A dimensionality reduction approach for Many-Objective Markov Decision Processes: application to a water reservoir operation problem

M. Giuliani[a,*], S. Galelli[b], R. Soncini-Sessa[a]

[a]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza L. da Vinci, 32, I-20133 Milano, Italy*

[b]*Pillar of Engineering Systems & Design, Singapore University of Technology and Design, 20 Dover Drive, 138682, Singapore*

Abstract

The operation of complex environmental systems usually accounts for multiple, conflicting objectives, whose presence imposes to explicitly consider the preference structure of the parties involved. Multi-Objective Markov Decision Processes are a useful mathematical framework for the resolution of such sequential, decision-making problems. However, the computational requirements of the available optimization techniques limit their application to problems involving few objectives. In real-world applications it is therefore common practice to select few, representative objectives with respect to which the problem is solved. This paper proposes a dimensionality reduction approach, based on the Non-negative Principal Component Analysis (NPCA), to aggregate the original objectives into a reduced number of principal components, with respect to which the optimization problem is solved. The approach is evaluated on the daily operation of a multi-purpose water

*Corresponding author. Tel.: +39-02-2399-4030.
E-mail address: matteo.giuliani@polimi.it

reservoir (Tono Dam, Japan) with 10 operating objectives, and compared against a 5-objectives formulation of the same problem. Results show that the NPCA-based approach provides a better representation of the Pareto front, especially in terms of consistency and solution diversity.

Keywords: Many-objective Optimization, Markov Decision Processes, Non-negative Principal Component Analysis, Visual analytics, Water Resources Management

---

1. Introduction

Contemporary environmental decision-making problems are often framed in heterogeneous socio-economic and ecologic contexts that involve multiple, conflicting and non-commensurable objectives. In such multi-objective contexts, the traditional concept of optimality is replaced by that of Pareto efficiency, which imposes the need to explicitly consider the preference structure of the parties involved (Zagonari and Rossi, 2013). When the number of objectives is equal or larger than four units, the problems are considered to take a *many-objective* nature, in contrast to multi-objective problems having three or less objectives (Farina and Amato, 2002; Fleming et al., 2005). For example, the design of an operating policy for a water reservoir with water quantity objectives (e.g. hydropower production and irrigation supply) requires considering few objectives only, but accounting for in-reservoir and downstream water quality targets can easily increase the number of operating objectives to ten or more units (Chaves and Kojiri, 2007).

Multi-objective Markov Decision Processes (MOMDPs) provide a useful mathematical framework for both analysis and resolution of these sequential decision-making problems (White, 1982, 1988). The traditional approach to solve a MOMDP is to convert a multi-objective problem to a family of single-objective problems, by emphasising one particular Pareto efficient solution at a time. Then, the problem can be solved by means of standard single-objective optimization techniques, such as Dynamic Programming (DP) family methods (Powell, 2007; Busoniu et al., 2010). The two most common scalarization techniques are the *weighted sum* and *ε-constraint methods* (Gass and Saaty, 1955; Haimes et al., 1971). The former is based on a linear combination of the objectives, while with the latter the conversion to a set of single-objective problems is obtained by transforming all the objectives, but one, into constraints. The main drawback of this approach stands in its computational intensity: the repetitions of single-objective problems scales exponentially with the number of objectives, thus making the approach feasible only for problems characterised by few objectives. Moreover, the accuracy in the approximation of the Pareto front might be scarce, with a limited solution diversity due to the non-linear relationships between the values of the weights (or constraints) and the corresponding objectives values.

An interesting alternative stands in the extension of single-objective Reinforcement Learning (RL) techniques (single-policy) to multi-objective problems (multi-policy). While the former aims to learn the single policy that best satisfies a set of preferences between objectives, as specified by a user or derived from the problem domain, the latter seeks to find a set of policies which approximates the Pareto front (Vamplew et al., 2011). Barrett and Narayanan (2008) and Lizotte et al. (2010) recently proposed two multi-objective RL methods that find in parallel the operating policies lying on the Pareto convex hull without an explicit search in the weights space. Pianosi et al. (2013) and Castelletti et al. (2013a) applied multi-objective RL to environmental systems by proposing a multi-objective extension of the Fitted Q-Iteration algorithm (Ernst et al., 2005; Castelletti et al., 2010) to design a two-objective reservoir operating policy. Other applications to environmental and water resources systems were proposed by Bone and Dragicevic (2009) and Shabani (2009). The main advantage of multi-objective RL stands in its capability of handling simultaneously multiple-objectives, although its effectiveness is currently limited to few objectives (Vamplew et al., 2011).

When dealing with MDPs characterised by several objectives, it is therefore common practice to select a priori few, representative objectives with respect to which the problem is then solved. This is done by studying the correlation between the objectives, or by direct interaction with the stakeholders (Soncini-Sessa et al., 2007). Although a conflict exists between some objectives, it is possible that others behave in a non-conflicting manner and some objectives can be discarded to obtain a lower-dimensional problem. In other terms, the original many-objective problem is simplified and re-formulated as a multi-objective one. However, this simplification comes at a price, as including all the objectives gives a number of benefits. First, transitioning to higher dimensional many-objective formulations may reveal that lower dimensional results represent extreme corners of the objective space that have little interest for decision-makers (see Kollat et al. (2011); Woodruff et al.

(2013), and references therein). Second, many-objective representations of tradeoffs help in reducing the negative impacts from two forms of decision bias (Brill. et al., 1990; Reed et al., 2013), namely *cognitive myopia* (Hogarth, 1981) and *cognitive hysteresis* (Gettys and Fisher, 1979). An example of how many-objective optimization is used to overcome these decision biases is given by Kasprzyk et al. (2012, 2013).

Another approach to the resolution of MOMDPs stands in the adoption of Multi-Objective Evolutionary Algorithms (MOEAs). The idea is to re-formulate the policies design problem as a Parameterization-Simulation-Optimization one (Koutsoyiannis and Economou, 2003), in which the policy is parameterized with an appropriate family of functions, and a MOEA is used to search for the best Pareto-efficient parameterizations (Kim et al., 2008). The main advantage of this approach is that MOEAs simultaneously handle many objectives (Reed et al. (2013) and references therein), and indeed they have been adopted for a broad spectrum of environmental and water resources problems, e.g. management of groundwater resources (Giustolisi et al., 2008), design of water distribution systems (Wu et al., 2013), hydrologic model calibration (Zhang et al., 2013), air quality planning (Carnevale et al., 2012) and design of wastewater treatment plants (Hakanen et al., 2013). Yet, their application is often limited to relatively simple problems, where an appropriate family of functions for the operating policy is chosen by relying on the empirical knowledge of the system behaviour. When dealing with complex systems, the empirical knowledge cannot guide this choice, since the operating policy has multiple inputs (large system state) and outputs (several control points). Selecting an unsuitable family of functions can then strongly influence the final result, with no guarantees on the optimality of the polices obtained as with DP or RL methods (Castelletti et al., 2013a).

The purpose of this paper is to propose a dimensionality reduction approach that assists DP and RL methods in the resolution of many-objective MDPs. As discussed in Galelli et al. (2011), the approach relies on the idea of exploiting the numerical correlation between the objectives to aggregate them into a reduced number of *principal components*, which are linear combinations of the original objectives. The reduced-dimensional MDP problem is then solved with respect to these components, and the value of the original objectives is eventually computed. The idea of reducing the complexity of many-objective optimization problems by exploiting the correlation between some objectives has been explored for the development of some MOEAs, which adopt Principal Component Analysis (PCA) techniques to progress iteratively from the interior of the search space towards the Pareto-optimal region by adaptively finding the correct lower-dimensional interactions (see Brockhoff and Zitzler (2006); Deb and Saxena (2006a); Brockoff and Zitzler (2007); López Jaimes et al. (2008); Brockhoff and Zitzler (2009); López Jaimes et al. (2009)). Yet, all these methods are developed for numerical, non dynamic, case studies. In this study, Non-negative Principal Component Analysis (NPCA, Zass and Shashua (2007)), which provides a combination of the original objectives with all the coefficients defined as positive, is not used to select the most relevant objectives, but rather to combine them in a reduced number of components. The advantage of the proposed approach is threefold: *i*) although being aggregated and projected into a lower dimensional space, all the original objectives of the many-objective MDP problem are considered, with the direction of optimization guaranteed by the positive coefficients; *ii*) the approach can be applied to any many-objective MDP with little a priori knowledge of the system behaviour, since it is based on the numerical correlation between the objectives; *iii*) the reduction of the number of objectives allows solving the MDP problem by means of DP and RL methods as it reduces the computational complexity of the many-objective MDP.

