Policy Representation Learning for multiobjective reservoir policy design with different objective dynamics

Marta Zaniolo[1], **Matteo Giuliani**[1], **and Andrea Castelletti**[1]

[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

Key Points:

- We introduce a novel method to define an optimal input set for a multipurpose dam operating policy that varies with the objective trade-off.
- Better informed policies are able to mitigate conflicts between water users and achieve system-wide benefits.
- The addition of information in policy design increases the policies robustness towards extreme hydrological conditions.

Corresponding author: Marta Zaniolo, marta.zaniolo@polimi.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2020WR029329.

Abstract

Most water reservoir operators make use of forecasts to inform their decisions and enhance water systems flexibility and resilience by anticipating hydrological extremes. Yet, despite numerous candidate hydro-meteorological variables and forecast horizons may potentially be beneficial to operations, the best information set for a given problem is often not evident. Additionally, in multi-purpose systems characterized by multiple demands with varying vulnerabilities and temporal scales, this information set might change according to the objective tradeoff. In this work, we contribute a novel method to learn the optimal policy representation (i.e., policy input set) by combining a feature selection routine with a multi-objective Direct Policy Search framework in order to retrieve the best policy input set online (i.e., while learning the policy) and dynamically with the objective trade-off. The selected policy search routine is the Neuro-Evolutionary Multi-Objective Direct Policy Search (NEMODPS) which generates flexible policy shapes adaptive to online changes in the input set. This approach is demonstrated on the case study of Lake Como (Italy), where the operating objectives are highly heterogeneous in their dynamics (fast and slow) and vulnerabilities (wet and dry extremes). We show how varying objectives, and tradeoffs therein, benefit from a different policy representation, ultimately yielding remarkable results in terms of conflict mitigation between different users. More informed policies, moreover, show higher robustness when re-evaluated across a suite of different hydrological conditions.

1 Introduction

Water reservoirs have long been fundamental components of coupled human-water systems worldwide, providing communities with green and affordable electricity, water supply for agricultural and urban consumption, and flood protection. Yet, lately, new concerns are arising regarding the reliability of water systems as climate change increases the likelihood of extreme events, and economic development exacerbates water demands and conflicts (*Fletcher et al.*, 2019; *Herman et al.*, 2020). One way of increasing resilience and reliability of water systems is to build more, larger, infrastructures, however, this hard path to capacity expansion is costly and often yields unintended cross-sectoral externalities (*Gleick*, 2003). An alternative, soft-path towards resilience advocates the improvement of the operating rules used to control the existing water infrastructures to enhance their capability to anticipate weather extremes, and timely prepare for them.

Traditionally, the operating policy of water reservoirs was conditioned upon very limited information systems comprising reservoir storage and a cyclostationary time index (*Hejazi et al.*, 2008). More recently, *Turner et al.* (2019) showed that most water system operators across the US make use of streamflow forecasts to further improve operations. The employed forecast horizon is however reservoir-specific, and, when official guidelines are absent, operators seem to rely on their expert judgment to identify their forecast horizon of choice. In the water resources literature, few studies have tackled the issue of the optimal selection of streamflow forecast horizon for a single-objective reservoir operated for water supply (*Anghileri et al.*, 2016), hydropower (*Hamlet et al.*, 2002; *Block*, 2011; *Xu et al.*, 2014), or for a generic concave objective function (*Zhao et al.*, 2014, *Zhao et al.*, 2019). Additionally, the breath of information sources that was demonstrated to be valuable to inform reservoir operations is by no means limited to streamflow forecasts, but includes the previous period's inflow (*Gal*, 1979; *Maidment and Chow*, 1981), available hydrological observations (*Denaro et al.*, 2017), traditional (*Hejazi and Cai*, 2011) or basin-specific (*Zaniolo et al.* 2018, *Zaniolo et al.*, 2019) drought indexes, measures of snow abundance (*Desreumaux et al.*, 2014; *Giuliani et al.*, 2016a), shifts in hydrological regimes (*Turner and Galelli*, 2016), teleconnection indices (*Libisch-Lehner et al.*, 2019), and sea surface temperature measured in appropriate locations (*Giuliani et al.*, 2019; *Zaniolo et al.*, 2021a).

While these studies are a great demonstration of the potential of using unconventional policy representations in policy design, none of them attempts at automatizing representation learning in a portable framework. Additionally, no attention has been given to a major challenge to learning an optimal policy representation, i.e., the coexistence of multiple operating objectives. In fact, previous studies either consider systems operated for a single purpose (i.e., reservoirs operated just for hydropower), or specify only one policy representation for the entire tradeoff space. In multi-purpose water reservoir systems, however, common operating targets, e.g., flood protection and water supply, can be vastly heterogeneous in their dynamics and vulnerabilities. Flood events are generally caused by the onset of fast and intense wet meteorological extreme events, while water supply failures are the result of a prolonged period of water shortage caused by slow-developing dry hydrological extremes, i.e., droughts. In these systems, defining an appropriate policy representation becomes more intricate. On the one hand, a flood-conservative policy benefits from a short lead time look-ahead information that conveys peak inflow magnitude and timing, on the other, a water supply-prone policy seeks predictors that are relevant for the onset of a prolonged water shortage to timely activate hedging strategies. The tradeoff space between these two opposite solutions is populated by an ensemble of policies diversely balancing opposite control targets. Such behavior is shown for a fixed policy representation via sensitivity analysis to policy inputs for alternative tradeoffs (*Quinn et al.*, 2017; *Doering et al.*, 2021).

In this work, for the first time, we hypothesize and quantitatively demonstrate that in Multi-Objective (MO) problems different objective tradeoffs require different information, and selection of policy representation should be tradeoff-specific. Our results demonstrate that one policy input set is inadequate to represent the entire space of different control behaviors that may emerge for alternative tradeoffs.

Part of the reason why a tradeoff dynamic selection was never performed is that traditional policy search routines only support static and prespecified input sets, thereby not allowing the evolution of a population with heterogeneous input sets. In this work, we propose a novel technique to automatically learn a Pareto front of optimal policies and their representations for a multipurpose water system. The method is applicable to large and heterogeneous datasets of candidate policy inputs, from meteorological and hydrological forecasts with disparate horizons, to observational data. The framework, namely SINEPS, Selection of Information for NeuroEvolutionary Policy Search, combines automatic feature selection with NEMODPS (NeuroEvolutionary Multi-Objective Direct Policy Search, *Zaniolo et al.* (2021b)), a policy search routine that can accomodate changes in the policy input set. SINEPS starts with a simple operating policy and a minimal policy representation and gradually includes new inputs to the policy representations while automatically adjusting the policy processing capacity. For every Pareto efficient policy, the selected input is the one that explains most of the information gap between the policy itself, and an ideal, deterministic, Perfect Operating Policy, designed under the assumption of perfect knowledge of future disturbance.

This framework is tested on the real-world case study of the multi-purpose Lake Como, operated to meet two conflicting and heterogeneous objectives of flood protection and water supply, mainly for irrigation. The flood objective is characterized by fast dynamics and vulnerability towards wet extremes, while irrigation supply is characterized by a slow dynamic and vulnerability towards dry extremes. In this paper, the dataset of candidate policy inputs is composed of perfect streamflow forecasts at different lead times.

1.1 Literature review on policy representation learning

The problem of learning a policy representation is not unique to water resources management, on the contrary, it is widely addressed in the control community, finding

applications in diverse fields, from spatial path scheduling (*Whiteson et al.*, 2005), stock index trading (*Si et al.*, 2017), to virtually any autonomous robot control task (e.g., *Hachiya and Sugiyama*, 2010; *Lesort et al.*, 2018). In this section, we propose a literature review on policy representation learning that goes beyond the existing experience in dam policy design in order to present and discuss the wider background and challenges the inspired the design of SINEPS, and motivate its algorithmic choices.

When designing an operating policy for a given system, defining the policy representation corresponds to selecting its input set. Such problem is generally tackled by pairing Feature Extraction with Policy Search (*Liu et al.*, 2015; *Lesort et al.*, 2018). Feature Extraction refer to a family of techniques that transform an original dataset into a more compact, while still highly informative dataset (*Cunningham*, 2008). Policy Search methods aim at learning an optimal operating policy for a system (e.g., a release policy from a reservoir) with respect to its objective functions (e.g., flood and water supply). In the proposed taxonomy, we identify *a priori*, *a posteriori*, and *online* approaches to pairing feature extraction and policy search for learning a policy and its representation.

In the first *a priori* approach, the feature extraction step is antecedent and independent from the policy search step. First, the feature extraction routine reduces the dimensionality of the dataset of candidate features for example extracting few relevant features from the dataset, removing irrelevant ones, or generating new features by appropriately combining existing ones. The reduced dataset represents the selected policy representation, and is used for policy search. The dimension reduction is generally achieved via i) data compression techniques, e.g., autoencoders (e.g., *Morimoto et al.*, 2008), or Principal Component Analysis (*Nouri and Littman*, 2010), that map the initial dataset into a lower dimensional latent space that retains most of its information content, ii) using a target control sequence to identify relevant policy drivers (*Kroon and Whiteson*, 2009; *Giuliani et al.*, 2015; *Denaro et al.*, 2017), or, iii) via expert-based feature selection (e.g., *Akrour et al.*, 2012) or extraction (e.g., *Sturtevant and White*, 2006; *Giuliani and Castelletti*, 2019) to design a problem-specific representation. In general, a priori approach to policy representation is advisable whenever there is sufficient knowledge of the task to confidently devise an appropriate feature set. This very low computationally demanding approach, in fact, does not offer any guarantees on the optimality of the chosen representation (*Lesort et al.*, 2018).

