	-
1	26	3	14	12	-	0.3359	0.6479	0.6588	0.7019	0.6849	-
1	26	2	14	3	-	0.3639	0.6506	0.6821	0.7044	0.6980	-
1	26	3	14	-	-	0.3339	0.6364	0.6748	0.6687	-	-
1	26	3	14	11	-	0.3655	0.6785	0.6813	0.7105	0.6861	-
2	14	26	-	-	-	0.3114	0.4978	0.6107	-	-	-
1	26	3	11	14	10	0.3824	0.6610	0.6748	0.6830	0.7103	0.6713
1	26	3	11	-	-	0.3669	0.6659	0.6806	0.6778	-	-
1	26	3	14	12	-	0.3566	0.6659	0.6772	0.7017	0.6750	-
1	26	3	-	-	-	0.3702	0.6736	0.6453	-	-	-
1	26	4	14	12	3	0.3765	0.6363	0.6715	0.6985	0.7122	0.6435
1	26	3	14	12	10	0.3542	0.6516	0.6778	0.6896	0.6920	0.6819
1	26	14	11	-	-	0.3580	0.6573	0.7001	0.6628	-	-
3	14	26	12	-	-	0.3016	0.5039	0.6088	0.5689	-	-
1	26	3	11	-	-	0.3940	0.6619	0.6745	0.6522	-	-

Table A.5: Performances of Input Variable Selection (continuous).

Selected variable						R2					
1	26	3	14	6	-	0.3475	0.6312	0.6913	0.6965	0.6916	-
1	26	3	11	-	-	0.3472	0.6528	0.6791	0.6727	-	-
1	26	3	14	11	10	0.3928	0.6642	0.6890	0.7077	0.7100	0.6873
1	26	2	14	12	-	0.3771	0.6653	0.6859	0.7174	0.7028	-
1	26	3	-	-	-	0.3785	0.6594	0.6588	-	-	-
1	26	2	14	11	-	0.3615	0.6526	0.6741	0.7068	0.7002	-
1	26	3	14	6	10	0.3648	0.6492	0.6729	0.6811	0.6840	0.6881
1	26	3	14	-	-	0.3798	0.6630	0.6925	0.6896	-	-
1	26	14	11	-	-	0.3682	0.6810	0.7041	0.6517	-	-
3	14	26	12	-	-	0.3516	0.5251	0.6006	0.5900	-	-
1	26	3	11	-	-	0.3516	0.6390	0.6747	0.6666	-	-
1	26	2	14	11	-	0.3534	0.6371	0.6885	0.6966	0.6874	-
1	26	2	14	-	-	0.3697	0.6413	0.6957	0.6945	-	-
1	26	14	11	-	-	0.3568	0.6588	0.6948	0.6630	-	-
1	26	3	14	11	-	0.3747	0.6408	0.6787	0.6925	0.6892	-
1	26	2	14	11	6	0.3596	0.6544	0.6825	0.6923	0.6998	0.6773
1	26	3	-	-	-	0.3688	0.7010	0.6734	-	-	-
1	26	3	14	12	-	0.3446	0.6404	0.6870	0.7077	0.6812	-
1	26	3	11	-	-	0.3892	0.6467	0.6779	0.6745	-	-
1	26	3	7	-	-	0.3429	0.6303	0.6994	0.6809	-	-
1	26	2	14	11	-	0.3716	0.6615	0.6837	0.7005	0.6809	-
26	1	2	14	12	-	0.3110	0.6516	0.6772	0.7025	0.6863	-
1	26	3	14	-	-	0.3500	0.6575	0.7048	0.7018	-	-
1	26	3	-	-	-	0.3625	0.6727	0.6704	-	-	-
1	26	3	-	-	-	0.3753	0.6623	0.6554	-	-	-