The NPCA-based approach is evaluated on a real-world case study, namely the daily operation of Tono Dam (Japan), a water reservoir managed for both quantity and quality targets, with up to 10 operating objectives. The evaluation of the results is performed in two stages. Firstly, we compare the results obtained in this study against those presented by Castelletti et al. (2013b), who previously considered a 5-objectives formulation of the same problem. Given the high-dimensional solution sets, the results are graphically analysed by means of visual analytics techniques (Kollat and Reed, 2007), which are becoming a common tool in environmental decision-making since the seminal work of Lotov et al. (2004). Secondly, we provide a multi-criteria assessment to account for convergence, consistency, and diversity of the obtained solutions (Reed et al., 2013).

2. Methods and Tools

2.1. Problem formulation

A discrete-time, continuous MOMDP is described as a tuple $< X, U, P, R, \gamma, \mu >$, where $X \subset \mathbb{R}^{N_x}$ is the state space, $U \subset \mathbb{R}^{N_u}$ the control (decision) space, $P(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)$ the conditional probability distribution of state $\mathbf{x}_{t+1}$ given the couple $\mathbf{x}_{t+1}, \mathbf{u}_t$ (i.e., Markov property), $R(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1}) = [g^1_{t+1}(\cdot), \ldots, g^k_{t+1}(\cdot)]$ a k-dimensional vector of immediate cost functions specifying the costs associated to the transition from state $\mathbf{x}_t$ to state $\mathbf{x}_{t+1}$ under the control $\mathbf{u}_t$, $\gamma \in (0, 1]$ a discount factor, and μ the initial state distribution from which the initial state is drawn. A control (operating) policy is a mapping from states to controls, i.e. $\pi : X \to U$, so that $\mathbf{u}_t = \pi(\mathbf{x}_t)$. For example, in a water reservoirs system the state variables are the storage and water quality levels in each reservoir, the control variables are the release decisions at each dam gate, the transition density is the probability of the next storage and water quality level $\mathbf{x}_{t+1}$ given the current state $\mathbf{x}_t$ and control $\mathbf{u}_t$, and $R(\cdot)$ accounts for the immediate costs associated to the different water-related interests, e.g. hydropower production, flood prevention, irrigation supply, and water quality maintenance.

The cost of following a certain policy π starting from state $\mathbf{x}_t$ at time t up to the end of the design horizon is formalized by the set of value functions $V^\pi(\mathbf{x}_t) = [V^{\pi,1}(\mathbf{x}_t), \ldots, V^{\pi,k}(\mathbf{x}_t)]$, with the i-th element defined as:

$$V^{\pi,i}(\mathbf{x}_t) = \int_X \left(g^i_{t+1}(\mathbf{x}_t, \pi(\mathbf{x}_t), \mathbf{x}_{t+1}) + \gamma V^{\pi,i}(\mathbf{x}_{t+1}) \right) P(\mathbf{x}_{t+1}|\mathbf{x}_t, \pi(\mathbf{x}_t)) d\mathbf{x}_{t+1} \tag{1}$$

Given the initial-state distribution μ, the i-th objective is defined as the expected return of the policy π from time $t = 0$ on, i.e.

$$J_\mu^{\pi,i} = \int_X V^{\pi,i}(\mathbf{x}_0)\mu(d\mathbf{x}_0) \tag{2}$$

and the vector of objectives is $\mathbf{J}_\mu^\pi = [J_\mu^{\pi,1}, \ldots, J_\mu^{\pi,k}]$. With this formulation, the expected cost is the statistic used to filter the uncertainty due to the presence of stochastic disturbances (e.g., precipitation, inflows).

Solving a MOMDP means finding the set of Pareto-optimal policies Π^* that maps onto the Pareto front in the space of the objectives $\mathcal{J}^* = \{\mathbf{J}^{\pi^*} | \pi^* \in \Pi^*\}$, meaning that a solution cannot be improved in a given objective without degrading its performance in another objective. The traditional approach to solve a MOMDP is to transform it into a family of single-objective problems by combining the k different immediate costs with some *scalarizing function* $\psi : \mathbb{R}^k \to \mathbb{R}$ (Perny and Weng, 2010). The most common approach to choose ψ is a convex combination of the immediate costs (weighting method) using a vector of weights $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_k] \in \Lambda^{k-1}$, where Λ^{k-1} is the unit $(k-1)$-dimensional simplex (so that $\sum_{i=1}^k \lambda^i = 1$ and $\lambda^i \geq 0 \ \forall i$). Each vector of weights $\boldsymbol{\lambda}$ therefore defines a single-objective MDP with the immediate cost function defined as

$$R_{\boldsymbol{\lambda}}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1}) = \sum_{i=1}^k \lambda^i g_{t+1}^i(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1}) \tag{3}$$

The single-objective MDP is then solved by finding the operating policy that minimises the value function $V_{\boldsymbol{\lambda}}(\cdot)$ in each state. In control problems, it is usually better to consider the *action-value function*, i.e. the value of taking the control $\mathbf{u}_t$ in state $\mathbf{x}_t$ and following the policy π thereafter. The optimal action-value function is the solution of the Bellman equation (Bellman, 1957) reformulated as:

$$Q^*_{\boldsymbol{\lambda}}(\mathbf{x}_t, \mathbf{u}_t) = \int_X \left(R_{\boldsymbol{\lambda}}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1}) + \gamma \min_{\mathbf{u}_{t+1} \in U} Q^*_{\boldsymbol{\lambda}}(\mathbf{x}_{t+1}, \mathbf{u}_{t+1}) \right) P(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t) d\mathbf{x}_{t+1} \tag{4}$$

Given the optimal action-value function, the associated optimal operating policy is the one that takes, in each state, the control with the lowest value, i.e.

$$\pi^* = \arg \min_{\mathbf{u}_t \in U} Q^*_{\boldsymbol{\lambda}}(\mathbf{x}_t, \mathbf{u}_t) \tag{5}$$

Each single-objective MDP yields one solution on the Pareto front. Since all the optimal policies of the single-objective MDPs are provably Pareto-optimal solutions of the original MOMDP (Chatterjee et al., 2006), the Pareto front is estimated by computing the set of objective vectors for all the possible values of $\boldsymbol{\lambda}$. In practice, an approximation of the set of Pareto-optimal policies Π^*, and the corresponding Pareto front, is obtained by considering a finite number n_λ of weight combinations and solving the associated n_λ single-objective MDPs. The main advantage of using the weighting method is that it computes Pareto efficient solutions only, which can be found by means of DP or RL methods. However, the repetition of single-objective problems increases exponentially with the number of immediate costs (or objectives) k, and this makes the computational complexity of the whole optimization process impractical for values of k larger than few units. Another limitation of this approach is that some Pareto-optimal policies may not be found, regardless of how many combinations of weights are used, if they lie in concave regions of the Pareto front (Vamplew et al., 2008).

Interactive, adaptive approaches (e.g., reference point method (Wierzbicki, 1980), Pareto race (Korhonen and Wallenius, 1988)) have been developed in order to interactively explore the Pareto front without having to fully compute it in advance, thus mitigating the associated computational burden (e.g., Deb et al., 2006b). Yet, the complexity and high number of questions to be posed to the DM remain an unsolved problem (Larichev, 1992).

2.2. Objective Reduction via Non-negative PCA

A feasible approach to reduce the problem complexity stands in aggregating the original k objectives into n linear combinations (with $n < k$), which then act as objectives in a lower dimensional MOMDP problem. An effective, yet informative, reduction may be obtained with PCA (Joliffe, 2002), a dimensionality reduction technique that provides linear combinations of the original variables with the coefficients of the combinations (the principal vectors) forming a low-dimensional sub-space corresponding to the directions of the maximal variance in the original data. Few (say n) principal components explain a high percentage of the variance of the original k variables. Moreover, the representation of the data in the projected space is uncorrelated, thus providing a useful tool for physical and statistical interpretations. Finally, from a computational point of view, PCA is quickly performed via an eigenvalue decomposition of the data covariance matrix. However, the adoption of PCA to reduce the dimensionality of the objective vector in a MOMDP is limited by the fact that the coefficients defining the components can be both positive and negative, with no guarantee on the direction of optimization of the original objectives, when these latter are replaced by the principal components (Galelli et al., 2011). This drawback can be eliminated by adding a non-negativity constraint to the original formulation of PCA, leading to the Non-negative Principal Component Analysis (NPCA, see Zass and Shashua (2007)).