The *a posteriori* approach evaluates the suitability of a policy representation by assessing the performance of the policy conditioned upon it. Multiple policies are designed with alternative input sets, and the desired representation is identified as the one generating the best performing policy. In principle, the entire combinatorial space of features subsets could be exhaustively explored, yielding to an optimal solution albeit resulting computationally intractable for non-trivial datasets (see, e.g., *Gaudel and Sebag*, 2010). Alternatively, for modest datasets, hill-climbing approaches incrementally add features to the representation retaining the most successful ones (*Wright et al.*, 2012; *Zhang*, 2009; *Tan et al.*, 2013). Finally, an initial *a priori* reduction can be applied to select a limited number of candidate representations that are then exhaustively compared *a posteriori* (*Giuliani et al.*, 2016a; *Castelletti et al.*, 2016). In general, *a posteriori* feature representation is significantly more computationally burdensome than the a priori counterpart. Yet, an exhaustive *a posteriori* search can be performed with virtually no preexisting knowledge of the task, and guarantees the optimality of the derived feature representation. Both *a priori*, and *a posteriori* approaches in general rely on heavy expert-based manual engineering in defining potentially appropriate policies representations to implement or test (*Bengio et al.*, 2013).

The third, *online* approach, interleaves feature extraction phases throughout the policy search process, using progressively refined feature representations to support policy learning. Representations are updated during the search via supervised learning, by extracting features that approximate the state space (*Curran et al.*, 2016; *Alvernaz and*

Togelius, 2017), state-transition space (*Assael et al.*, 2015; *Van Hoof et al.*, 2016), or the reward trajectory (*Munk et al.*, 2016; *Oh et al.*, 2017) of the policy learned thus far (for a comprehensive review, see *Lesort et al.*, 2018). The adjusted representation is then employed to refine policy search in a feedback loop between the two routines. Computationally, *online* approaches are more expensive than *a priori*, but less than *a posteriori* methods, while handling significantly larger datasets of candidate information.

1.2 SINEPS

In this work, we present a novel method for *online* dynamic policy representation called SINEPS, Selection of Information for NeuroEvolutionary Policy Search. It requires the selection of i) a feature extraction method, ii) a policy search routine, and iii) a strategy to interface the two.

1. **Feature extraction method**: Several *online* policy representation routines employ Feature extraction techniques that reduce the dimensionality of the representation by projecting the initial feature space into a lower dimensional latent space that preserves information content. However, such an approach does not guarantee that any candidate feature is actually excluded from the problem formulation (*Loscalzo et al.*, 2015) . As a result, while the operating policy can actually benefit from a lower-dimensional representation, the actual problem size remains unchanged. In an operational setting, this implies that the entire dataset of initial features must be retrieved continuously. Alternatively, Feature Selection methods are a subset of the feature extraction techniques that reduces the dataset size by identifying a subset of the initial features. Some authors suggest the use of feature selection routines, rather than information encoders, for representation learning, in order to effectively restrict the number of candidate variables included in the problem formulation (e.g., *Loscalzo et al.*, 2015). The representation obtained through variable selection, moreover, highlights relevant policy drivers, is easily interpretable, and can thus generate insights on the task at hand. Within Feature Selection techniques, the iterative online framework can accommodate simple correlation-based variable filtering (i.e., the variables that are most correlated with the target are selected), as well as non-linear model-based selection routines (e.g., IIS, *Castelletti et al.*, 2010). Here, we use a correlation-based filtering approach, where the correlation is measured in Symmetric Uncertainty (SU, *Blum and Langley*, 1997). SU is a normalized version of the Mutual Information metric (MI, *Shannon*, 1948) that quantifies the degree of similarity between two variables, or, more specifically, the amount of information that can be obtained on one variable by observing the other. Entropy-based techniques like SU are model-free and generalizable to any modeling context, as they do not require to assume any functional relationship between the variables (*MacKay*, 2003), contrary to simpler metrics such as correlation coefficients that assume a linear dependence. The use of SU is supported in the information theoretic literature and was demonstrated to outperform several other feature selection methods on a suite of 15 benchmark feature selection problems (*Zhang and Chen*, 2021).
   Note that SU is employed as a screening tool that allows to detect promising policy representations by identifying candidate variables with high information content across different objectives. This is intended to avoid an exhaustive approach that would test every possible candidate representation in policy search, which would be computationally untractable. The policy search step, described below, evolves policies with different representations to generate a Pareto front of optimal policies and applies further selection pressure onto alternative policy representations thereby further refining the representation selection in a policy search context.
2. **Policy Search Method**: Direct Policy Search (DPS) is emerging as one of the most effective, and widely applied methods to design optimal operating policies

for multi-purpose reservoir operations, given its multi-objective nature, flexibility in problem and objective formulation, and data-driven nature that allows to use trajectories of non-modeled information in policy design (*Giuliani et al.*, 2016b). DPS defines the operating policy within a prespecified class of functions and solves a problem of optimal functional parameterization with respect to the problem's objectives (*Zatarain et al.*, 2017; *Quinn et al.*, 2018; *Giuliani et al.*, 2019; *Quinn et al.*, 2019). Flexible universal approximators such as Neural Networks (NNs) are generally employed to parameterize the operating policy in order not to restrict the parametrical search to a small functional subspace that may not contain skillful solutions (*Giuliani et al.*, 2014, *Giuliani et al.*, 2018). The architecture of a NN employed for policy design includes as many input nodes as the number of features in the policy representation, and as many output nodes as the decisions to be taken on the system, e.g., reservoir release decisions. Finally, the internal NN complexity, i.e., number of hidden nodes, connections, and layers, is crucial to determine the network processing capability and training requirements. The a priori definition of the optimal network complexity for a given problem would require a perfect knowledge of the operational task, which is in general unavailable. Therefore, in practical application, the network architecture is selected by the modeler via few manual trials and errors balancing the network approximation capacity, training costs, and overfitting tendency. Given its rigid, prespecified, policy structure, DPS techniques do not support dynamic changes in the dimensionality of the policy feature representation.

A promising alternative that obviates to policy rigidity is represented by NeuroEvolution (NE), a set of techniques that employs evolutionary algorithms to evolve neural networks in terms of their architectures and parameters. These techniques generally begin with a population of simple networks and progressively build more sophisticated ones by applying new architectural elements (nodes and connections). The evolutionary competition ultimately determines the optimal network complexity. By pairing NE with DPS, it is possible to derive policy search routines that support online changes in policy architecture. Popular NE algorithms (e.g., NEAT *Stanley and Miikkulainen*, 2002) are, however, strictly applicable to single-objectives problems. The here employed NeuroEvolutionary Multi-Objective Direct Policy Search (NEMODPS), is the first NE routine specifically designed to solve MO problems in one algorithmic iteration (*Zaniolo et al.*, 2021b). NEMODPS is here employed for the first time to jointly evolve policies with different feature representations. In general, not all the policy representations identified in the feature selection step will survive the evolution pressure, thus refining the selection of optimal representations via policy competition. NEMODPS will be briefly introduced in Section 2.2 of the Methods. The reader is referred to *Zaniolo et al.* (2021b) for a more detailed analysis of NEMODPS, and its benchmarking against traditional DPS in terms of performance and computational costs.

3. **Interfacing strategy**: in many applications, the selection of relevant features is performed via supervised learning using as target the state, state-transition, state-value spaces, or the cost trajectory produced by the policy learned thus far (for a review, see *Lesort et al.*, 2018). Cost-based selection is generally recognized as more effective in identifying task-oriented policy representations (*Loscalzo et al.*, 2015), however, in multi-objective problems, the coexistence of multiple cost signals complicates the cost-based selection process. In SINEPS, we propose a novel interfacing strategy that is both task-tailored, and suitable for MO problems. In particular, we use as reference a deterministic Perfect Operating Policy (POP) that assumes full knowledge of future system disturbance. For a given state, we contrast the actions extracted from the POP to those extracted from the policy under design. We assume that the difference in actions is due to the information gap in the policies representations, and thus surrogates the information that the designed policy would require to meet the POP performance. The trajectory of ac-

tion residuals is used as an interfacing strategy, and employed as target for feature selection. Such target can be considered task-relevant, as it is a proxy of the policy information deficiency for a given task. Additionally, it can be applied to MO problems by contrasting each Pareto efficient policy with the corresponding perfect counterpart supporting a tradeoff dynamic feature selection.

To summarize, SINEPS combines feature selection, neuroevolution and an original interfacing strategy. The choices made in the selection and development of the building tools of SINEPS target the overarching goal of designing the first multi-objective feature representation learning routine that automatically specifies an optimal policy representations for each tradeoff.

This paper is organized as follows. The next section presents the methods of this work, by presenting the methodological Framework 2.1, and expanding on the key concepts and tools employed in the methodology, including NEMODPS 2.2. Section 3 is dedicated to the presentation of the case study and experimental settings. Results are discussed in Section 4, and in the following Section 5 we draw conclusions and introduce some discussion points.

2 Methods

In this work, we consider a water reservoir system modeled as a discrete-time, periodic, non-linear, stochastic process defined by a state variable s_t (reservoir storage), a control variable u_t representing the release decision from the dam gates, stochastic disturbances $\boldsymbol{\varepsilon}_{t+1}$ (net reservoir inflow), and a state-transition function $f(\cdot)$: $s_{t+1} = f(s_t, r_{t+1}, \boldsymbol{\varepsilon}_{t+1})$ where the effective release r_{t+1} coincides with the release decision u_t corrected, where appropriate, with a non-linear release function $R_t(s_t, u_t, \boldsymbol{\varepsilon}_{t+1})$ determining the minimum and maximum releases feasible for the time interval $[t, t+1)$ to respect physical and legal constraints. The operating policy π determines the release decision from the water reservoir $u_t = \pi(\cdot)$ at each time step t over the simulation horizon H. The objective of this work is to design the optimal operating policy and relative representation for this system by solving a minimization problem formulated as follows:

$$\min_{\pi, \mathbf{I}_t, \zeta(\theta)} \mathbf{J}(\pi, s_0, \varepsilon_1^H) \tag{1}$$

where we search the minimum of the multidimensional objective function $\mathbf{J}$, here interpreted as cost, with respect to the closed loop operating policy π, its representation $\mathbf{I}_t$, functional class ζ and relative parameterization θ. In particular, the operating policy π is conditioned upon basic information (i.e., the reservoir storage s_t a time index d_t), and an additional vector of information $\mathbf{I}_t$ searched within the dataset of candidate information as in $\pi = \pi(s_t, d_t, \mathbf{I_t})$. Among the available policy search methods, parametric approaches define π within a class of functions ζ, and search its optimal parameterization θ. The employed NEMODPS technique supports the conjunct search of the optimal functional class ζ and relative parameters as in $\zeta(\theta)$.

