Curses, tradeoffs, and scalable management: advancing evolutionary multi-objective direct policy search to improve water reservoir operations

Matteo Giuliani[1]; Andrea Castelletti[23]; Francesca Pianosi[4]; Emanuele Mason[5]; Patrick M. Reed[6]

ABSTRACT

Optimal management policies for water reservoir operation are generally designed via stochastic dynamic programming (SDP). Yet, the adoption of SDP in complex real-world problems is challenged by the three curses of dimensionality, of modeling, and of multiple objectives. These three curses considerably limit SDP's practical application. Alternatively, in this study, we focus on the use of evolutionary multi-objective direct policy search (EMODPS), a simulation-based optimization approach that combines direct policy search, nonlinear approximating networks and multi-objective evolutionary algorithms to design Pareto approximate closed-loop operating policies for multi-purpose water reservoirs. Our analysis explores the technical and practical implications of using EMODPS through a careful diagnostic assessment of the effectiveness and reliability of the overall EMODPS solution design as well as of the resulting Pareto approximate operating policies. The EMODPS approach is evaluated using the multi-purpose Hoa Binh water reservoir in Vietnam, where water operators are seeking to balance the conflicting objectives of maximizing hydropower

[1]PhD, Dept. of Electronics, Information, and Bioengineering, Politecnico di Milano, P.za Leonardo da Vinci, 32, 20133 Milano, Italy. E-mail: matteo.giuliani@polimi.it.

[2]Associate Professor, M. ASCE, Dept. of Electronics, Information, and Bioengineering, Politecnico di Milano, P.za Leonardo da Vinci, 32, 20133 Milano, Italy. E-mail: andrea.castelletti@polimi.it.

[3]Senior scientist, Institute of Environmental Engineering, ETH Zurich, Ramistrasse 101, 8092 Zurich Switzerland.

[4]Research Associate, Dept. of Civil Engineering, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, UK. Email: francesca.pianosi@bristol.ac.uk.

[5]PhD student, Dept. of Electronics, Information, and Bioengineering, Politecnico di Milano, P.za Leonardo da Vinci, 32, 20133 Milano, Italy. E-mail: emanuele.mason@polimi.it.

[6]Professor, M. ASCE, School of Civil and Environmental Engineering, University of Cornell, 211 Hollister Hall, Ithaca, USA.E-mail: patrick.reed@cornell.edu.

production and minimizing flood risks. A key choice in the EMODPS approach is the selection of alternative formulations for flexibly representing reservoir operating policies. In this study, we distinguish the relative performance of two widely used nonlinear approximating networks, namely Artificial Neural Networks and Radial Basis Functions. Our results show that RBF solutions are more effective than ANN ones in designing Pareto approximate policies for the Hoa Binh reservoir. Given the approximate nature of EMODPS, our diagnostic benchmarking uses SDP to evaluate the overall quality of the attained Pareto approximate results. Although the Hoa Binh test case's relative simplicity should maximize the potential value of SDP, our results demonstrate that EMODPS successfully dominates the solutions derived via SDP.

Keywords: water management, direct policy search, multi-objective evolutionary algorithm

INTRODUCTION

Climate change and growing populations are straining freshwater availability worldwide (McDonald et al. 2011) to the point that many large storage projects are failing to produce the level of benefits that provided the economic justification for their development (Ansar et al. 2014). In a rapidly changing context, operating existing infrastructures more efficiently, rather than planning new ones, is a critical challenge to balance competing demands and performance uncertainties (Gleick and Palaniappan 2010). Yet, most major reservoirs have had their operations defined in prior decades (U.S. Army Corps of Engineers 1977; Loucks and Sigvaldason 1982), assuming "normal" hydroclimatic conditions and considering a restricted number of operating objectives. The effectiveness of these rules is however limited, as they are not able to adapt release decisions when either the hydrologic system deviates from the assumed baseline conditions or additional objectives emerge over time. On the contrary, closing the loop between operational decisions and evolving system conditions provides the adaptive capacity needed to face growing water demands and increasingly uncertain hydrologic regimes (Soncini-Sessa et al. 2007).