To introduce the mathematical formulation of NPCA, let $\mathbf{J}^1, \ldots, \mathbf{J}^N \in \mathbb{R}^k$ form a zero-mean collection of N data points (i.e. N evaluations of the k-dimensional objective vector $\mathbf{J}$), arranged as the columns of the matrix $\mathcal{G} \in \mathbb{R}^{k \times N}$, and $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^n \in \mathbb{R}^k$ be the desired n principal components, arranged as the columns of the matrix $\mathcal{P} \in \mathbb{R}^{k \times n}$. Adding a non-negative constraint to the PCA formulation, which maximises the explained variance by principal components, and relaxing the orthonormality constraint on the desired components, which prevents the computation of a disjoint matrix $\mathcal{P}$ (for further details see Zass and Shashua (2007)), gives the following problem, whose solution is $\mathcal{P}$:

$$\max_{\mathcal{P}} \frac{1}{2}\|\mathcal{P}^T \cdot \mathcal{G}\|_{fr}^2 - \frac{\alpha}{4}\|I - \mathcal{P}^T \cdot \mathcal{P}\|_{fr}^2 \tag{6a}$$

subject to

$$\mathcal{P} \geqslant 0 \tag{6b}$$

where $\|\cdot\|_{fr}^2$ is the square Frobenius norm, I the identity matrix, $\|I - \mathcal{P}^T \cdot \mathcal{P}\|_{fr}^2$ a non-negative orthonormality distance measure that vanishes if $\mathcal{P}$ is orthonormal (like in the original PCA formulation), and α (≥ 0) a parameter balancing between data reconstruction and orthonormality. The higher the value of α, the higher is the importance of the orthonormality distance, potentially forcing the the orthogonality of the principal components. On the other side, the lower the value of α, the lower is the importance given to orthonormality, thus allowing more overlapping among the components yielding to a better reconstruction of the original data. Notice that relaxing the disjoint property of NPCA implies a relaxation in the maximum variance property of PCA, with the parameter α allowing the exploration of the tradeoff. A more detailed discussion on the role of the parameter α is reported in Appendix A.

The resolution of problem (6) yields a set of non-negative and partially overlapping principal components $[\mathbf{p}^1, \ldots, \mathbf{p}^n]$ that can effectively replace the k-dimensional objectives in the original MOMDP problem. This latter is then solved by means of DP or RL methods, and the optimal policies so obtained are Pareto-optimal solutions of the problem defined with respect to the n non-negative principal components. Finally, the values of the original k objectives are evaluated.

3. Case study: Tono Dam

3.1. System description

Tono Dam is located at the confluence of Kango and Fukuro rivers (Figure 1a), in the western part of Japan. The construction works were completed in 2011. With a height of 75 m (Figure 1b), the dam forms an impounded reservoir of 12.4 x 10^6 m^3 (gross capacity), with a surface area of 0.64 km^2 and fed by a 38.1 km^2 catchment. The construction of the dam aims at supporting agriculture, enhancing the recreational value of the reservoir and protecting the riverine ecosystems potentially threatened by the dam's operation. Due to the region's local climate, the reservoir is characterized by prolonged periods of stratification that negatively impact the water quality both in-reservoir and in the reservoir's outflow. The dam was therefore equipped with a Selective Withdrawal System (SWS, see Bohan and Grace (1973)). Fifteen vertically stacked siphons allow the dam to release water at different depths with different physico-chemical properties, and blending is allowed. The obtained flexibility in the selection of the outlet offers advantages in order to meet water quality targets when the reservoir is stratified or to respond to short term inflow events (Gelda and Effler, 2007). The possibility of designing a multi-purpose operating strategy for the SWS is studied in Castelletti et al. (2013b). Indeed, the operation of the dam directly impacts on different water sectors, which are classified as in *in-reservoir*, affected by level variations, and *downstream*, dependent on the release. Two sectors belong to the first class: recreation, aiming to keep high reservoir levels and prevent algal blooms, and silting, whose objective is to maximize the sediments evacuation. Two sectors belong to the second class: irrigation, aiming to reduce the water supply deficit (which has a direct effect on the seasonal harvest), and environment, whose goal is to protect the downstream riverine ecosystem, potentially threatened by large deviations of the water temperature from the seasonal natural patterns.

3.2. Operating objectives

In order to evaluate alternative SWS operating strategies, one (or more) immediate cost function $g^i_{t+1}(\cdot)$ is (are) defined for each sector. The i-th operating objective $J^i(\cdot)$ is then defined as the daily average of the corresponding immediate cost $g^i_{t+1}(\cdot)$. The definitions of the immediate cost functions are as follows:

- *Level*: the squared positive difference of reservoir level h_{t+1} with respect to the reference level $\bar{h} = 182.8$ m a.s.l.:

$$g_{t+1}^{Lev} = \left(\max\left(\bar{h} - h_{t+1}, 0\right)\right)^2 \tag{7}$$

- *Algae*: the daily average hourly maximum concentration of chlorophyll-a (Chl-a) in the see-through layer:

$$g_{t+1}^{Algae} = \frac{1}{24}\sum_{\tau=1}^{24} \max_{z_\tau \in z_E}(chla_\tau(z_\tau)) \tag{8}$$

where $chla_\tau$ is the Chl-a concentration [μg/L] at the τ-th hour of day t, z_τ is the depth with respect to the reservoir surface, z_E is the see-through layer depth set at 7 m below water surface (where the thermocline is generally formed in summer).

- *Sedimentation*: the daily volume of sediment expelled with the release, which has to be maximized in order to reduce the silting of the reservoir and increase its expected life:

$$g_{t+1}^{Sed} = TSS_{t+1}^{out} \tag{9}$$

where TSS_{t+1}^{out} is the amount of Total Suspended Solid [g/day] in the reservoir outflow between t and $t+1$ computed as

$$TSS_{t+1}^{out} = \sum_{i=1}^{n} tss_{t+1}^{i} r_{t+1}^{i} + tss_{t+1}^{spill} r_{t+1}^{spill} \tag{10}$$

where tss_{t+1}^{i} is the average TSS concentration [g/m^3] of the water released by the i-th controlled siphon, and tss_{t+1}^{spill} is the average TSS concentration [g/m^3] of the water released by the spillway, and r^i_{t+1} and r^{spill}_{t+1} are the corresponding released volumes [m^3/day].

- *Irrigation*: the squared water daily deficit with respect to the agricultural water demand w_t:

$$g^{Irr1}_{t+1} = \beta_t \left(\max\left(w_t - (r_{t+1} - q^{MEF}_{t+1}), 0\right)\right)^2 \tag{11}$$

where r_{t+1} is the total release from the dam (including SWS and spillway), q^{MEF}_{t+1} is the minimum environmental flow, and β_t is a time-varying coefficient taking into consideration the different relevance of the water deficit in different periods of the year. In particular, the immediate cost is elevated to the second power to favour operating policies that reduce severe deficits in a single time step, while allowing for more frequent, small shortages, which cause less damage to the crop. This ensures that vulnerability is a minimum (Hashimoto et al., 1982).

In addition, four other immediate costs are introduced: the first one (g^{Irr2}_{t+1}) is the daily deficit expressed as m^3/s (i.e., $g^{Irr2}_{t+1} = (w_t - (r_{t+1} - q^{MEF}_{t+1}))^+$). The remaining (i.e. g^{Irr3}_{t+1}, g^{Irr4}_{t+1}, g^{Irr5}_{t+1}) are defined in the same way, but they consider a shorter inter-annual period, namely winter (from December 21st to March 20th), May and summer (from June 21st to September 21st).

- *Temperature*: the squared difference between the inflow and outflow temperature (as in Fontane et al. (1981) and Baltar and Fontane (2008)):

$$g^{Temp1}_{t+1} = (T^{out}_{t+1} - T^{in}_{t+1})^2 \tag{12}$$

where T^{out}_{t+1} is the average temperature in a section just downstream of dam outlet and $T^{in}_{t+1} = \frac{T^K_{t+1} a^K_{t+1} + T^F_{t+1} a^F_{t+1}}{a^K_{t+1} + a^F_{t+1}}$ with T^K and T^F being the average temperature [°C] of the inflow respectively in the Kango and Fukuro rivers, and a^K_{t+1} and a^F_{t+1} the corresponding flows.

As for the case of the irrigation objectives, a more intuitive immediate cost g^{Temp2}_{t+1} is defined as the daily difference of temperature between the inflow and the outflow, expressed in °C.

The optimal operation of Tono Dam SWS requires accounting for the above ten immediate cost functions and the associated operating objectives, i.e. J^{Lev}, J^{Algae}, J^{Sed}, J^{Irr1}, J^{Irr2}, J^{Irr3}, J^{Irr4}, J^{Irr5}, J^{Temp1}, J^{Temp2} (see Figure 2 for a schematic representation of the hierarchy of water sectors and objectives). A first, approximate solution to this problem is described in Castelletti et al. (2013b) and Giuliani et al. (2013), who selected five operating objectives considered representative of the water sectors.