In general, in MO problems, conflicts occur between different operating objectives, and the solution is constituted by a set of non-dominated (or Pareto optimal) solutions $\mathcal{P}^* = \{\pi^* | \nexists \pi \prec \pi^*\}$, which maps onto the Pareto front $\mathcal{F}^* = \{\mathbf{J}(\pi^*, \mathbf{x}_0, \boldsymbol{\varepsilon}_1^H) | \pi^* \in \mathcal{P}^*\}$. For a more complete problem formulation please refer to the *Detailed Problem Formulation* section of the Supplementary Information.

2.1 Framework

In this section, we present the flowchart of the proposed SINEPS framework employed to approach Problem 1, reported in Figure 1 and organized in numbered blocks.

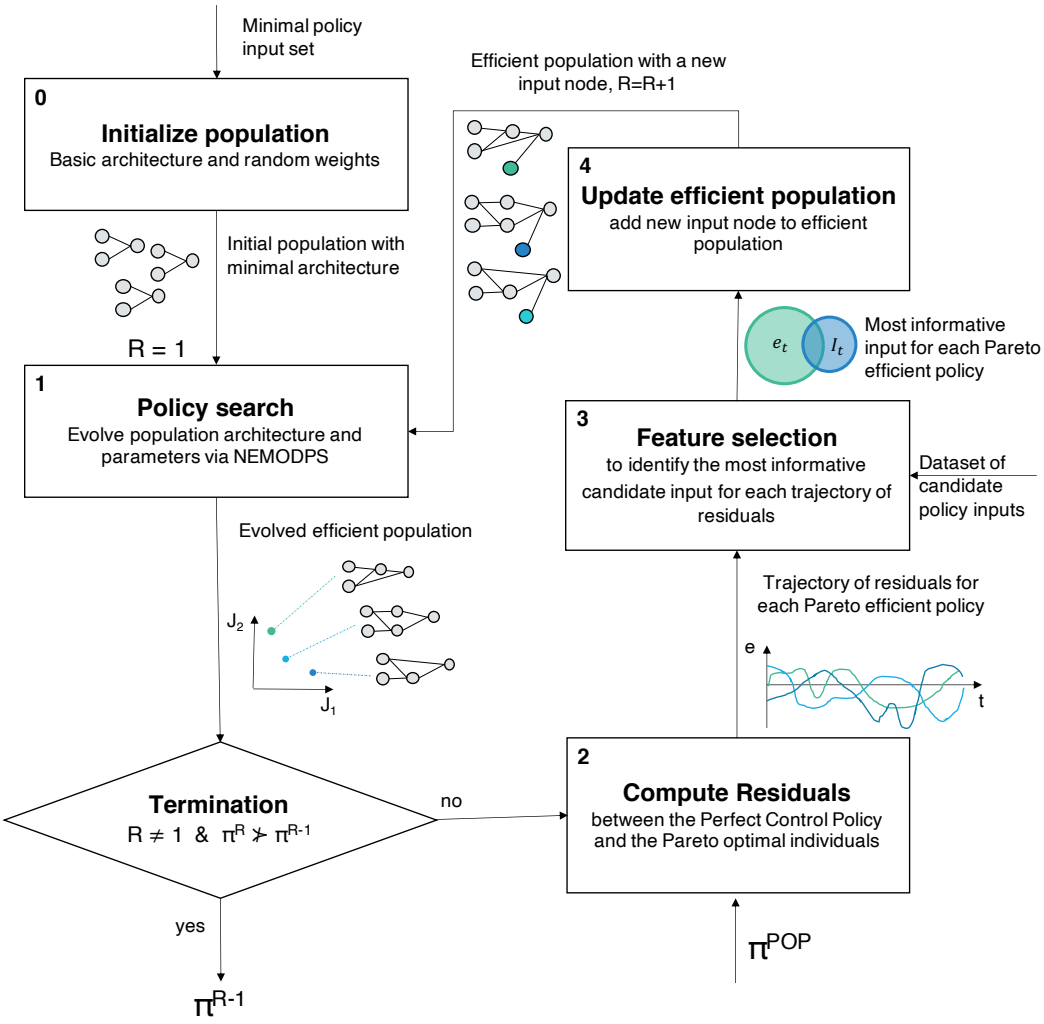

Figure 1. SINEPS flowchart. By looping through the building blocks of this flowchart, the procedure complexifies the initial population in terms of feature representation and policy architecture.

0: The procedure begins in round R1, with the initialization of a population of simple neural networks, a minimal architecture, and random weights. At this stage, the policy representation is also minimal, comprising a cyclostationary time index d_t and the reservoir storage s_t, namely, $\pi^{R1} = \pi^{R1}(d_t, s_t)$.

1: This population is the input to the Policy Search building block that employs NEMODPS. For a given input set, NEMODPS evolves policies' architecture and parameters in a MO problem (more details in the dedicated Subsection 2.2). The output of this step is an ensemble of Pareto efficient operating policies, each specified with a tailored architecture, resulting in an architecturally heterogeneous population.

2: In the first round, the flowchart proceeds to the building block named Compute Residuals. In this step, we contrast the operating decisions produced by each Pareto efficient policy with the decisions given by a Perfect Operating Policy (POP) extracting the trajectories of decision residuals e_t, i.e., the difference in the decisions selected by the minimally informed policy under design π^{R1}, and the perfectly informed policy π^{POP}. The calculated residuals are assumed to be due to their information gap (more details in the dedicated Section 2.3).

3: In the Feature Selection step, we search the dataset of candidate policy inputs D to identify the most informative feature for π^{R1}. For this purpose, we compute the SU metric between each vector of residual trajectory in e_t, i.e., the policys information gap, and the candidate policy input dataset D. SU quantifies the amount of information shared between e_t and each candidate input, allowing to identify the most promising feature representation by selecting the feature that explains most of the policy information gap. SU is defined in [0,1] and can be computed for two variables X and Y as:

$$SU(X,Y) = 2 * \frac{MI(X,Y)}{H(X) + H(Y)} = 2 * \frac{H(X) + H(Y) - H(X,Y)}{H(X) + H(Y)} \tag{2}$$

Where H(X) and H(Y) are the entropy of the variable X and Y, and H(X,Y) is their joint entropy.

Because the trajectory of residuals is computed independently for each efficient policy, the inputs selected are policy-specific, and may vary across the tradeoff space.

4: Each efficient policy is then updated by including the selected feature in the input set, with a single input-output connection and a randomly initialized weight. The population of policies is now heterogeneous in its feature representation. Such population will now enter round R2 of SINEPS, with an update representation that includes the tailored information I_t, $\pi^{R2} = \pi^{R2}(d_t, s_t, I_t)$. In step 1 of the second round R2, this population is further evolved via NEMODPS. Individuals will appropriately complexify their architecture by genetic evolution to adapt to the newly inserted input, and learn how to make use of its information content. Neuro-evolutionary competition will further filter feature representation, causing only the fittest representations to survive in the efficient policies of round R2. Note that this framework performs a joint optimization of the policy inputs and architecture which cannot be decoupled. In particular, the input layer contains the information that a policy can access, while the policy structure governs how this information is used and translated into a control decision. Therefore, on the one hand, a policy with an inadequate input layer wont be able to make good control decisions because poorly informed, no matter how well the policy structure can translate input into decisions. Similarly, when a new input is added to an existing policy, the structural optimization is necessary for the policy to learn how to use the input, i.e., to build the structural elements (connection, nodes), that will enable it to appropriately use it to make more informed control decisions. Without a structural optimization, the new input would be unused.

SINEPS proceeds analogously until the Termination check is positive, namely when the efficient Pareto set at Round R does not significantly dominate the Pareto set in the previous round: $\pi^{*R} \not\prec \pi^{*R-1}$. More details on the termination criterion are presented in Section 2.4. Upon termination, we retain as efficient solutions the Pareto set generated at the previous round $R-1$, as it achieves virtually the same performance as round R with a simpler representation.

2.2 NEMODPS

In this section, we give an overview of the main components of NEMODPS, the policy search routine employed in this study. NEMODPS builds on a recent Reinforcement Learning branch called Neuro-Evolution (NE) (*Stanley and Miikkulainen*, 2003; *Floreano et al.*, 2008), which employs Evolutionary Algorithms to optimize neural network architectures and parameters. NEMODPS algorithm is inspired by NEAT (*Stanley and Miikkulainen*, 2002), and the subsequent literature of NEAT improvements targeting complex control problems, vast decision spaces, and noisy environments. Additionally, NEMODPS contains original strategies to address the specific complexities of multi-objective optimization problems, which make NEMODPS the first multi-objective NE algorithm. An in-depth explanation of NEMODPS can be found in Zaniolo, 2021, but here we discuss the main algorthmic components.

Key elements of NEMODPS are (1) a process of evolutionary complexification, (2) the use of parametrical and topological operators, and (3) an architecture-based competition scheme that sustains solution diversity and avoids premature convergence.