In the literature, the design problem of closed-loop operating policies for managing water storages has been extensively studied since the seminal work by Rippl (1883). From the first applications by Hall and Buras (1961), Maass et al. (1962), and Esogbue (1989), Dynamic Programming (DP) and its stochastic extension (SDP) are probably the most widely used methods for designing optimal operating policies for water reservoirs (for a review, see Yeh (1985); Labadie (2004); Castelletti et al. (2008), and references therein). SDP formulates the operating policy design problem as a sequential decision-making process, where a decision taken now produces not only an immediate reward, but also affects the next system state and, through that, all the subsequent rewards. The search for optimal policies relies on the use of value functions defined over a discrete (or discretized) state-decision space, which are obtained by looking ahead to future events and computing a backed-up value. In principle, SDP can be applied under relatively mild modeling assumptions (e.g., finite domains of state, decision and disturbance variables, time-separability of objective functions and constraints). In practice, the adoption of SDP in complex real-world water resources problems is challenged by three curses that considerably limit its use, namely the *curse of dimensionality*, the *curse of modeling*, and the *curse of multiple objectives*.

The *curse of dimensionality*, first introduced by Bellman (1957), means that the computational cost of SDP grows exponentially with the state vector dimensionality. SDP would be therefore inapplicable when the dimensionality of the system exceeds 2 or 3 storages (Loucks et al. 2005). In addition, particularly in such large systems, the disturbances (e.g., inflows) are likely to be both spatially and temporally correlated. While including space variability in the identification of the disturbance's probability distribution function (pdf) can be sometimes rather complicated, it does not add to SDP's computational complexity. Alternatively, properly accounting for temporal correlation requires using a dynamic stochastic model, which contributes additional state variables and exacerbates the curse of dimensionality.

The *curse of modeling* was defined by Tsitsiklis and Van Roy (1996) to describe the SDP requirement that, in order to solve the sequential decision-making process at each stage in a step-based optimization, any information included into the SDP framework must be explicitly modeled to fully predict the one-step ahead model transition used for the estimation of the value function. This information can be described either as a state variable of a dynamic model or as a stochastic disturbance, independent in time, with an associated pdf. As a consequence, exogenous information (i.e., variables that are observed but are not affected by the decisions, such as observations of inflows, precipitation, snow water equivalent, etc.), which could potentially improve the reservoir operation (Tejada-Guibert et al. 1995; Faber and Stedinger 2001), cannot be explicitly considered in conditioning the decisions, unless a dynamic model is identified for each additional variable, thus adding to the curse of dimensionality (i.e., additional state variables). Moreover, SDP cannot be combined with high-fidelity process-based simulation models (e.g., hydrodynamic and ecologic), which require a warm-up period and cannot be employed in a step-based optimization mode.

The *curse of multiple objectives* (Powell 2007) is related to the generation of the full set of Pareto optimal (or approximate) solutions to support a posteriori decision making (Cohon and Marks 1975) by exploring the key alternatives that compose system tradeoffs, providing decision makers with a broader context where their preferences can evolve and be exploited opportunistically (Brill. et al. 1990; Woodruff et al. 2013). Most of the DP-family methods relies on single-objective optimization algorithms, which require a scalarization function (e.g., convex combination or non-linear Chebyshev scalarization) to reduce the dimensionality of the objective space to a single-objective problem (Chankong and Haimes 1983; ReVelle and McGarity 1997). The single-objective optimization is then repeated for every Pareto optimal point generated by using different scalarization values (Soncini-Sessa et al. 2007). However, this process is computationally very demanding in many-objective optimization problems, namely when the number of objectives grows to three or more (Fleming et al. 2005), and the accuracy in the approximation of the Pareto front might be degraded given the non-linear relationships between the scalarization values and the corresponding objectives values.

Approximate Dynamic Programming (Powell 2007) and Reinforcement Learning (Busoniu et al. 2010) seek to overcome some or all the SDP curses through three different approaches: (*i*) value function-based methods, which compute an approximation of the value function (Bertsekas and Tsitsiklis 1996); (*ii*) on-line methods, which rely on the sequential resolution of multiple open-loop problems defined over a finite receding horizon (Bertsekas 2005); (*iii*) policy search-based methods, which use a simulation-based optimization to iteratively improve the operating policies based on the simulation outcome (Marbach and Tsitsiklis 2001). However, the first two approaches still rely on the estimation (or approximation) of the value function with single-objective optimization algorithms. Simulation-based optimization, instead, represents a promising alternative to reduce the limiting effects of the three curses of SDP by first parameterizing the operating policy using a given family of functions and, then, by optimizing the policy parameters (i.e., the decision variables of the problem) with respect to the operating objectives of the problem. This approach is generally named direct policy search (DPS, see Rosenstein and Barto (2001)) and is also known in the water resources literature as parameterization-simulation-optimization by Koutsoyiannis and Economou (2003), where has been adopted in several applications (Guariso et al. 1986; Oliveira and Loucks 1997; Cui and Kuczera 2005; Dariane and Momtahen 2009; Guo et al. 2013).