4. Experimental setting

4.1. Models

The design and evaluation of different management alternatives requires modeling the main hydrodynamic and ecological processes characterizing the reservoir. To this purpose, we adopted the coupled 1D DYRESM-CAEDYM model (Hipsey et al., 2006; Imerito, 2007). The 1D hydrodynamic model DYRESM (Dynamic Reservoir Simulation Model) simulates the vertical distribution of temperature, salinity and density in the reservoir, while the aquatic ecosystem model CAEDYM (Computational Aquatic Ecosystem Dynamics Model) simulates a range of biological, chemical and physical processes, commonly related with water quality characteristics (such as total phosphorus, total nitrogen, chlorophyll-a, etc.). The SWS ability to release water at different depth is modeled by two decision variables, u^{-3} and u^{-13}, representing the volumes to be released in a decision time-step (i.e., one day) at 3 and 13 meters below the water surface. In both cases, the decision is defined with respect to the water body surface (see Figure 1b). These water depths should correspond, respectively, to the epilimnium and the hypolimnium of the stratified reservoir. As in Castelletti et al. (2013b), we do not model all the fifteen outlets as this would make the problem computationally impracticable.

4.2. Data-set Generation

In order to identify n principal components, a zero-mean collection of N data-points is required. To this purpose, the 1D DYRESM-CAEDYM model was run over the hydro-meteorological period 1995-2006 under 100 different release scenarios pseudo-randomly generated with the aim of exploring the state-decision space as more homogeneously as possible. In particular, the decision vectors $\mathbf{u}_t$ were generated with probability equal to 1/3 of opening the siphon at -3 m only, the same probability for the siphon at -13 m and, finally, probability equal to 1/3 of opening both the controlled siphons. The sampling was performed using quasi-random sequences and an irregular grid with lower probability assigned to high release values in order to reduce the occurrence of full reservoir drawdown. For each of the 100 simulations, the ten objectives are computed as the daily average of the immediate costs $g^i_{t+1}(\cdot)$ (with $i = 1, \dots, 10$) defined in Section 3.2. The normalized realisations of the objective vector (i.e., zero mean and unit standard deviation) are arranged in the matrix $\mathcal{G} \in \mathcal{R}^{10\times100}$ from which the principal components are extracted, as described in Section 5.1.

4.3. Optimization Algorithm

To design the operation of Tono Dam an optimization algorithm able to consider water quality and quantity targets is needed. In this work, in order to compare the results against those found in Castelletti et al. (2013b), the same batch-mode RL algorithm, i.e. Fitted Q-iteration (Ernst et al., 2005; Castelletti et al., 2010), is adopted. The algorithm combines RL concepts of off-line learning and functional approximation of the value function, from which the policy is derived, using tree-based regression (Geurts et al., 2006; Galelli and Castelletti, 2013). The optimal operating policy is determined on the basis of experience samples represented as a finite data-set $\mathcal{F}$ of tuples of the form $< t, \mathbf{x}_t, \mathbf{u}_t, t+1, \mathbf{x}_{t+1}, g_{t+1} >$, where the state variables $\mathbf{x}_t$ are the reservoir level h_t, the temperature T_t^i and the total suspended solid TSS_t^i in the 1D model layer corresponding to the outlet controlled by the decision variables u_t^i (with i = -3; -13). In this study, the adopted version of the Fitted Q-iteration algorithm solves one single-objective problem at each optimization run, so the immediate costs g_{t+1} are defined according to the weighting method as in eq. (3), using the same weights as in Castelletti et al. (2013b). The data-set $\mathcal{F}$ has to be previously collected from the system or simulations thereof, i.e. a variety of system conditions experienced by the system under different combinations of release decisions and external driver realizations with the associated resulting immediate costs. In order to construct the data-set $\mathcal{F}$, we used the 100 simulations of the 1D DYRESM-CAEDYM model with pseudo-random release scenarios. In synthesis, the overall modeling and optimization procedure is represented in Figure 3.

4.4. Performance Evaluation

In order to provide a quantitative evaluation of the obtained solutions (i.e., a 10-objective Pareto front), it is necessary to consider multiple criteria that account for different aspects, such as the proximity of a set of solutions to the Pareto optimal front (or its best known approximation) or the capacity of representing the full extent of tradeoffs. In this work we adopt three metrics, i.e. generational distance, additive ε-indicator and hypervolume, which respectively account for convergence, consistency and diversity (Knowles and Corne, 2002; Zitzler et al., 2003).

The generational distance I_{GD} measures the average Euclidean distance between the points in an approximation set S and the nearest corresponding points in the reference set $\bar{S}$, and it is defined as

$$I_{GD}(S, \bar{S}) = \frac{\sqrt{\sum_{\mathbf{s} \in S} d_{\mathbf{s}}^2}}{n_S} \tag{13a}$$

where n_S is the number of points in S, and $d_{\mathbf{s}}$ the minimum Euclidean distance between each point in S and $\bar{S}$. Assuming that the two sets S and $\bar{S}$ correspond to two sets of objectives $J^i(\mathbf{s})$ and $J^i(\bar{\mathbf{s}})$ $(i = 1, \dots, k)$, the distance $d_{\mathbf{s}}$ is defined as

$$d_{\mathbf{s}} = \min_{\bar{\mathbf{s}} \in \bar{S}} \sqrt{\sum_{i=1}^{k} [J^i(\mathbf{s}) - J^i(\bar{\mathbf{s}})]^2} \tag{13b}$$

I_{GD} is a pure measure of convergence, so it requires only a single solution close to the reference set to attain ideal performance.

The additive ε-indicator I_ε measures the worst case distance required to translate an approximation set solution to dominate its nearest neighbour in the reference set. It is defined as

$$I_\varepsilon(S, \bar{S}) = \max_{\bar{\mathbf{s}} \in \bar{S}} \min_{\mathbf{s} \in S} \max_{1 \le i \le k} (J^i(\mathbf{s}) - J^i(\bar{\mathbf{s}})) \tag{14}$$

This metric is very sensitive to gaps in tradeoffs and is viewed as a measure of consistency.

Finally, the hypervolume I_H measures the volume of objective space dominated by an approximation set, i.e.

$$I_H(S, \bar{S}) = \frac{\int \alpha_S(\mathbf{s}) ds}{\int \alpha_{\bar{S}}(\bar{\mathbf{s}}) d\bar{s}} \tag{15a}$$

with

$$\alpha(\mathbf{s}) = \begin{cases} 1 & \text{if } \exists \mathbf{s}' \in S \text{ such that } \mathbf{s}' \preceq \mathbf{s} \\ 0 & \text{otherwise} \end{cases} \tag{15b}$$

This metric captures both convergence and diversity.

Overall, a good solution is characterised by low values of the first two criteria and a high value of the third one.

5. Application Results

5.1. NPCA Analysis

5.1.1. Analysis of the correlation matrix

The correlation matrix of the ten objectives evaluated over the 100 management scenarios is reported in Table 1. In particular, J^{Irr1} is positively correlated with all the other irrigation objectives, and this somewhat justifies the choice of considering it representative of this sector (Castelletti et al., 2013b). Indeed, J^{Irr1} has a strong correlation with both J^{Irr2} and J^{Irr5} and a weaker correlation with J^{Irr3} and J^{Irr4}. This seems to suggest that the five irrigation objectives, although correlated, capture different information: the irrigation deficits of J^{Irr1} and J^{Irr2} are mainly related to the deficit in summer J^{Irr5}, while high deficits in either winter or May are not completely reflected in high values of J^{Irr1}. A strong correlation exists between J^{Temp1} and J^{Temp2}, and these latter are also correlated to J^{Algae}. Indeed, releasing large volumes of water reduces the concentration of nutrients in the reservoir, thus preventing algal blooms, and maintains similar temperature patterns between inflow and outflow. J^{Lev} and J^{Sed} are weakly correlated and have no relevant positive correlations with the other objectives. The most relevant conflict is between J^{Lev} on one side and J^{Algae}, J^{Temp1}, J^{Temp2} on the other. This conflict is not surprising as the high releases that produce low values of J^{Algae}, J^{Temp1} and J^{Temp2} tend to drawdown the reservoir level. Moreover, both J^{Lev} and J^{Sed} are anti-correlated with all the irrigation objectives, since releasing small volumes of water keeps the reservoir at high levels but produces significant irrigation deficits, while releasing large volumes of water flushes out the sediments but reduces the water availability for irrigation supply. Finally it is worth noting that J^{Irr3} and J^{Irr4} have no either positive or negative correlations. They seem quite independent with respect to the other objectives, probably because the specific criteria they account for (i.e., the irrigation deficit in winter and May, respectively) are not captured by the other objectives.