1. **Evolutionary complexification:** NEMODPS begins with a population of uniform simple networks, i.e., neural networks composed of just input and output layers, fully connected, with randomly initialized connection weights. As the evolution proceeds, neural architectures gradually complexify by including more architectural elements (nodes and connections) in the network's hidden layer, which connects inputs to outputs. These elements are randomly generated by topological evolutionary operators and selected by evolutionary pressure.
2. **Parametrical and topological operators:** EAs use evolutionary operators such as mutation and crossover to recombine existing individual parameters to generate new individuals. NE evolves individual architectures along with their parameters, and therefore it includes both parametrical, and topological mutation and crossover. In particular, the topological mutation operator performs a randomized addition of a node (sigmoidal or Gaussian) or a connection to an individual. Topological crossover assigns the offspring a mix of the parents' architectures. NEMODPS coordinates the topological and parametrical search in a dual timescale: parametrical mutation and crossover takes place every generation, while topological variations happen on a slower timescale, every few generations, to allow the competition scheme to protect solution diversity.
3. **Competition scheme:** at every generation, the population is divided into species of individuals with similar topologies. Species compete among each other for their ability to reproduce, so that a larger offspring is assigned to well performing ones. A fitness sharing mechanism penalizes numerous species preventing them from taking over the entire population causing loss of topological diversity and premature convergence. NEMODPS generalizes the fitness sharing strategy for MO problems, rewarding species with Pareto efficient individuals, and penalizing species whose individuals are located in crowded region of the objectives space in order to encourage the exploration of the entire tradeoff space.

2.3 Extraction of optimal decision from a Perfect Operating Policy

Following *Giuliani et al.* (2015), the Perfect Operating Policy π^{POP} is designed by solving Problem 1 under the hypothesis of deterministic knowledge of the trajectory ε_1^H of external drivers over the entire evaluation horizon H at any given time step, $\pi^{POP} = \pi^{POP}(s_t, t, \varepsilon_1^H)$ and can be solved via various open loop deterministic control methods (examples can be found in, e.g., *Dobson et al.*, 2019; *Macian-Sorribes and Pulido-Velazquez*, 2020). Here, we solve the problem with Deterministic Dynamic Programming (DDP). Such a deterministic policy can be considered the optimal reference for improving a basic policy design, but cannot be realistically implemented in a real-world system (e.g., *Denaro et al.*, 2017). In order to obtain the trajectory of decision residuals e_t, we compare the decisions extracted from the π^{POP} with those extracted from the efficient policy π^R at a given round R, referring to the same state trajectory produced by the simulation of π^R. The difference in decisions extracted by the policy under design π^R, and the perfectly informed policy π^{POP}, is assumed to be due to their information gap. In a MO problem, π^R and π^{POP} are constituted by a set of Pareto efficient policies, therefore, each π^R policy is associated with the POP solution that displays the most similar tradeoff.

2.4 Termination criterion

SINEPS terminates at round $R > 1$ when the efficient Pareto set at Round R does not significantly dominate the Pareto set in the previous round: $\pi^R \not\succ \pi^{R-1}$, according to an appropriate metric. Several metrics could in principle be used to express dominance in a Pareto sense. Here, as suggested in *Giuliani et al.* (2015), we use the hypervolume indicator (HV), which captures both the convergence of the Pareto front under examination $\mathcal{F}$ to the optimal one $\mathcal{F}^*$, as well as the representation of the full extent of tradeoffs in the objective space. The hypervolume metric allows set-to-set evaluations, measuring the volume of objective space Y dominated ($\preceq$) by the considered approximate set. HV assumes values between 0 to 1, where Pareto fronts with higher HV are considered better. For this study, we consider the search terminated when the HV increase from round R-1 to round R is lower than 5%. Policies in round R are characterized by an increased complexity in the input layer, that however doesn't yield a significant performance increase. Therefore, round R is discarded, and the policies produced at round R-1 are considered final.

3 Case Study and Data

We consider the case study of Lake Como, a multipurpose regulated lake located in the southern Alpine belt, Italy (Fig. 2). The main tributary, and only emissary of the lake is the Adda river, whose sublacual reach originates in the southeastern branch of Lake Como, crosses the Po valley, and eventually serves as a tributary to the Po river downstream. In its course, part of its waters are withdrawn to irrigate four agricultural districts. The southwestern branch of Lake Como constitutes a dead end, and exposes the city of Como to flooding events. The Lake Como basin hydrological regime is snow-rainfall dominated, characterized by scarce winter and summer inflows, a large snowmelt peak in late spring, and a secondary rainfall peak in autumn.

The lake regulation has two conflicting aims of supplying water to downstream users by storing spring snowmelt peak, and minimizing flood risk on the lake shores by maintaining the lake level as low as possible, therefore, $\mathbf{J}$ in eq. 1 is a bidimensional vector. On the basis of previous works (e.g., *Castelletti et al.*, 2010), these two objectives are defined as:

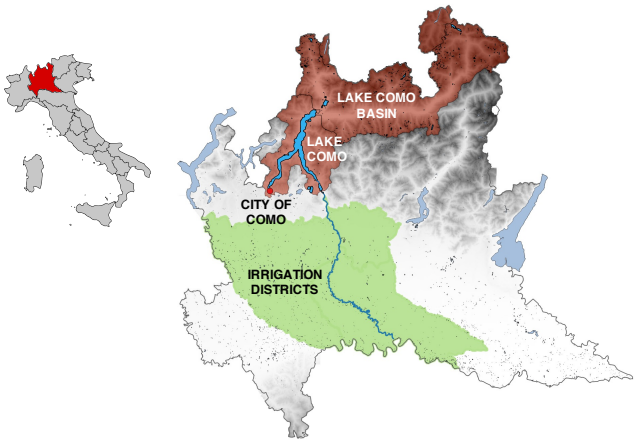

Figure 2. On the left, Lombardy region is highlighted in a map of Italy. On the right, a physical map of Lombardy, comprising Lake Como basin, in red, Lake Como, the city of Como, and the irrigation district downstream the lake.

Flood days: the average number of annual flood days, defined as days in which the lake level h_t is above the flood threshold $\bar{h} = 1.24$ m, i.e.,

$$J^{flood} = \frac{1}{N_y} \sum_{t=0}^{H-1} g_{t+1}^{flood}; \quad g_{t+1}^{flood} = \begin{cases} 1 & \text{if } h_{t+1} \geq \bar{h} \\ 0 & \text{if } h_{t+1} < \bar{h} \end{cases} \tag{3}$$

where N_y is the number of years in the simulation horizon.

Water supply deficit: the daily average squared water deficit with respect to the daily downstream demand w_t, subject to the minimum flow constraint $q^{MEF} = 5$ m^3/s to guarantee environmental stakes. Downstream demand is mainly driven by irrigation and is highest during the crop growing season of spring and summer. The quadratic formulation is selected with the aim of penalizing severe deficits in a single time step, while allowing for more frequent, small shortages. i.e.,

$$J^{irr} = \frac{1}{H} \sum_{t=0}^{H-1} (max(w_t - (r_{t+1} - q^{MEF}), 0))^2 \tag{4}$$

The release decision is conditioned on an annual cyclostationary time index, and thus the decision at the end of the time horizon is no different than during the equivalent period of all previous years. For this application, we used Lake Como inflow data for a 10 year optimization horizon from 1997 to 2006 included. This time span contains a diverse range of hydrological conditions, including average and extreme years, from the 2005 record drought to the late 2000 high inflow pulses. Optimal policies are then tested on three validation chunks: an extended 20-years validation from 1977-1996, a combination of extreme dry years (1949, 1962, 1990, 1994, 2007), and wet years (1951, 1960, 1977, 2008, 2014) selected by searching the driest and wettest years from the available historical record of inflows to Lake Como (1947-2014), discarding the calibration years.

The set of candidate policy inputs employed in this analysis includes perfect forecasts of the lake inflow computed over the historical timeseries at different lead times, ranging from one day to over 6 months (Table 1). The forecasts are of two types: i) Cumulated inflows, which represent the cumulative inflows over a given lead time, and ii) Inflow Anomaly, which corresponds to the anomalies in inflow with respect to the inflow cyclostationary mean, cumulated over a given lead time. As argued in the introduction,

the aim of this methodological contribution is to demonstrate that the optimal policy representation varies with the objective tradeoff, and, therefore, one single policy representation is inadequate to represent the entire tradeoff space. The risk of using real forecasts in order to prove this concept is that the forecast bias may introduce noise and errors, and ultimately alter the information selection. Therefore, as per previous works (*Zhao et al.*, 2011; *Denaro et al.*, 2017), we made the modeling choice of using perfect forecasts with the aim of searching the optimal policy representation for the system, given its hydrology, physical characteristics, and objectives, and without being biased by errors in forecast products.

Lead time	**Feature name**	
	Cumulated inflow	**Inflow Anomaly**
1	C1	A1
2	C2	A2
3	C3	A3
5	C4	A5
7	C7	A7
14	C14	A14
21	C21	A21
28	C28	A28
51	C51	A51
62	C62	A62
75	C75	A75
90	C90	A90
120	C120	A120
145	C145	A145
200	C200	A200

Table 1. Dataset of candidate policy inputs comprising perfect inflow forecasts in terms of cumulated inflows and anomalies at various lead times.

3.1 Experimental Settings

SINEPS was run for 20 independently initialized and randomized seeds, In each seed, the termination criterion (described in Section 2.4) is met at the 4th round, which is responsible for no tangible advancement in the Pareto front, (lower than 5%), therefore, we retain as efficient solutions those generated at round 3. At each round, NEMODPS is run for a Number of Function Evaluations (NFE) equal to 600 thousands, with populations of 600 individuals. When new policy inputs are selected in step 4 of the methods, these are connected to the previously optimized policy architectures with an input-output connection. This set of individuals constitutes the initial population of the new round of NEMODPS optimization, in step 1 of round R2.