The simulation-based nature of DPS offers some advantages over the DP-family methods. First, the variable domain does not need to be discretized, thus reducing the curse of dimensionality. The complexity of the operating policy (i.e., the number of policy inputs/outputs) however depends on the dimensionality of the system. The higher the number of reservoirs, the more complex is the policy to design, which requires a large number of parameters. Second, DPS can be combined with any simulation model and does not add any constraint on modeled information, allowing the use of exogenous information in conditioning the decisions. Third, when DPS problems involve multiple objectives, they can be coupled with truly multi-objective optimization methods, such as multi-objective evolutionary algorithms (MOEAs), which allow estimating an approximation of the Pareto front in a single run of the algorithm.

Following Nalbantis and Koutsoyiannis (1997), DPS can be seen as an optimization-based generalization of well known simulation-based, single-purpose heuristic operating rules (U.S. Army Corps of Engineers 1977). The New York City rule (Clark 1950), the spill-minimizing “space rule” (Clark 1956), or the Standard Operating Policy (Draper and Lund 2004) can all be seen as parameterized single-purpose policies. Many of these rules are based largely on empirical or experimental successes and they were designed, mostly via simulation, for single-purpose reservoirs (Lund and Guzman 1999). In more complex systems, such as networks of multi-purpose water reservoirs, the application of DPS is more challenging due to the difficulties of choosing an appropriate family of functions to represent the operating policy. Since DPS can, at most, find the best possible solution within the prescribed family of functions, a bad approximating function choice can strongly degrade the final result. For example, piecewise linear approximations have been demonstrated to work well for specific problems, such as hedging rules or water supply (Oliveira and Loucks 1997). In other problems (e.g., hydropower production), the limited flexibility of these functions can however restrict the search to a subspace of policies that, likely, does not contain the optimal one. In many cases, the choice of the policy architecture can not be easily inferred either from the experience of the water managers, who may not be operating the system at full attainable efficiency, or a priori on the basis of empirical considerations, when the system is under construction and data about the historical regulation are not yet available. A more flexible function, depending on a larger number of parameters, has hence to be selected to ensure the possibility of approximating the unknown optimal solution of the problem to any desired degree of accuracy. In this work, we have adopted two widely used nonlinear approximating networks (Zoppoli et al. 2002), namely Artificial Neural Networks (ANNs) and Radial Basis Functions (RBFs), which have been demonstrated to be universal approximators under mild assumptions on the activation functions used in the hidden layer (for a review see Tikk et al. (2003) and references therein).

The selected policy parameterization strongly influences the selection of the optimization approach, which is often case study dependent and may require ad-hoc tuning of the optimization algorithms. Simple parameterizations, defined by a limited number of parameters, can be efficiently optimized via ad-hoc gradient-based methods (Peters and Schaal 2008; Sehnke et al. 2010). On the contrary, gradient-free global optimization methods are preferred when the complexity of the policy parameterization, and consequently the number of parameters to optimize, increases. In particular, evolutionary algorithms (EAs) have been successfully applied in several policy search problems characterized by high-dimensional decision spaces as well as noisy and multi-modal objective functions (Whitley et al. 1994; Moriarty et al. 1999; Whiteson and Stone 2006; Busoniu et al. 2011). Indeed, EAs search strategies, which are based on ranking of candidate solutions, better handle the performance uncertainties than methods relying on the estimation of absolute performance or performance gradient (Heidrich-Meisner and Igel 2008). This property is particular relevant given the stochasticity of water resources systems. In this work, we address the challenges posed by multi-objective optimization under uncertainty by using the self-adaptive Borg MOEA (Hadka and Reed 2013). The Borg MOEA has been shown to be highly robust across a diverse suite of challenging multi-objective problems, where it met or exceeded the performance of other state-of-the-art MOEAs (Reed et al. 2013). In particular, the Borg MOEA overcomes the limitations of tuning the algorithm parameters to the specific problems by employing multiple search operators, which are adaptively selected during the optimization based on their demonstrated probability of generating quality solutions. In addition, it automatically detects search stagnation and self-adapts its search strategies to escape local optima (Hadka and Reed 2012; Hadka and Reed 2013).