5.1.2. *Identification of the components*

Given the matrix $\mathcal{G}$ of the ten objectives realizations and the corresponding correlation matrix, the NPCA algorithm requires defining the number n of components to extract. Choosing the 'exact' value of n is not straightforward, because it is necessary to balance the dimensionality reduction with the effective representation of the original variables (objectives). Few components substantially reduce the dimension of the objective vector, but may not take into account all the information contained in $\mathcal{G}$. On the other hand, considering many components tends to decrease the effectiveness of the reduction process. Figure 4 represents the percentage of variance explained by the principal components as a function of n. The results are reported for both the non-negative principal components (red bars) and the principal components obtained with the original PCA formulation (blue bars). In the case of NPCA, the value of the parameter α is defined via trial and error analysis (further details are given in Appendix A). The variance explained via PCA is reported as a benchmark, since it represents the maximum variance that could be explained. Indeed, the non-negative constraint introduced by the NPCA, along with the relaxation of the orthonormality constraint of PCA, reduces the variance explained by the non-negative principal components. Assuming the value of 75% as a reference for a good representation of the original objectives (Joliffe, 2002), five non-negative principal components are extracted. Also, this choice allows the development of an effective comparison with the results discussed in Castelletti et al. (2013b), where the problem is solved with the same number of objectives.

The values of the coefficients defining the five components are reported in Table 2. The coefficients reflect the correlation between the objectives reported in Table 1: the first component seems to represent the irrigation sector, having high coefficients for J^{Irr1}, J^{Irr2} and J^{Irr5}, which are indeed all strongly correlated. The second one is mainly related to J^{Algae}, J^{Temp1} and J^{Temp2}, thus confirming that these objectives are physically correlated. The third and fourth components are basically related to J^{Irr4} and J^{Irr3} respectively, possibly because the deficit in winter and May represent a different process with respect to the other irrigation objectives. Finally, J^{Sed} and J^{Lev} are projected on the fifth component, even though they are not strongly correlated.

5.2. Design of the operating policies

The optimal set of daily, periodic (with period equal to one year) release policies are obtained by solving the MOMDP problems with the Fitted Q-iteration algorithm, with the five operating objectives considered in Castelletti et al. (2013b) replaced by the five non-negative principal components. The weighting method is used to transform the 5-objective problem into a family of single-objective problems, with the same 36 combinations of weights as in Castelletti et al. (2013b). According to the procedure depicted in Figure 3 (dashed line), the 10 original objectives are eventually evaluated via simulation over the hydro-meteorological period 1990-1995. The results analysis is performed in three steps: firstly, we compare the solutions focusing only on the five objectives selected in Castelletti et al. (2013b) (Section 5.2.1); secondly, the same solutions are compared with respect to the remaining five objectives (Section 5.2.2); thirdly, the two approaches are compared with respect to the entire set of ten objectives (Section 5.2.3).

5.2.1. *First comparison -* J^{Algae}, J^{Temp1}, J^{Lev}, J^{Irr1} *and* J^{Sed}

Figure 5 shows the solutions with respect to the five objectives optimized in Castelletti et al. (2013b) (*selection-based formulation* in the followings), with the red and grey cones associated to the NPCA and selection-based formulation respectively. For both formulations it is evident that J^{Algae} and J^{Temp1} are not conflicting, and it is possible to minimize simultaneously the two objectives as there are many cones in the bottom-left part of the figure. Moreover, the best performing alternatives with respect J^{Algae} and J^{Temp1} negatively impact on J^{Lev}. This is because the optimal operation with respect to the first two objectives tends to release large volumes of water to flush out the nutrients and maintain similar temperatures between inflow and outflow, but it generates a drawdown of the reservoir level. Looking at the grey cones, it is possible to observe that J^{Algae} and J^{Temp1} are only partially conflicting with J^{Sed}: although the cones in the bottom-left corner have an intermediate inclination, some cones pointing upward are not far from that corner, and are characterized by small values of J^{Algae} and J^{Temp1}. On the other hand, J^{Sed} is in conflict with J^{Lev} as most of the cones on the right part of the figure, characterized by low values of J^{Lev}, point downward. The tradeoffs with respect to J^{Irr1} are more evident looking at the red cones: again, the conflict between J^{Algae} and J^{Temp1} seems weak, with the cones in the bottom-left corner having intermediate sizes. The smallest cones, characterizing the best solutions for J^{Irr1}, are in the center of the objective space and are horizontally oriented, meaning that a good performance for J^{Irr1} does not have a negative impact on the other objectives.

It can be observed that the NPCA-based solutions do not assume worse values than the selection-based ones, except for J^{Lev}. On average, the NPCA-based solutions produces better solutions with respect to J^{Algae} and J^{Temp1}, being most of the cones in the bottom-left part of the figure red. The second principal component, which has high coefficients for J^{Algae} and J^{Temp1}, is therefore effective in representing both objectives. Furthermore, also the best solutions with respect to J^{Irr1}, i.e. the smallest cones, are red. This is somewhat expected, since three of the five components are mainly related to irrigation objectives (see Table 2). The presence of grey as well as red cones with upward orientation indicates that a good performance in terms of J^{Sed} is obtained with both formulations. With respect to the NPCA-based solutions this means that the parameterisation of the fifth principal component (see Table 2) adequately represents this objective. On the other hand, the performance of the NPCA-based solutions is lower than the selection-based ones with respect to J^{Lev}. Unlike J^{Sed}, the fifth component does not effectively represent J^{Lev} due to the low coefficient assigned to this objective.

More details regarding the conflict between J^{Lev}, J^{Algae} and J^{Temp1} are illustrated in Figure 6a, which shows that most of the NPCA-based solutions (red points) are in the top part of the figure, with associated high values of J^{Lev}. Moreover, the best NPCA-based solution for this objective is set around the middle of the J^{Lev}-axis, thus confirming that these solutions penalise the water level objective. Figure 6b shows the superiority of the NPCA-based solutions according to J^{Algae} and J^{Temp1}, with most of the points in the bottom-left corner being red and, conversely, most of the grey points set on right half of the figure, corresponding to poor performance with respect to J^{Algae}.

5.2.2. Second comparison - J^{Irr2}, J^{Irr3}, J^{Irr4}, J^{Irr5} and J^{Temp2}

In Figure 7 the comparison is performed with respect to the five objectives that are not considered in Castelletti et al. (2013b), with the red and grey cones associated to the NPCA and selection-based solutions respectively. For both formulations most of the cones in Figure 7 are in the bottom-left corner, meaning that the objectives on the three primary axes are not significantly conflicting, and many alternatives produce good performance with respect to all these objectives. Note that there are many alternatives that are optimal for J^{Irr4} and have different values for J^{Irr3}, and viceversa. This is because these objectives are not strongly correlated. Looking at the orientation and the dimension of the cones, J^{Irr5} and J^{Irr2} do not appear to be strongly conflicting. These two objectives seem to be instead conflicting with J^{Temp2}, as the smallest and downward oriented cones are in the top half of the objective space. A weak conflict exists also between J^{Irr3} and J^{Irr4} with respect to J^{Irr5}, as the cones in the bottom-left corner are slightly upward oriented.

The NPCA-based solutions significantly outperform the selection-based ones for three of the five objectives, namely J^{Irr3}, J^{Irr4} and J^{Temp2}, with most of the cones in the bottom-left part of the figure being red. Moreover, the red cones are on average smaller than the grey ones, meaning that also the performance with respect to J^{Irr2} is more satisfactory. Finally, the results with respect to J^{Irr5} seem similar for the two formulations. Therefore, the proposed NPCA-based aggregation seems effective in enhancing the system operation with respect to the objectives that are not selected in the selection-based case. In particular, it is worth noting the differences in performance with respect to J^{Irr3} and J^{Irr4} (Figure 8a), which are the irrigation objectives less correlated to J^{Irr1}. In the selection-based formulation these objectives are considered redundant and the irrigation sector is represented by J^{Irr1} only. Yet, the information content of J^{Irr3} and J^{Irr4} (the water deficit in winter and May) is different from J^{Irr1} and their exclusion produces poorly performing alternatives. Furthermore, even though the correlation between J^{Temp1} and J^{Temp2} is high, the better performance of the NPCA-based solutions with respect to this latter (Figure 8b) suggests that also the information captured by these objectives is slightly different and it is not sufficient to optimize with respect to only one of them.