4 Results

4.1 Feature selection and policy design

Figure 3 reports the Pareto fronts resulting from 3 optimization rounds of SINEPS with respect to the two objectives of Water supply deficit (vertical axis) and Flood days (horizontal axis), both to be minimized as indicated by the arrows. The black square in the bottom left corner of the graph represents the ideal performance of the POP. In accordance to other studies on the same water system, we find that the conflicts between water supply and flood objectives in Lake Como disappear under the assumption of per-

fect knowledge of future inflow (*Denaro et al.*, 2017). An operating policy with full foresight is able to guarantee a sufficient flood pool to buffer the peak inflow and avoid overflow when physically possible, while storing in the lake any excess of water to be used for irrigation purposes during the dry season. Therefore, the deterministic solution of this MO problem does not yield a Pareto front of efficient solutions, but collapses to a single optimal point into the objective space. However, in the absence of a perfect future foresight, we expect that the addition of tailored information can reduce conflicts between water users.

The first round of NEMODPS optimization, conditioned upon basic information only, produces the Pareto front of white circles that lays in the top right portion of the objective space in Figure 3a, showing a sharp conflict between the two operating objectives. Additionally, a concavity can be recognized in the central region of the Pareto front, for values of the Flood objective between 20 to 80. Concave regions of the front are usually regarded as disadvantageous tradeoffs, as one objectives degrades more than proportionally to the second objective's improvement. The normalized HV indicator (panel b) relative to round R1 scores 0.142, indicating a large space for improvement between POP and R1.

Prior to the second NEMODPS optimization round, a feature selection routine identifies the most suitable variables to inform the operating policies via a two-step selection process. First, promising features are identified based on their correlation, measured in SU, with the policy error trajectory, representative of its information gap (Figure 1, box 3). Figure S2 of the Supplementary Information shows examples of error trajectories against the forecast anomaly lead time that scores the highest SU for different objective tradeoffs. Second, a population comprising all the promising features is evolved via NEMODPS, and the fittest representations prevail through evolutionary competition (Figure 1, box 1 for R>1). In particular, only a subset of the policy representations preselected via SU is likely to survive the evolutionary selection pressure, meanwhile new individuals are generated by recombining existing ones and enabling well-performing representations to survive in future generations and establish in the final Pareto front. Figure S1 of the Supplementary Information reports the intermediate results of the two-fold Feature Selection process, highlighting that evolutionary competition is key to identify a contained and relevant feature set for policy representation.

The result of the second NEMODPS optimization round are represented in Figure 3a with colored triangles. The more informed policies significantly outperform R1, scoring an over 3-fold increase in the HV metric. The color of the triangle corresponds to the new feature added to the policy representation, and divides the R2 front in two, around its middle and in correspondence to the persisting concavity in the Pareto front. The analysis of the selected information may uncover unexpected results: flood-inclined policies do not select short term predictions of fast inflow peaks, but long forecasts lead times (75 days). Vice versa, water supply-inclined policies select, in comparison, slightly shorter lead times (62 days) instead of preferring season-long look-ahead. This behavior can be explained from the point of view of conflict mitigation. A minimally-represented flood-inclined policy has, in fact, already developed a solid strategy to prevent floods when physically possible, namely, keeping a low lake level for the most part of the year to always count on a buffer pool to accommodate incoming inflow peaks. This strategy is valid from a lakeshore protection perspective, yet, comes at a remarkable price in terms of water supply. Such policy, therefore, does not require any additional information on upcoming inflow peaks, as the lake is virtually always ready to buffer them. On the contrary, it can significantly benefit from a longer term information on how to improve irrigation while still remaining strongly flood risk-adverse, thereby alleviating water supply deficit downstream, and mitigating conflicts between water users. In fact, by comparing flood conservative policies of R1 and R2 (left region of the Pareto fronts), we notice that the added information has the effect of improving the policies in the direction

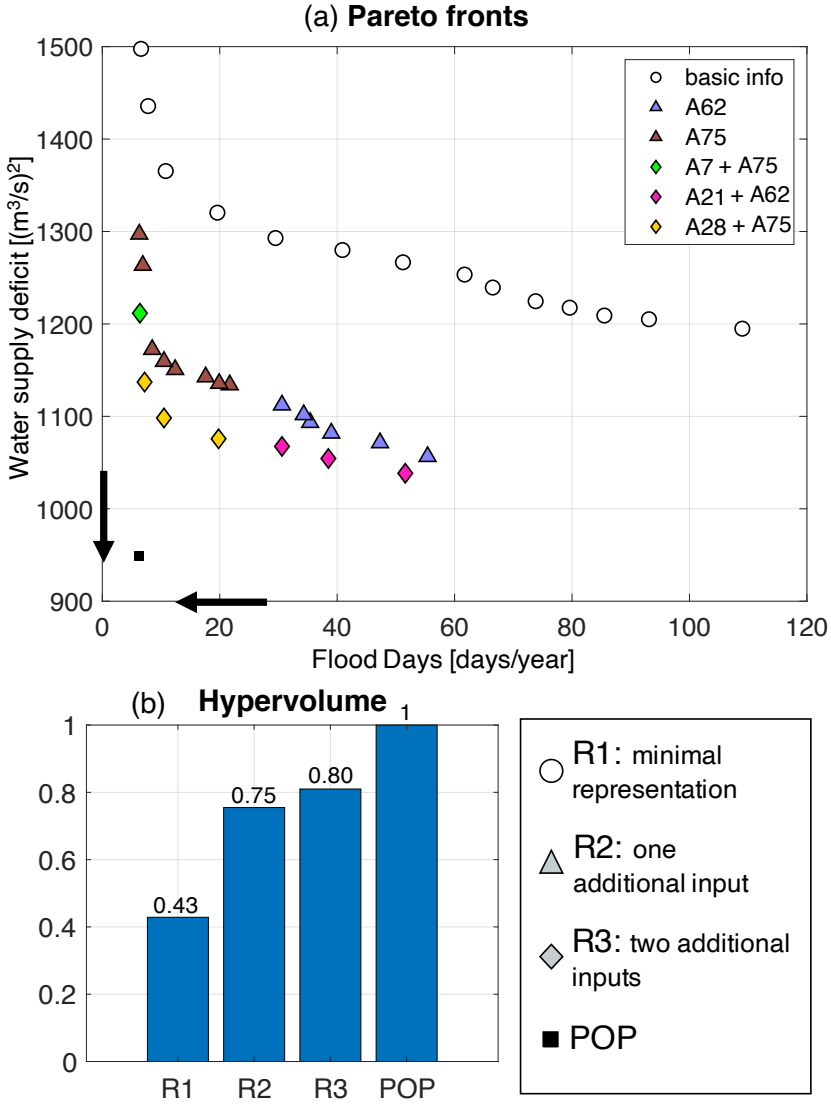

Figure 3. Panel (a): Performance obtained by different Lake Como operating policies with respect to the two cost objectives of water supply deficit (vertical axis) and Flood days (horizontal axis). The black square indicates the ideal performance of the POP, white circles the performance of efficient policies designed at round R1, triangles refer to policies at round R2, and diamonds at round R3. For rounds R2 and R3, the shape color is associated with the information added to the feature representation. Panel (b) shows the improvements in the Hypervolume indicator across different rounds, normalized to the value of hypervolume scored by POP.

of a significantly lower irrigation deficit, at no cost for the flood objective. The long lead time information selected by flood oriented policies is thus employed to minimize objectives conflicts, rather than further improve the flood objective. The other half of the Pareto front selects a shorter lead time, which allows policies to move both in the direction of a reduced flood and irrigation damage. Overall, however, this first round of information selection produces the largest improvement in the reduction of the water supply deficit by employing forecast with a long lead time (2 months or more). This selection is coherent with the multi-seasonal nature of the water supply operations in a snow-dominated system like the one considered in this study. In particular, the reservoir is used to create the seasonal storage by impounding the spring snowmelt-driven inflow peak and distribute it throughout the irrigation season, from spring to autumn, when water supply demand is highest. Forecast lead times of 2+ months are thus used to plan summer irrigation and inform the implementation of effective hedging rules when natural water availability does not meet demand. Lastly, policies select the anomaly in cumulated flow (A75, A62), rather than the flow cumulation, as it is a better indication of whether the system is entering a dry season and hedging strategies should be activated.

The third optimization round includes a second additional information in the policy input set generating further improvement in the HV indicator. The Pareto front of round R3 not only dominates the fronts of the previous rounds, but also resolves their concavity generating a fully convex front, where it is possible to identify a knee. Contrary to the previous round, the front shift between R2 and R3 is mainly horizontal, i.e., contributing to a Flood objective improvement rather than an water supply improvement. Accordingly, the policy inputs selected in this round have a much shorter lead time, between 1 and 4 weeks. The solutions that at this round select the longer lead time, 4 weeks, are those showing a diagonal improvement that unfolds in both objective directions. We note that the by using perfect forecasts to inform the policies, the results shown in our work are upper bounds of what could be achievable with real forecasts in the system. For a demonstrative comparison of the performance using real forecasts instead of perfect forecasts, refer to section S4 of the SI.

It is worth noting that the optimal representations always select the anomaly in flow cumulation, over the flow cumulation. Cumulation time-series are analogous to their anomalies except for an additive cyclostationary, term which corresponds to the annual climatology and expresses the standard hydrological seasonality. However, the policy minimal representation $\pi^R 1 = \pi^R 1(d_t, s_t)$ already contains a cyclostationary time index d_t, which encapsules the climatology. As a consequence, it seems rational for the policy to prefer the selection of an anomaly information over a partially redundant cumulative information. Additionally, it is common for medium-to-long term forecasts products to produce forecast anomalies rather than cumulation (*Crochemore et al.*, 2020).