In this paper, we first contribute a complete formalization of the evolutionary multi-objective direct policy search (EMODPS) approach to design closed-loop Pareto approximate operating policies for multi-purpose water reservoirs by combining DPS, nonlinear approximating networks, and the Borg MOEA. Secondly, we propose a novel EMODPS diagnostic framework to comparatively analyze the effectiveness and reliability of different policy approximation schemes (i.e., ANNs and RBFs), in order to provide practical recommendations on their use in water reservoir operating problems independently from any case-study specific calibration of the policy design process (e.g., preconditioning the decision space, tuning the optimization algorithm). Finally, we systematically review the main limitations of DP family methods in contrast to using the EMODPS approach for understanding the multi-objective tradeoffs when evaluating alternative operating policies.

The Hoa Binh water reservoir system (Vietnam) is used to demonstrate our framework. The Hoa Binh is a multi-purpose reservoir that regulates the flows in the Da River, the main tributary of the Red River, and is mainly operated for hydropower production and flood control in Hanoi. This case study represents a relatively simple problem which, in principle, should maximize the potential of SDP. As a consequence, if EMODPS met or exceeded the SDP performance, we can expect that the general value of the proposed EMODPS approach would increase when transitioning to more complex problems. The rest of the paper is organized as follows: the next section defines the methodology, followed by the description of the Hoa Binh case study. Results are then reported, while final remarks, along with issues for further research, are presented in the last section.

METHODS AND TOOLS

In this section, we first introduce the traditional formulation of the operating policy design problem adopted in the DP family methods and contrast it with the EMODPS formulation. The EMODPS framework has three main components: (*i*) direct policy search, (*ii*) nonlinear approximating networks, and (*iii*) multi-objective evolutionary algorithms. This section concludes with a description of the diagnostic framework used to distinguish the relative performance of ANN and RBF implementations of the proposed EMODPS approach.

Stochastic Dynamic Programming

Water reservoir operation problems generally require sequential decisions $\mathbf{u}_t$ (e.g., release or pumping decisions) at discrete time instants on the basis of the current system conditions described by the state vector $\mathbf{x}_t$ (e.g., reservoir storage). The decision vector $\mathbf{u}_t$ is determined, at each time step, by an operating policy $\mathbf{u}_t = p(t, \mathbf{x}_t)$. The state of the system is then altered according to a transition function $\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\varepsilon}_{t+1})$, affected by a vector of stochastic external drivers $\boldsymbol{\varepsilon}_{t+1}$ (e.g., reservoir inflows). In the adopted notation, the time subscript of a variable indicates the instant when its value is deterministically known. Since SDP requires that the system dynamics are known, the external drivers can only be made endogenous into the SDP formulation either as state variables, described by appropriate dynamic models (i.e., $\boldsymbol{\varepsilon}_{t+1} = f_t(\boldsymbol{\varepsilon}_t, \cdot)$), or as stochastic disturbances, represented by their associated pdf (i.e., $\boldsymbol{\varepsilon}_{t+1} \sim \phi_t$).

The combination of states and decisions over the time horizon $t = 1, \ldots, H$ defines a trajectory τ , which allows evaluating the performance of the operating policy p as follows:

$$J_p = \Psi[R(\tau)|p] \tag{1}$$

where $R(\tau)$ defines the objective function of the problem (assumed to be a cost) and $\Psi[\cdot]$ is a filtering criterion (e.g., the expected value) to deal with uncertainties generated by $\boldsymbol{\varepsilon}_{t+1}$. The optimal policy p^* is hence obtained by solving the following problem:

$$p^* = \arg\min_p J_p \tag{2}$$

subject to the dynamic constraints given by the state transition function $\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\varepsilon}_{t+1})$.