5.2.3. Third comparison - Full set of objectives

The parallel-coordinates plot in Figure 9 provides a comprehensive view of the solutions obtained with the two formulations with respect to the entire set of ten objectives. For illustration purposes the objectives are standardized (zero mean and unit standard deviation) and each axis is oriented so that the direction of preference is always downward. The ideal solution would be a horizontal line running along the bottom of all the axes. The tradeoff relationships among the objectives are represented by crossing line segments between two adjacent axes, see for example the large number of crossing lines between J^{Temp1} and J^{Lev} representing the strong conflict between these two objectives as discussed in Section 5.2.1. The placement of the axes has therefore a key role in highlighting the tradeoffs. Since the purpose of this section is not to discuss the different conflicts (as done in Section 5.2.1 and 5.2.2), but rather show the overall performance of the two approaches on the whole set of objectives, we arbitrarily set one specific configuration, namely the five objectives explicitly considered in Castelletti et al. (2013b) on the first five axes, followed by the remaining objectives. Besides highlighting some key tradeoffs between adjacent axes (e.g., J^{Temp1} and J^{Lev}), the information provided by the parallel-coordinates plot confirms the general findings discussed in the previous sections: the NPCA-based solutions (red lines) seem to be not inferior to the selection-based ones (grey lines) with respect to the five objectives explicitly considered in Castelletti et al. (2013b), other than J^{Lev}. The two approaches indeed cover the same range of performance on the first five axes, with no clear distinction between red and grey solutions. On the other hand, the NPCA-based solutions are clearly better than the selection-based ones with respect to the remaining five objectives, which are the ones not considered in Castelletti et al. (2013b). Most of the red solutions in the right-hand half of the figure are indeed placed lower than the grey ones, thus attaining better performance in these objectives.

A more detailed comparison can be done by focusing on two specific compromise alternatives, designated by the dashed and solid black lines in Figure 9. Their selection is a subjective evaluation by the authors and aims only at providing more details with respect to the representation of the entire set of Pareto efficient alternatives. With the purpose of equally accounting for all the objectives, we analyze in details the solutions obtained by setting $\lambda^i = 0.2$ (for $i = 1, \dots, 5$) in both formulations. Figure 10 reports the daily average value of the immediate cost functions computed over the period

1990-1995. The performance obtained for these alternatives further confirms that the proposed method seems effective in enhancing the system operation with respect to the objectives not considered in the selection-based formulation (right part of the figure), at the cost of very small worsening in the ones originally optimized (left part of the figure). Indeed, the NPCA-based solution (red line) is significantly better than the selection-based one (grey line) with respect to J^{Algae}, J^{Temp1} and J^{Irr1}. The performance of the two alternatives is similar with respect to J^{Sed}, while the NPCA-based solution is poorly performing for J^{Lev}. As discussed in Section 5.2.1, this is due to the low coefficient assigned to this objective in the definition of the fifth component. On the other hand, looking at the objectives not considered in the selection-based formulation, the NPCA-based solution is significantly better than the selection-based one with respect to J^{Irr3}, J^{Irr4} and J^{Temp2}, while it obtains similar irrigation deficit in J^{Irr2} and J^{Irr5}, which are more correlated with J^{Irr1}.

5.2.4. Multi-criteria assessment

Finally, a quantitative evaluation is obtained by computing the multiple criteria introduced in Section 4.4. The reference set, representing the best approximation of the 10-objective Pareto front, is defined as the set of non-dominated solutions selected in the union of the NPCA-based and selection-based Pareto optimal sets. A good solution should be characterized by low values of the first two metrics, namely generational distance I_{GD} and additive ε-indicator I_{ε}, and a high value in the hypervolume indicator I_H. As shown in Figure 11, the selection-based formulation has a better performance in terms of generational distance, meaning that it produces at least one or more solutions close to the reference set. This is not surprising, since the aggregation performed with NPCA does not allow the design of the extreme points of the Pareto front, i.e. the policies obtained by setting to zero all the weights but for one. These solutions, which for construction belong to the reference set being not-dominated by any compromise solution, are obtained with the selection-based formulation only and, therefore, the value of generational distance is very low. On the other side, the NPCA-based solutions have better performance with respect to both the additive ε-indicator and the hypervolume metrics. The selection-based solutions are indeed characterized by gaps in the tradeoffs involving the non-selected objectives, yielding to high values of additive ε-indicator. Furthermore, they are Pareto efficient with respect to five objectives only, thus reducing the volume dominated in the 10-objective space that is represented by low values of the hypervolume indicator.

6. Computational requirements

In order to ensure that the shape of the Pareto front is reasonably represented, the number M of Pareto efficient solutions is a priori selected. In particular, M is defined according to the following permutation (Ross, 2013)

$$M = \sum_{i=1}^{k} \frac{k!}{i!(k-i)!} + k \tag{16}$$

where k is the number of objectives considered. The underlying idea is to explore the Pareto front by computing the k extreme solutions, obtained by setting to zero all weights but for one, and some compromise solutions by relaxing the extremes and assigning the same weight to few objectives. The exploration of a ten-objective Pareto front thus requires designing 1033 Pareto optimal alternatives. Conversely, the adoption of the NPCA-based aggregation method allows exploring an approximation of the 10-objective Pareto front by solving a MOMDP whose objectives are the five non-negative principal components. Therefore, the number of alternatives to be generated is reduced to 36 only. Considering that the time required to design and simulate an operating policy on a 3.16 Ghz Intel Xeon QuadCore with 16 GB Ram is about 20 hours for each alternative, the exploration of the ten-objective Pareto front would require 20,660 hours (about 861 days, 2.4 years), while the 36 NPCA-based solutions require 720 hours (30 days).

7. Conclusions

In this work we presented a dimensionality reduction approach to solve many-objective Markov Decision Processes (MDPs) problems in environmental contexts. The approach relies on Non-negative Principal Component Analysis (NPCA), which is used to identify a lower dimensional representation of the original objectives and to obtain an approximated solution of the many-objective problem. The approach is demonstrated on the daily operation of a multi-purpose water reservoir (Tono Dam, Japan) involving 10 operating objectives. The comparison of the NPCA-based solutions with the ones obtained by selecting a subset of 5-objectives shows that the proposed approach is able to provide a better representation of the 10 objectives Pareto front, especially in terms of consistency and solution diversity. Moreover, the combination of this approach with visual analytics techniques makes it possible to explore the high dimensional formulation of the decision-making problem and attain insight about management alternatives that can be hidden in lower dimensional formulations. The proposed approach, being based on the numerical correlation between the objectives, can in principle be applied to any many-objective MDP with little a priori knowledge of the system behaviour, and therefore combined with any DP or RL method.

An important aspect of the NPCA-based approach that requires further investigation is the sub-optimality of the obtained solutions. As discussed in Franssen (2005), the optimization of aggregate measures does not optimise the individual performance criteria themselves, and aggregating preference across multiple criteria will always favour some criteria over others in a manner that is difficult to ascertain a priori. Thus, the resulting solutions can be biased towards a subset of performance objectives in ways that cannot be known a priori by decision-makers (Woodruff et al., 2013). Another aspect that will be considered is the interpretation of the aggregated objectives (principal components), which are designed to maximise the performance with respect to a particular set of preferences, but not to support the direct understanding of the solutions.

8. Acknowledgement

The first and second author are partially supported by *Fondazione Fratelli Confalonieri* and SRG ESD 2013 061 Start-up Research project respectively. The authors are also grateful to Matthew Woodruff, Christa Kelleher and Hiroshi Yajima.

References

Baltar, A., Fontane, D., 2008. Use of Multiobjective Particle Swarm Optimization in Water Resources Management. Journal of Water Resources Planning and Management 134 (3), 257–265.

Barrett, L., Narayanan, S., 2008. Learning all optimal policies with multiple criteria. In: Proceedings of the 25th International Conference on Machine Learning. Omnipress, pp. 41–47.

Bellman, R., 1957. Dynamic programming. Princeton University Press, Princeton.

Bohan, J., Grace, J., 1973. Selective withdrawal from man-made lakes. Tech. Rep. H-73-4, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississipi.

Bone, C., Dragicevic, S., 2009. Gis and intelligent agents for multiobjective natural resource allocation: A reinforcement learning approach. Transactions in GIS 13 (3), 253–272.

Brill., E., Flach, J., Hopkins, L., Ranjithan, S., 1990. MGA: A decision support system for complex, incompletely defined problems. IEEE Transactions on Systems, Man, and Cybernetics 20 (4), 745–757.

Brockhoff, D., Zitzler, E., 2006. Are all objectives necessary? on dimensionality reduction in evolutionary multiobjective optimization. In: Runarsson, T., Beyer, H.-G., Burke, E., Merelo-Guervas, J., Whitley, L., Yao, X. (Eds.), Parallel Problem Solving from Nature. Vol. 4193 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 533–542.