4.2 The role of information for conflict mitigation

In Figure 4 we explore how added information is employed by progressively informed policies for a given tradeoff. This analysis focuses on the solutions located along the lilac vertical line in panel (a), corresponding to an average of 6.3 flood days a year. This tradeoff was chosen in order to compare the 4 Pareto fronts only in terms of the water supply objective, for a given flood performance. A common cyclostationary behavior emerges for different policy representations in panel (b). The lake recharges in May, in correspondence to the onset of the irrigation season, reaches a level peak around late June, followed by an emptying phase lasting for the entire irrigation season until September/October, when abundant rains cause a new level increase. In the POP, perfect future foresight informs the policy on the exact onset of inflow peaks, allowing to timely generate an adequate flood pool to contain them, while keeping, on average, a high lake level that ensures water availability to supply downstream irrigation demand. Whenever the full trajectory of future disturbance is not available, policies have to be more conservative to-

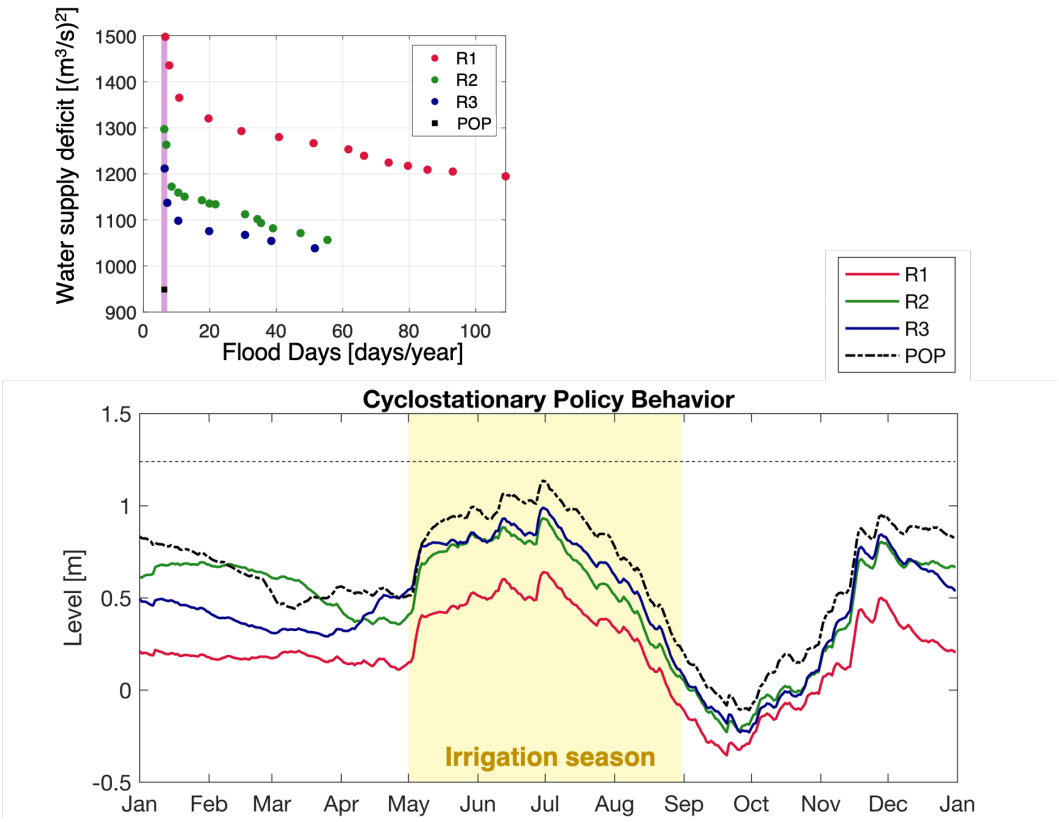

Figure 4. Cyclostationary behavior of efficient policies across different optimization rounds. The investigated policies are aligned along the lilac line in the Pareto front of panel (a) and yield an average number of flood days equal to 6.3, and different values with respect to the water supply objective. In panel (b), their cyclostationary behavior is shown, and contrasted with the Perfect Operating Policy.

wards flood events, thereby keeping a lower lake level to buffer possible incoming inflow peaks, at the expense irrigation availability. This behavior is sharper in the minimally informed round R1 (red line), while more informed policies can confidently maintain a fuller lake during the summer, resulting in a smaller water deficit downstream, without damaging the flood objective. Cyclostationary behaviors outside the irrigation season are fairly divergent, however, the system's winter downstream demand is almost negligible with respect to summer demand, thereby not contributing significantly to the water supply objective performance.

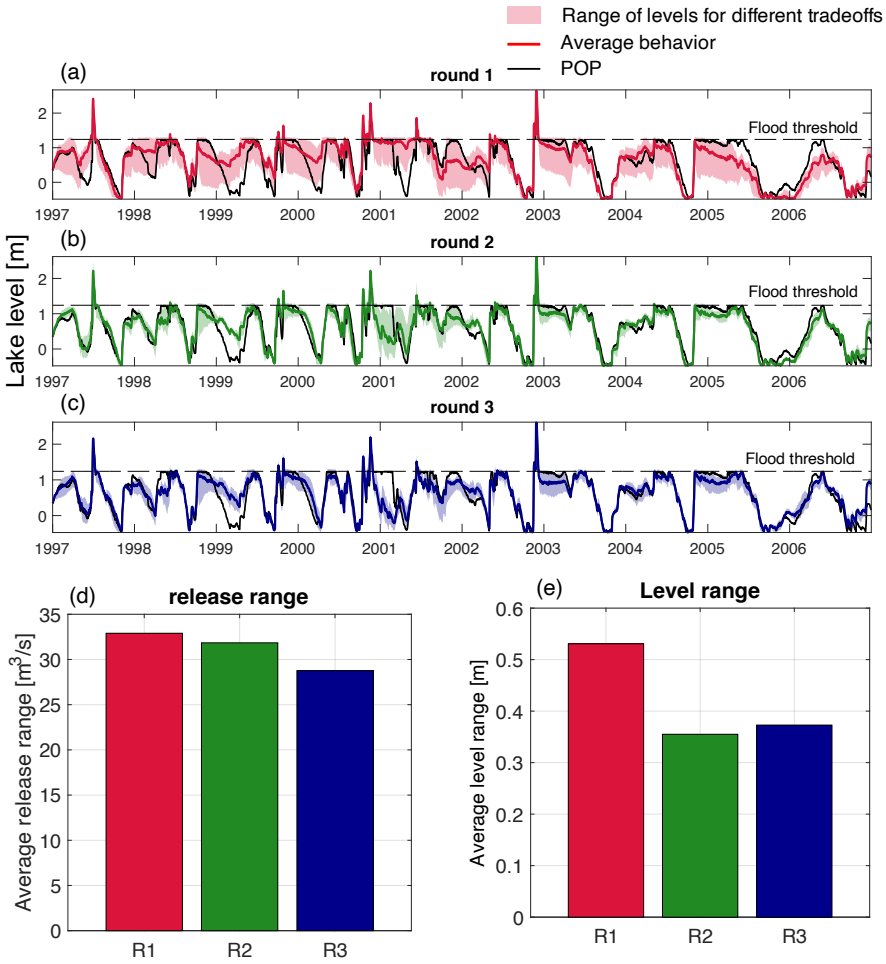

Figure 5. Conflict mitigation. Panels (a), (b), and (c) report the range of lake levels yielded by all the Pareto efficient policies designed at the given optimization round across different tradeoffs. The optimal trajectory is reported in every panel in black for reference. The average round-specific release range is quantified in the barplot of panel (d), while the lake level range is shown in panel (e).

In Figure 5, we analyze how a refinement in policy representation operationally modifies lake regulation towards conflict mitigation. The shaded area in panels (a), (b), and (c) delimits the ensemble of lake level trajectories associated to the set of Pareto efficient policies produced in a given round, while the central colored bold line represents the average behavior. The optimal POP trajectory is reported in black for reference. The wideness of the shaded area indicates the range of variability in operations spanned by the

efficient policies, where a thick area indicates that different tradeoffs are associated with diverse operations, and a narrow area suggests similar operations even across opposite tradeoffs. The plots show a visible narrowing in the operational variability from the first round to the following ones. Operationally, this translates into a mitigated conflict between water users, as different interests tend to converge towards a common efficient policy. This convergence is quantified in the barplots showing the average daily range in levels (panel e) and releases (panel d) associated to different policies in the Pareto set resulting from a given round. The addition of information in the policy representation shows a consistent reduction in release variability. Level variability significantly drops from round R1, where Lake Como is operated at an average difference of more than 53 cm for different tradeoffs, to about 35 cm in round R2. R3 shows a slight increase in variability that is however below 2 cm, and can be considered negligible.