The DP family methods solve Problem (2) by estimating the expected long-term cost of a policy for each state $\mathbf{x}_t$ at time t by means of the value function

$$Q_t(\mathbf{x}_t, \mathbf{u}_t) = \mathrm{E}_{\boldsymbol{\varepsilon}_{t+1}}[g_t(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\varepsilon}_{t+1}) + \gamma \min_{\mathbf{u}_{t+1}} Q_{t+1}(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})] \tag{3}$$

where $Q_t(\cdot)$ is defined over a discrete grid of states and decisions, $g_t(\cdot)$ represents the immediate (time separable) cost function associated to the transition from state $\mathbf{x}_t$ to state $\mathbf{x}_{t+1}$ under the decision $\mathbf{u}_t$, and $\gamma \in (0, 1]$ a discount factor. With this formulation, the expected value is the statistic used to filter the uncertainty (i.e., $\Psi[\cdot] = \mathrm{E}[\cdot]$). The optimal policy is then derived as the one minimizing the value function, namely $p^* = \arg\min_p Q_t(\mathbf{x}_t, \mathbf{u}_t)$.

The computation of the value function defined in eq. (3) requires the following modeling assumptions (Castelletti et al. 2012): (*i*) the system is modeled as a discrete automaton with finite domains of state, decision, and disturbance variables, with the latter described as stochastic variables with an associated pdf; (*ii*) the objective function must be time-separable along with the problem's constraints; (*iii*) the disturbance process must be uncorrelated in time. Although these assumptions might appear to be restrictive, they can be applied to the majority of the water resources systems by properly enlarging the state vector dimensionality (Soncini-Sessa et al. 2007). For example, a duration curve can be modeled as time-separable by using an auxiliary state variable accounting for the length of time. Unfortunately, the resulting computation of $Q_t(\mathbf{x}_t, \mathbf{u}_t)$ becomes very challenging in high-dimensional state and decision spaces. Let n_x, n_u, n_ε be the number of state, decision, and disturbance variables with N_x, N_u, N_ε the number of elements in the associated discretized domains, the computational complexity of SDP is proportional to $(N_x)^{n_x} \cdot (N_u)^{n_u} \cdot (N_\varepsilon)^{n_\varepsilon}$.

When the problem involves multiple objectives, the single-objective optimization must be repeated for every Pareto optimal point by using different scalarization values, such as changing the weights used in the convex combination of the objectives (Gass and Saaty 1955). The overall cost of SDP to obtain an approximation of the Pareto optimal set is therefore much higher, as a linear increase in the number of objectives considered yields a factorial growth of the number of sub-problems to solve (i.e., a four objective problem requires to solve also 4 single-objective sub-problems, 6 two-objective sub-problems, and 4 three-objective sub-problems (Reed and Kollat 2013; Giuliani et al. 2014)). It follows that SDP cannot be applied to water systems where the number of reservoirs as well as the number of objectives increases. Finally, it is worth noting that the adoption of a convex combination of the objectives allows exploring only convex tradeoff curves, with gaps in correspondence to concave regions. Although concave regions can be explored by adopting alternative scalarization functions, such as the ε-constraint method (Haimes et al. 1971), this approach cannot be applied in the SDP framework because it violates the requirement of time-separability.

Direct Policy Search

Direct policy search (DPS, see Sutton et al. (2000); Rosenstein and Barto (2001)) replaces the traditional SDP policy design approach, based on the computation of the value function, with a simulation-based optimization that directly operates in the policy space. DPS is based on the parameterization of the operating policy p_θ and the exploration of the parameter space Θ to find a parameterized policy that optimizes the expected long-term cost, i.e.

$$p_\theta^* = \arg\min_{p_\theta} J_{p_\theta} \quad \text{s.t. } \theta \in \Theta; \quad \mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \boldsymbol{\varepsilon}_{t+1}) \tag{4}$$

where the objective function J_{p_θ} is defined in eq. (1). Finding p_θ^* is equivalent to find the corresponding optimal policy parameters θ^*.

As reviewed by Deisenroth et al. (2011), different DPS approaches have been proposed and they differ in the methods adopted for the generation of the system trajectories τ used in the estimation of the objective function and for the update and evaluation of the operating policies. Among them, in order to avoid the three curses of SDP and to advance the design of operating policies for multi-purpose water reservoirs, we focus on the use of an evolutionary multi-objective direct policy search (EMODPS) approach (see Figure 1) with the following features:

- *Stochastic trajectory generation*: the dynamic model of the system is used as simulator for sampling the trajectories τ used for the estimation of the objective function. In principle, given the stochasticity of water systems, the model should be simulated under an infinite number of disturbance realizations, each of infinite length, in order to estimate the value of the objective function defined in eq. (1). In practice, the expected value over the probability distribution of the disturbances can be approximated with the average value over a sufficiently long time series of disturbances' realizations, either historical or synthetically generated (Pianosi et al. 2011). An alternative is represented by the analytic computation of the system trajectories (i.e., the dynamics of the state vector probability distributions with the associated decisions). However, this latter is computationally more expensive than sampling the trajectories from the system simulation, even though it can be advantageous for the subsequent policy update, as it allows the analytic computation of the gradients.