Brockhoff, D., Zitzler, E., 2009. Objective reduction in evolutionary multi objective optimization: Theory and applications. Evolutionary Computation 17 (2), 135–166.

Brockoff, D., Zitzler, E., September 25–28 2007. Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods. In: Proceedings of 2007 IEEE Congress on Evolutionary Computation. Singapore.

Busoniu, L., Babuska, R., De Schutter, B., Ernst, D., 2010. Reinforcement Learning and Dynamic Programming Using Function Approximators. CRC Press, New York.

Carnevale, C., Finzi, G., Guariso, G., Pisoni, E., Volta, M., 2012. Surrogate models to compute optimal air quality planning policies at a regional scale. Environmental Modelling & Software 34, 44–50.

Castelletti, A., Galelli, S., Restelli, M., Soncini-Sessa, R., 2010. Tree-based batch-mode reinforcement learning for optimal water reservoir operation. Water Resources Research 46 (9), W09507, doi: 10.1029/2009WR008898.

Castelletti, A., Pianosi, F., Restelli, M., 2013a. A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run. Water Resources Research 49, doi: 10.1002/wrcr.20295.

Castelletti, A., Yajima, H., Giuliani, M., Soncini-Sessa, R., Weber, E., 2013b. Planning the optimal operation of a multi–outlet water reservoir with water quality and quantity targets. Journal of Water Resources Planning and Management -, doi: 10.1061/(ASCE)WR.1943-5452.0000348.

Chatterjee, K., Majumdar, R., Henzinger, T., 2006. Markov decision processes with multiple objectives. In: Proceedings of STACS 2006. Vol. 3884 of Lecture notes in computer science. pp. 325–336.

Chaves, P., Kojiri, T., 2007. Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. Advances in Water Resources 30 (5), 1329–1341.

Deb, K., Saxena, D., July 16–21 2006a. Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In: Proceedings of 2006 IEEE Congress on Evolutionary Computation. Vancouver, CDN.

Deb, K., Sundar, J., Udaya Bhaskara Rao, N., Chaudhuri, S., 2006b. Reference point based multi–objective optimization using evolutionary algorithms. International Journal of Computational Intelligence Research 2 (3), 273–286.

Ernst, D., Geurts, P., Wehenkel, L., 2005. Tree-based batch mode reinforcement learning. Journal of Machine Learning Research 6 (1), 503–556.

Farina, M., Amato, P., June 2002. On the optimal solution definition for many-criteria optimization problems. In: Proceedings of the NAFIPS-FLINT International Conference '2002. IEEE Service Center, Piscataway, Ney Jersey, pp. 233–238.

Fleming, P., Purshouse, R., Lygoe, R., 2005. Many-Objective optimization: an engineering design perspective. In: Proceedings of the Third international conference on Evolutionary Multi-Criterion Optimization. Guanajuato, Mexico, pp. 14–32.

Fontane, D., Labadie, J., Loftis, B., 1981. Optimal Control of Reservoir Discharge Quality Through Selective Withdrawal. Water Resources Research 12 (6), 1594–1604.

Franssen, M., 2005. Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. Research in Engineering Design 16 (1), 42–56.

Galelli, S., Castelletti, A., 2013. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. Hydrology and Earth System Sciences 17, 2669–2684.

Galelli, S., Giuliani, M., Soncini-Sessa, R., 28 August - 2 September 2011. Dealing with many-objectives problems in water resources planning and management. In: Proceedings of the 18th IFAC World Congress. Milan, Italy.

Gass, S., Saaty, T., 1955. Parametric objective function - Part II. Operations Research 3, 316–319.

Gelda, R., Effler, S., 2007. Simulation of Operations and Water Quality Performance of Reservoir Multilevel Intake Configurations. Journal of Water Resources Planning and Management 133 (1), 78–86.

Gettys, C., Fisher, S., 1979. Hypothesis plausibility and and hypothesis generation. Organizational Behavior and Human Performance 24 (1), 93–110.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63 (1), 3–42.

Giuliani, M., Castelletti, A., Galelli, S., Soncini-Sessa, R., Weber, E., 2013. Many-objective operation of selective withdrawal reservoirs including water quality targets. In: Proceedings of the 2013 World Environmental and Water Resources Congress. American Society of Civil Engineers, pp. 1581–1590.

Giustolisi, O., Doglioni, A., Savic, D., Di Pierro, F., 2008. An evolutionary multiobjective strategy for the effective management of groundwater resources. Water Resources Research 44 (1).

Haimes, Y., Lasdon, L., Wismer, D., 1971. On a bicriterion formulation of the problems of integrated system identification and system optimization. IEEE Transactions on Systems, Man and Cybernetics 1, 296–297.

Hakanen, J., Sahlstedt, K., Miettinen, K., 2013. Wastewater treatment plant design and operation under multiple conflicting objective functions. Environmental Modelling & Software 46, 240–249.

Hashimoto, T., Stedinger, J., Loucks, D., 1982. Reliability, resilience, and vulnerability criteria for water resource system performance evaluation. Water Resources Research 18 (1), 14–20.

Hipsey, M., Romero, J., Antenucci, J., Hamilton, D., 2006. Computational

Aquatic Ecosystem Dynamics Model: CAEDYM v2.3 Science Manual. Centre for Water Research, University of Western Australia.

Hogarth, R., 1981. Beyond discrete biases: Functional and dysfunctional aspects of judgemental heuristics. Psychological Bulletin 90 (1), 197–217.

Imerito, A., 2007. Dynamic Reservoir Simulation Model: DYRESM Science Manual. Centre for Water Research, University of Western Australia.

Joliffe, I., 2002. Principal Component Analysis. Springer, New York, N.Y.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., Lempert, R. J., 2013. Many objective robust decision making for complex environmental systems undergoing change. Environmental Modelling & Software 42, 55 – 71.

Kasprzyk, J. R., Reed, P. M., Characklis, G. W., Kirsch, B. R., 2012. Many-objective de novo water supply portfolio planning under deep uncertainty. Environmental Modelling & Software 34, 87 – 104.

Kim, T., Heo, J., Bae, D., Kim, J., 2008. Single-reservoir operating rules for a year using multi objective genetic algorithms. Journal of Hydroinformatics 10 (2), 163–179.

Knowles, J., Corne, D., 2002. On metrics for comparing non–dominated sets. In: Proceedings of the 2002 World Congress on Computational Intelligence (WCCI). IEEE Computer Society, pp. 711–716.

Kollat, J., Reed, P., 2007. A framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VIDEO). Environmental Modelling & Software 22 (12), 1691–1704.

Kollat, J., Reed, P., Maxwell, R., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble kalman filtering, evolutionary optimization, and visual analytics. Water Resources Research 47, W02529, doi:10.1029/2010WR009194.

Korhonen, P., Wallenius, J., 1988. A pareto race. Naval Research Logistics (NRL) 35 (6), 615–623.

Koutsoyiannis, D., Economou, A., 2003. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. Water Resources Research 39 (6), doi: 10.1029/2003WR002148.

Larichev, O., 1992. Cognitive validity in design of decision–aiding techniques. Journal of Multi–Criteria Decision Analysis 1 (3), 127–128.

Lizotte, D., Bowling, M., Murphy, S., 2010. Efficient reinforcement learning with multiple reward functions for randomised controlled trial analysis. In: Proceedings of the 27th International Conference on Machine Learning. Omnipress, pp. 695–702.

López Jaimes, A., Coello, C., Uraas Barrientos, J., 2009. Online objective reduction to deal with many-objective problems. In: Ehrgott, M., Fonseca, C., Gandibleux, X., Hao, J.-K., Sevaux, M. (Eds.), Evolutionary Multi-Criterion Optimization. Vol. 5467 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 423–437.

López Jaimes, A., Coello Coello, C. A., Chakraborty, D., 2008. Objective reduction using a feature selection technique. In: Proceedings of the 10th annual conference on Genetic and evolutionary computation. GECCO '08. ACM, New York, NY, USA, pp. 673–680.

Lotov, A., Bushenkov, V., Kamenev, G., 2004. Interactive Decision Maps Approximation and Visualization of Pareto Frontier. Springer-Verlag, Heidelberg, D.

Perny, P., Weng, P., 2010. On Finding Compromise Solutions in Multiobjective Markov Decision Processes. In: Coelho, H. and Studer, R. and Wooldridge, M. (Ed.), ECAI 2010 - 19th European Conference on Artificial Intelligence. Vol. 215 of Frontiers in Artificial Intelligence and Applications. pp. 969–970.

Pianosi, F., Castelletti, A., Restelli, M., 2013. Tree-based fitted Q-iteration for multi-objective markov decision processes in water resources management. Journal of Hydroinformatics 15 (2), 258–270.