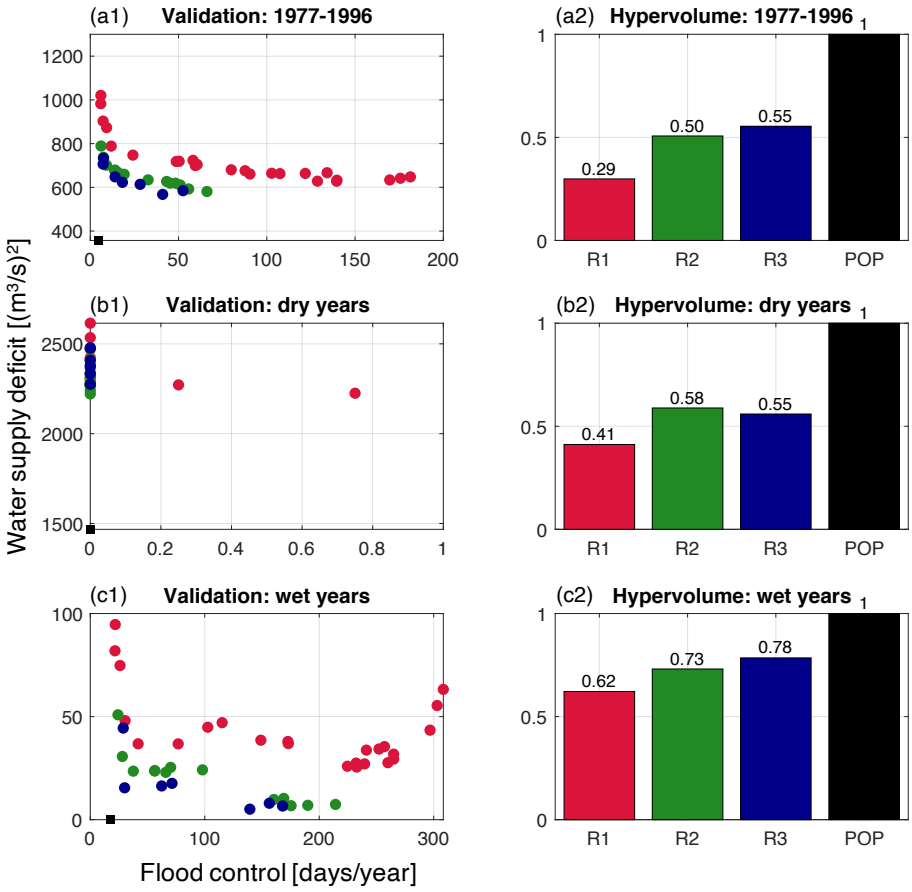

Figure 6. Validation of optimal policies for the three rounds of SINEPS for a 20-year evaluation horizon (panel a1 and a2), and two 5-year evaluation horizons composed of extreme dry (panels b1 and b2) and wet years (panels c1 and c2).

4.3 Policy validation in uncertain hydrological conditions

Figure 6 shows the re-evaluation of the optimal policies on three inflow trajectories, an extended 20-years horizon 1977-1996 (panel a1), an extreme dry (panel b1), and extreme wet horizon (panels c1). Panels a2, b2, and c2 report the value of the HV in-

dicator computed for the different Rounds for the corresponding validation period. The POP performance is reported for reference is each panel colored in black. The most informed round R3 outperforms the other two in the 1977-1996 and wet-years datasets, as quantified by the HV indicator and evident by the Pareto front of optimal validation policies, which is composed by R3 solutions except for sporadic instances of R2 solutions in panel c1. In the dry years dataset, one R2 solution achieve slightly lower water supply deficit compared to R3, but with a fairly negligible difference, under 3%. This analysis shows that the performance improvement resulting from an enhanced information set persists in validation proving the robustness of the information selection technique across highly diverse hydrological conditions.

5 Conclusions

In the past, reservoir operating rules were conditioned upon basic information systems comprising time index and reservoir storage (*Hejazi et al.*, 2008). However, the potential of enhancing the performance of water system operations using information on current or future water availability has long been recognized by researchers and practitioners alike. Despite many features can contribute to operations to some extent, it is in general unclear what is the most effective information set to condition a given water system, for a given tradeoff.

Moreover, previous studies have generally overlooked how defining one single policy representation to characterize the entire tradeoff space of multi-purpose systems can be insufficient. The coexistence of fast and slow process dynamics, and different vulnerabilities requires the search of a tradeoff-tailored policy representation. In this work, we demonstrate for the first time that one input set is inadequate to inform the entire Pareto front of efficient policies that constitutes the solution to a multi-objective problem. In fact, when the policy search routine is allowed to evolve heterogeneous input sets, the selected optimal policy representation will vary Pareto-dynamically with the tradeoff. In this work, we propose SINEPS, a novel framework for automatic, tradeoff-dynamic feature representation and policy learning. SINEPS starts with a population of minimal policies and gradually complexifies their feature representation by selecting variables that surrogate the policy information deficit, measured by comparison to a Perfect Operating Policy. Policies' architectures are adjusted accordingly, in order to accomodate new inputs and support more complex behaviors. We apply SINEPS to the case study of Lake Como, characterized by conflicting heterogeneous objectives, and we use a dataset of deterministic inflow forecasts at different lead times as candidate policy inputs.

Results show that different objective tradeoffs benefit from different information sets with unexpected, but insightful, outcomes. Flood-conservative policies select forecasts with long lead times, thereby improving water supply performance without increasing flood failures. water supply-inclined policies select, in comparison, shorter lead times achieving better flood and water supply results. Not only we notice a trend in the information selected for different tradeoffs, but also across subsequent selection rounds. The first forecast included in the representation at the second round counts on a over 2 months-ahead lead time, and produces the largest improvement in the direction of a lower water supply deficit, and only partially, flood mitigation. In round three, lead times are shorter than a month, enhancing primarily flood mitigation skills. Overall, the search for a tradeoff-specific feature representation demonstrates the potential to significantly enhance the water system overall reliability, resilience towards both dry and wet extremes, while reducing conflicts across conflicting water uses.

Lastly, it is important to note that policy representation in water resources management should not be considered a static concept, but should dynamically adapt in response to variations in the ever-evolving boundary conditions that coupled human-natural systems are exposed to. In particular, the optimal policy representation could change

in response to variations in socio-economic drivers e.g., a water user experiencing unprecedented and more frequent failures; climatic drivers, i.e., an increased likelihood of one of more class of extreme event; and physical drivers, e.g., when a new water user or infrastructure is included in the system. When one or more of these drivers change, the previous policy representation may not be adequate to represent the new system conditions and should be updated accordingly. The SINEPS framework can be run frequently to monitor and adapt to such changes with a rolling calibration horizon that includes new observations as they become available. A critical challenge yet to address is to determine when and how to timely update the feature representation by means of appropriate triggers.

Data availability statement

The data used in this work are freely available upon request from Consorzio dellAdda at `https://addaconsorzio.it/`.

References

Akrour, R., M. Schoenauer, and M. Sebag (2012), April: Active preference learning-based reinforcement learning, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 116–131, Springer.

Alvernaz, S., and J. Togelius (2017), Autoencoder-augmented neuroevolution for visual doom playing, in *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8, IEEE.

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

Assael, J.-A. M., N. Wahlström, T. B. Schön, and M. P. Deisenroth (2015), Data-efficient learning of feedback policies from image pixels using deep dynamical models, *arXiv preprint arXiv:1510.02173*.

Bengio, Y., A. Courville, and P. Vincent (2013), Representation learning: A review and new perspectives, *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Block, P. (2011), Tailoring seasonal climate forecasts for hydropower operations., *Hydrology & Earth System Sciences*, *15*(4).

Blum, A. L., and P. Langley (1997), Selection of relevant features and examples in machine learning, *Artificial intelligence*, *97*(1-2), 245–271.

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2010), Tree-based reinforcement learning for optimal water reservoir operation, *Water Resources Research*, *46*(9).

Castelletti, A., R. Fedorov, P. Fraternali, and M. Giuliani (2016), Multimedia on the mountaintop: Using public snow images to improve water systems operation, in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 948–957.

Crochemore, L., M.-H. Ramos, and I. Pechlivanidis (2020), Can continental models convey useful seasonal hydrologic information at the catchment scale?, *Water Resources Research*, *56*(2), e2019WR025,700.

Cunningham, P. (2008), Dimension reduction, in *Machine learning techniques for multimedia*, pp. 91–112, Springer.

Curran, W., T. Brys, D. Aha, M. Taylor, and W. D. Smart (2016), Dimensionality reduced reinforcement learning for assistive robots, in *2016 AAAI Fall Symposium Series*.

Denaro, S., D. Anghileri, M. Giuliani, and A. Castelletti (2017), Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-

meteorological data, *Advances in water resources*, *103*, 51–63.

Desreumaux, Q., P. Côté, and R. Leconte (2014), Role of hydrologic information in stochastic dynamic programming: a case study of the kemano hydropower system in british columbia, *Canadian Journal of Civil Engineering*, *41*(9), 839–844.

Dobson, B., T. Wagener, and F. Pianosi (2019), An argument-driven classification and comparison of reservoir operation optimization methods, *Advances in Water Resources*, *128*, 74–86.

Doering, K., J. Quinn, P. M. Reed, and S. Steinschneider (2021), Diagnosing the time-varying value of forecasts in multiobjective reservoir control, *Journal of Water Resources Planning and Management*, *147*(7), 04021,031.

Fletcher, S., M. Lickley, and K. Strzepek (2019), Learning about climate change uncertainty enables flexible water infrastructure planning, *Nature communications*, *10*(1), 1–11.

Floreano, D., P. Dürr, and C. Mattiussi (2008), Neuroevolution: from architectures to learning, *Evolutionary Intelligence*, *1*(1), 47–62.

Gal, S. (1979), Optimal management of a multireservoir water supply system, *Water Resources Research*, *15*(4), 737–749.

Gaudel, R., and M. Sebag (2010), Feature Selection as a One-Player Game, in *International Conference on Machine Learning*, ICML 2010 Conference Proceedings Book, pp. 359–366, Haifa, Israel.

Giuliani, M., and A. Castelletti (2019), Data-driven control of water reservoirs using el niño southern oscillation indexes, in *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pp. 1–5, IEEE.

Giuliani, M., J. Herman, A. Castelletti, and P. Reed (2014), Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management, *Water Resources Research*, *50*(4), 3355–3377.

Giuliani, M., F. Pianosi, and A. Castelletti (2015), Making the most of data: an information selection and assessment framework to improve water systems operations, *Water Resources Research*, *51*(11), 9073–9093.

Giuliani, M., A. Castelletti, R. Fedorov, and P. Fraternali (2016a), Using crowdsourced web content for informing water systems operations in snow-dominated catchments, *Hydrology and Earth System Sciences*, *10*(5194), 20–5049.

Giuliani, M., A. Castelletti, F. Pianosi, E. Mason, and P. Reed (2016b), Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations, *Journal of Water Resources Planning and Management*, *142*(2), 04015,050.