- *Episode-based exploration and evaluation*: the quality of an operating policy p_θ (and of its parameter vector θ) is evaluated as the expected return computed on the whole episode (i.e., a system simulation from $t = 0, \ldots, H$) to allow considering non-time separable objectives and constraints (e.g., flow duration curves) without augmenting the state vector's dimensionality. On the contrary, the step-based exploration and evaluation assesses the quality of single state-decision pairs by changing the parameters θ at each time step. As in other on-line approaches, such as traditional model predictive control (Bertsekas 2005), this approach requires setting a penalty function on the final state (condition) of the system to account for future costs (Mayne et al. 2000). Yet, the definition of this penalty function requires the evaluation of the value function and, hence, suffers the same limitation of DP family methods.
- *Multi-objective*: although most of DPS approaches looks at a single measure of policy performance, optimized via single-objective gradient-based optimization methods (Peters and Schaal 2008; Sehnke et al. 2010), we replace the single-objective formulation (eqs. 1-4) with a multi-objective one, where J_{p_θ} and p_θ represent the objective and policy vectors, respectively, that can be solved via multi-objective evolutionary algorithms (MOEAs).

The core components of the EMODPS framework have been selected to alleviate the restrictions posed by the three main curses of SDP: (*i*) EMODPS overcomes the curse of dimensionality, as it avoids the computation of the value function $Q(\mathbf{x}_t, \mathbf{u}_t)$ (see eq. (3))

for each combination of the discretized state and decision variables, along with the biases introduced by the discretization of the state, decision, and disturbance domains (Baxter et al. 2001). In addition, episode-based methods are not restricted to time-separable cost functions, which can depend on the entire simulated trajectory τ. (*ii*) EMODPS overcomes the curse of modeling, as it can be combined with any simulation model as well as it can directly employ exogenous information (e.g., observed or predicted inflows and precipitation) to condition the decisions, without presuming either an explicit dynamic model or the estimation of any pdf. (*iii*) EMODPS overcomes the curse of multiple objectives, as the combination of DPS and MOEAs allows users to explore the Pareto approximate tradeoffs for up to ten objectives in a single run of the algorithm (Kasprzyk et al. 2009; Reed and Kollat 2013; Giuliani et al. 2014).

Beyond these practical advantages, the general application of EMODPS does not provide theoretical guarantees on the optimality of the resulting operating policies, which are strongly dependent on the choice of the class of functions to which they belong and on the ability of the optimization algorithm to deal with non-linear models and objectives functions, complex and highly constrained decision spaces, and multiple competing objectives. Some guarantees of convergence and the associated approximation bounds with respect to a known optimal solution have been defined for some classes of single-objective problems, characterized by time-separable and regular cost functions that can be solved with gradient-based methods (Zoppoli et al. 2002; Gaggero et al. 2014). Nonetheless, EMODPS can also be employed in multi-objective applications where a reference optimal solution cannot be computed due to the problem's complexity, facilitating potentially good approximations of the unknown optimum for a broader class of problems.

Nonlinear approximating networks

The definition of a parameterized operating policy provides a mapping between the decisions $\mathbf{u}_t$ and the policy inputs $\mathcal{I}_t$, namely $\mathbf{u_t} = p_\theta(\mathcal{I}_t)$. In the literature, a number of parameterizations of water reservoir operating rules have been proposed, defining the release decision as a function of the reservoir storage (Lund and Guzman 1999; Celeste and Billib 2009). However, most of these rules have been derived from empirical considerations and for single-objective problems, such as the design of hedging rules for flood management (Tu et al. 2003) or of water supply operations (Momtahen and Dariane 2007). Indeed, if prior knowledge about a (near-)optimal policy is available, an ad-hoc policy parameterization can be designed: parameterizations that are linear in the state variables can be used when it is known that a (near-)optimal policy is a linear state feedback. However, when the complexity of the system increases, more flexible structures depending on a high number of parameters are required to avoid restricting the search for the optimal policy to a subspace of the decision space that does not include the optimal solution. In addition, the presence of multiple objectives may require to condition the decisions not only on the reservoir storage, but also on additional information (e.g., inflows, temperature, precipitation, snow water equivalent (Hejazi and Cai 2009)). Two alternative approaches are available to this end: (i) identify a dynamic model describing each additional information and use the states of these models to condition the operating policies in a DP framework (Tejada-Guibert et al. 1995; Desreumaux et al. 2014); (ii) adopt approximate dynamic programming methods allowing the direct, model-free use of information in conditioning the operating policies (Faber and Stedinger 2001; Castelletti et al. 2010).