Powell, W., 2007. Approximate Dynamic Programming: solving the curse of dimensionality. Wiley-Interscience, New York.

Reed, P., Hadka, D., Herman, J., Kasprzyk, J., Kollat, J., 2013. Evolutionary multiobjective optimization in water resources: The past, present and future. Advances in Water Resources 51, 438–456.

Ross, S., 2013. A first course in probability (Fourth ed.). Macmillan College Publishing, New York.

Shabani, N., 2009. Incorporating flood control rule curves of the Columbia River hydroelectric system in a multireservoir reinforcement learning optimization model. Master's thesis, University of British Columbia, Vancouver, Canada.

Soncini-Sessa, R., Castelletti, A., Weber, E., 2007. Integrated and participatory water resources management: Practice. Elsevier, Amsterdam, NL.

Vamplew, P., Dazeley, R., Berry, A., Issabekov, R., Dekker, E., 2011. Empirical evaluation methods for multi objective reinforcement learning algorithms. Machine Learning 84, 51–80.

Vamplew, P., Yearwood, J., Dazeley, R., Berry, A., 2008. On the limitations of scalarization for multi-objective reinforcement learning of Pareto fronts. In: AI'08: The 21st Australasian Joint Conference on Artificial Intelligence. pp. 372–378.

White, D., 1982. Multi-objective infinite-horizon discounted markov decision processes. Journal of Mathematical Analysis and Optimization 89 (2), 639–647.

White, D., 1988. Further real applications of markov decision processes. Interfaces 18 (5), 55–61.

Wierzbicki, A., 1980. The use of reference objectives in multiobjective optimization. Springer.

Woodruff, M., Reed, P., Simpson, T., 2013. Many objective visual analytics: rethinking the design of complex engineered systems. Structural and Multidisciplinary Optimization 48 (1), 201–219.

Wu, W., Maier, H., Dandy, G., 2013. Multiobjective optimization of water distribution systems accounting for economic cost, hydraulic reliability, and greenhouse gas emissions. Water Resources Research 49 (3), 1211–1225.

Zagonari, F., Rossi, C., 2013. A heterogeneous multi-criteria multi-expert decision-support system for scoring combinations of flood mitigation and recovery options. Environmental Modelling & Software 49, 152–165.

Zass, R., Shashua, A., 2007. Nonnegative Sparse PCA. In: B. Schˆlkopf, J. Platt, and T. Hoffman (Ed.), Advances in Neural Information Processing Systems. Vol. 19. MIT Press, pp. 1561–1568.

Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R.C., Sadeghi, A., Thomson, A.M., Sahajpal, R., Srinivasan, R., Arnold, J.G., 2013. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. Environmental Modelling & Software 46, 208–218.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C., Grunert da Fonseca, V., 2003. Performance assessment of multiobjective optimizers: an analysis and review. IEEE Transactions on Evolutionary Computation 7 (2), 117–132.

Appendix A NPCA Setting

The NPCA approach requires setting two parameters, i.e. the number n of components and the value of α, the parameter balancing data reconstruction and orthonormality. As in the original PCA formulation, n is defined as the number of components allowing to explain a given threshold of the variance of the original variables (Joliffe, 2002). On the other hand, there are not similar criteria supporting the definition of α. According to Zass and Shashua (2007), α can be heuristically determined via trial-and-error, namely by selecting the value corresponding to the maximum explained variance. We tested different values of $\alpha \in [10^{-5}, 10^{10}]$ (for $n = 5$), with values of explained variance varying between 56% and 77%, with the maximum obtained for $\alpha = 1000$, which is the value adopted in this work.

Table 1: Correlation matrix for the ten objectives.

	In-reservoir			Downstream						
	J^{Lev}	J^{Algae}	J^{Sed}	J^{Irr1}	J^{Irr2}	J^{Irr3}	J^{Irr4}	J^{Irr5}	J^{Temp1}	J^{Temp2}
J^{Lev}	-	-0.67	0.11	-0.16	-0.18	0.03	-0.13	-0.13	-0.50	-0.58
J^{Algae}	-0.67	-	-0.12	0.36	0.31	-0.02	0.18	0.22	0.53	0.56
J^{Sed}	0.11	-0.12	-	-0.22	-0.23	-0.10	-0.04	-0.15	-0.13	-0.08
J^{Irr1}	-0.16	0.36	-0.22	-	0.88	0.13	0.51	0.62	0.38	0.23
J^{Irr2}	-0.18	0.31	-0.23	0.88	-	0.37	0.31	0.61	0.30	0.14
J^{Irr3}	0.03	-0.02	-0.10	0.13	0.37	-	0.11	-0.11	0.03	-0.03
J^{Irr4}	-0.13	0.18	-0.04	0.51	0.31	0.11	-	-0.13	0.20	0.09
J^{Irr5}	-0.13	0.22	-0.15	0.62	0.61	-0.11	-0.13	-	0.31	0.27
J^{Temp1}	-0.50	0.52	-0.13	0.38	0.30	0.03	0.20	0.31	-	0.88
J^{Temp2}	-0.58	0.56	-0.08	0.23	0.14	-0.03	0.09	0.27	0.88	-

Table 2: Values of the coefficients characterising the five principal vectors.

Objective	$\mathbf{p}^1$	$\mathbf{p}^2$	$\mathbf{p}^3$	$\mathbf{p}^4$	$\mathbf{p}^5$
J^{Lev}	0	0	0	0.0103	0.3789
J^{Algae}	0.0663	0.4822	0.0132	0	0
J^{Sed}	0	0	0	0	0.9254
J^{Irr1}	0.5573	0.0260	0.1275	0	0
J^{Irr2}	0.5986	0	0	0.0832	0
J^{Irr3}	0	0	0.0003	0.9964	0
J^{Irr4}	0	0.0043	0.9915	0.0124	0
J^{Irr5}	0.5702	0	0	0	0
J^{Temp1}	0.0405	0.6107	0.0234	0.0040	0
J^{Temp2}	0	0.6276	0	0	0

Figure 1: Tono Dam location in Western Japan (panel a), the main characteristics of the reservoir with the decision variables adopted in this study (panel b). Symbols are defined in Section 3.2.

Figure 2: The hierarchy of sectors and objectives of Tono dam management problem. The grey-shaded objectives are accounted for in the 5-objective formulation presented in Castelletti et al. (2013b).

Figure 3: Schematization of the optimization and simulation procedure. The black line is the optimization workflow, the dashed line is the evaluation via simulation of the optimal operating policies.

Figure 4: Explained variance as a function of the number of principal components extracted via NPCA (red bars) and PCA (blue bars).

Figure 5: Graphical comparison between the approximated Pareto fronts obtained by employing five components (red cones) or five selected objectives (grey cones). J^{Algae}, J^{Temp1} and J^{Lev} (in logarithmic scale) are plotted on the primary axes, with the black arrows indicating the directions of increasing preference. The orientation of the cones accounts for J^{Sed}, with the best solutions represented by upward cones. The dimension of the cones is proportional to J^{Irr1}, with the best solutions identified by small cones.

Figure 6: Graphical comparison between the approximated Pareto fronts obtained by employing five components (red points) or five selected objectives (grey points) projected in the plane J^{Algae}, J^{Lev} (panel (a)) and J^{Algae}, J^{Temp1} (panel (b)).

Figure 7: Graphical comparison between the approximated Pareto fronts obtained by employing five components (red cones) or five selected objectives (grey cones). J^{Irr3}, J^{Irr4} and J^{Temp2} are plotted on the primary axes, with the black arrows identifying the directions of increasing preference. The orientation of the cones represents J^{Irr5}, with the best solutions represented by downward cones. The dimension of the cones is proportional to J^{Irr2}, with the best solutions identified by small cones.

Figure 8: Graphical comparison between the approximated Pareto fronts obtained by employing five components (red points) or five selected objectives (grey points) projected in the plane J^{Irr3}, J^{Irr4} (panel (a)) and J^{Irr3}, J^{Temp2} (panel (b)).

Figure 9: Graphical comparison between the approximated Pareto fronts obtained with the NPCA-based and the selection-based approaches. For illustration purposes the objectives are standardized (zero mean and unit standard deviation) and each axis is oriented so that the direction of preference is always downward. The five objectives in bold are accounted for in the 5-objective formulation presented in Castelletti et al. (2013b).

Figure 10: Comparison of the average daily value of the immediate costs obtained with the selection-based (grey line) and NPCA-based (red line) compromise alternatives.

Figure 11: Performance of the selection-based (grey bars) and NPCA-based (red bars) approaches in terms of generational distances, additive ε-indicator and hypervolume indicator.