Giuliani, M., J. D. Quinn, J. D. Herman, A. Castelletti, and P. M. Reed (2018), Scalable multiobjective control for large-scale water resources systems under uncertainty, *IEEE Transactions on Control Systems Technology*, *26*(4), 1492–1499.

Giuliani, M., M. Zaniolo, A. Castelletti, G. Davoli, and P. Block (2019), Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations, *Water Resources Research*, *55*(11), 9133–9147.

Gleick, P. H. (2003), Global freshwater resources: soft-path solutions for the 21st century, *Science*, *302*(5650), 1524–1528.

Hachiya, H., and M. Sugiyama (2010), Feature selection for reinforcement learning: Evaluating implicit state-reward dependency via conditional mutual information, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 474–489, Springer.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier (2002), Economic value of long-lead streamflow forecasts for columbia river hydropower, *Journal of Water Resources Planning and Management*, *128*(2), 91–101.

Hejazi, M. I., and X. Cai (2011), Building more realistic reservoir optimization models using data mining–a case study of shelbyville reservoir, *Advances in water resources*, *34*(6), 701–717.

Hejazi, M. I., X. Cai, and B. L. Ruddell (2008), The role of hydrologic information in reservoir operation–learning from historical releases, *Advances in water resources*, *31*(12), 1636–1650.

Herman, J. D., J. D. Quinn, S. Steinschneider, M. Giuliani, and S. Fletcher (2020), Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty, *Water Resources Research*, *56*(2), e24,389.

Hundecha, Y., B. Arheimer, C. Donnelly, and I. Pechlivanidis (2016), A regional parameter estimation scheme for a pan-european multi-basin model, *Journal of Hydrology: Regional Studies*, *6*, 90–111.

Kroon, M., and S. Whiteson (2009), Automatic feature selection for model-based reinforcement learning in factored mdps, in *2009 International Conference on Machine Learning and Applications*, pp. 324–330, IEEE.

Lesort, T., N. Díaz-Rodríguez, J.-F. Goudou, and D. Filliat (2018), State representation learning for control: An overview, *Neural Networks*, *108*, 379–392.

Libisch-Lehner, C., H. Nguyen, R. Taormina, H. Nachtnebel, and S. Galelli (2019), On the value of enso state for urban water supply system operators: Opportunities, trade-offs, and challenges, *Water Resources Research*, *55*(4), 2856–2875.

Liu, D.-R., H.-L. Li, and D. Wang (2015), Feature selection and feature learning for high-dimensional batch reinforcement learning: A survey, *International Journal of Automation and Computing*, *12*(3), 229–242.

Loscalzo, S., R. Wright, and L. Yu (2015), Predictive feature selection for genetic policy search, *Autonomous Agents and Multi-Agent Systems*, *29*(5), 754–786.

Macian-Sorribes, H., and M. Pulido-Velazquez (2020), Inferring efficient operating rules in multireservoir water resource systems: A review, *Wiley Interdisciplinary Reviews: Water*, *7*(1), e1400.

MacKay, D. J. (2003), *Information theory, inference and learning algorithms*, Cambridge university press.

Maidment, D. R., and V. T. Chow (1981), Stochastic state variable dynamic programing for reservoir systems analysis, *Water Resources Research*, *17*(6), 1578–1584.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, and F. Vitart (2011), *The new ECMWF seasonal forecast system (System 4)*, vol. 49, European Centre for Medium-Range Weather Forecasts Reading.

Morimoto, J., S.-H. Hyon, C. G. Atkeson, and G. Cheng (2008), Low-dimensional feature extraction for humanoid locomotion using kernel dimension reduction, in *2008 IEEE International Conference on Robotics and Automation*, pp. 2711–2716, IEEE.

Munk, J., J. Kober, and R. Babuška (2016), Learning state representation for deep actor-critic control, in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 4667–4673, IEEE.

Nouri, A., and M. L. Littman (2010), Dimension reduction and its application to model-based exploration in continuous spaces, *Machine Learning*, *81*(1), 85–98.

Oh, J., S. Singh, and H. Lee (2017), Value prediction network, in *Advances in Neural Information Processing Systems*, pp. 6118–6128.

Pechlivanidis, I., L. Crochemore, J. Rosberg, and T. Bosshard (2020), What are the key drivers controlling the quality of seasonal streamflow forecasts?, *Water Resources Research*, *56*(6), e2019WR026,987.

Quinn, J. D., P. M. Reed, M. Giuliani, and A. Castelletti (2017), Rival framings: A framework for discovering how problem formulation uncertainties shape risk man-

agement trade-offs in water resources systems, *Water Resources Research*, *53*(8), 7208–7233.

Quinn, J. D., P. M. Reed, M. Giuliani, A. Castelletti, J. W. Oyler, and R. E. Nicholas (2018), Exploring how changing monsoonal dynamics and human pressures challenge multireservoir management for flood protection, hydropower production, and agricultural water supply, *Water Resources Research*, *54*(7), 4638–4662.

Quinn, J. D., P. M. Reed, M. Giuliani, and A. Castelletti (2019), What is controlling our control rules? opening the black box of multireservoir operating policies using time-varying sensitivity analysis, *Water Resources Research*, *55*(7), 5962–5984.

Shannon, C. E. (1948), A mathematical theory of communication, *The Bell system technical journal*, *27*(3), 379–423.

Si, W., J. Li, P. Ding, and R. Rao (2017), A multi-objective deep reinforcement learning approach for stock index future's intraday trading, in *2017 10th International symposium on computational intelligence and design (ISCID)*, vol. 2, pp. 431–436, IEEE.

Stanley, K. O., and R. Miikkulainen (2002), Efficient reinforcement learning through evolving neural network topologies, in *Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation*, pp. 569–577, Morgan Kaufmann Publishers Inc.

Stanley, K. O., and R. Miikkulainen (2003), A taxonomy for artificial embryogeny, *Artif. Life*, *9*(2), 93–130.

Sturtevant, N. R., and A. M. White (2006), Feature construction for reinforcement learning in hearts, in *International Conference on Computers and Games*, pp. 122–134, Springer.

Tan, M., R. Deklerck, J. Cornelis, and B. Jansen (2013), Phased searching with neat in a time-scaled framework: experiments on a computer-aided detection system for lung nodules, *Artificial intelligence in medicine*, *59*(3), 157–167.

Turner, S., and S. Galelli (2016), Regime-shifting streamflow processes: Implications for water supply reservoir operations, *Water Resources Research*, *52*(5), 3984–4002.

Turner, S., W. Xu, and N. Voisin (2019), Inferred inflow forecast horizons guiding reservoir release decisions across the united states, *Hydrology and Earth System Sciences Discussions*, pp. 1–25.

Van Hoof, H., N. Chen, M. Karl, P. van der Smagt, and J. Peters (2016), Stable reinforcement learning with autoencoders for tactile and visual data, in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3928–3934, IEEE.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The wfdei meteorological forcing data set: Watch forcing data methodology applied to era-interim reanalysis data, *Water Resources Research*, *50*(9), 7505–7514.

Whiteson, S., P. Stone, K. O. Stanley, R. Miikkulainen, and N. Kohl (2005), Automatic feature selection in neuroevolution, in *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, pp. 1225–1232.

Wright, R., S. Loscalzo, and L. Yu (2012), Embedded incremental feature selection for reinforcement learning, *Tech. rep.*, AIR FORCE RESEARCH LAB ROME NY INFORMATION DIRECTORATE.

Xu, W., C. Zhang, Y. Peng, G. Fu, and H. Zhou (2014), A two stage bayesian stochastic optimization model for cascaded hydropower systems considering varying uncertainty of flow forecasts, *Water Resources Research*, *50*(12), 9267–9286.

Yang, W., J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall (2010), Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, *Hydrology Research*,

41(3-4), 211–229.

Zaniolo, M., M. Giuliani, A. F. Castelletti, and M. Pulido-Velazquez (2018), Automatic design of basin-specific drought indexes for highly regulated water systems, *Hydrology and Earth System Sciences*, *22*(4), 2409–2424.

Zaniolo, M., M. Giuliani, and A. Castelletti (2019), Data-driven modeling and control of droughts, *IFAC-PapersOnLine*, *52*(23), 54–60.

Zaniolo, M., M. Giuliani, S. Sinclair, P. Burlando, and A. Castelletti (2021), When timing mattersmisdesigned dam filling impacts hydropower sustainability, *Nature communications* .

Zaniolo, M., M. Giuliani, and A. Castelletti (2021), Neuro-evolutionary direct policy search for multiobjective optimal control, *IEEE Transactions on Neural Networks and Learning Systems*.

Zatarain, J. S., P. M. Reed, J. D. Quinn, M. Giuliani, and A. Castelletti (2017), Balancing exploration, uncertainty and computational demands in many objective reservoir optimization, *Advances in Water Resources*, *109*, 196–210.

Zhang, L., and X. Chen (2021), Feature selection methods based on symmetric uncertainty coefficients and independent classification information, *IEEE Access*, *9*, 13,845–13,856.

Zhang, T. (2009), Adaptive forward-backward greedy algorithm for sparse learning with linear models, in *Advances in Neural Information Processing Systems*, pp. 1921–1928.

Zhao, Q., X. Cai, and Y. Li (2019), Determining inflow forecast horizon for reservoir operation, *Water Resources Research*, *55*(5), 4066–4081.

Zhao, T., X. Cai, and D. Yang (2011), Effect of streamflow forecast uncertainty on real-time reservoir operation, *Advances in water resources*, *34*(4), 495–504.

Zhao, T., J. Zhao, J. R. Lund, and D. Yang (2014), Optimal hedging rules for reservoir flood operation from forecast uncertainties, *Journal of Water Resources Planning and Management*, *140*(12), 04014,041.