In order to ensure flexibility to the operating policy structure and to potentially condition the decisions on several variables, we define the parameterized operating policy p_θ by means of two nonlinear approximating networks, namely Artificial Neural Networks and Gaussian Radial Basis Functions. These nonlinear approximating networks have been proven to be universal approximators (for a review see Tikk et al. (2003) and references therein): under very mild assumptions on the activation functions used in the hidden layer, it has been shown that any continuous function defined on a closed and bounded set can be approximated by a three-layered ANNs (Cybenko 1989; Funahashi 1989; Hornik et al. 1989) as well as by a three-layered RBFs (Park and Sandberg 1991; Mhaskar and Micchelli 1992; Chen and

Chen 1995). Since these features guarantee high flexibility to the shape of the parameterized function, ultimately allowing to get closer to the unknown optimum, ANNs and RBFs have become widely adopted as universal approximators in many applications (Maier and Dandy 2000; Buhmann 2003; de Rigo et al. 2005; Castelletti et al. 2007; Busoniu et al. 2011).

Artificial Neural Networks

Using ANNs to parameterize the policy, the k-th component in the decision vector $\mathbf{u}_t$ (with $k = 1, \dots, n_u$) is defined as:

$$u_t^k = a_k + \sum_{i=1}^{N} b_{i,k} \psi_i(\mathcal{I}_t \cdot \mathbf{c}_{i,k} + d_{i,k}) \tag{5}$$

where N is the number of neurons with activation function $\psi(\cdot)$ (i.e., hyperbolic tangent sigmoid function), $\mathcal{I}_t \in \mathbb{R}^M$ the policy inputs vector, and $a_k, b_{i,k}, d_{i,k} \in \mathbb{R}$, $\mathbf{c}_{i,k} \in \mathbb{R}^M$ the ANNs parameters. To guarantee flexibility to the ANN structure, the domain of the ANN parameters is defined as $-10,000 < a_k, b_{i,k}, \mathbf{c}_{i,k}, d_{i,k} < 10,000$ (Castelletti et al. 2013). The parameter vector θ is therefore defined as $\theta = [a_k, b_{i,k}, \mathbf{c}_{i,k}, d_{i,k}]$, with $i = 1, \dots, N$ and $k = 1, \dots, n_u$, and belongs to $\mathbb{R}^{n_\theta}$, where $n_\theta = n_u(N(M + 2) + 1)$.

Radial Basis Functions

In the case of using RBFs to parameterize the policy, the k-th decision variable in the vector $\mathbf{u}_t$ (with $k = 1, \dots, n_u$) is defined as:

$$u_t^k = \sum_{i=1}^{N} w_{i,k} \varphi_i(\mathcal{I}_t) \tag{6}$$

where N is the number of RBFs $\varphi(\cdot)$ and $w_{i,k}$ the weight of the i-th RBF. The weights are formulated such that they sum to one (i.e., $\sum_{i=1}^{N} w_{i,k} = 1$) and are non-negative (i.e., $w_{i,k} \geq 0 \quad \forall i, k$). The single RBF is defined as follows:

$$\varphi_i(\mathcal{I}_t) = \exp\left[-\sum_{j=1}^{M} \frac{((\mathcal{I}_t)_j - c_{j,i})^2}{b_{j,i}^2} \right] \tag{7}$$

where M is the number of policy inputs $\mathcal{I}_t$ and $\mathbf{c}_i, \mathbf{b}_i$ are the M-dimensional center and radius vectors of the i-th RBF, respectively. The centers of the RBF must lie within the bounded input space and the radii must strictly be positive (i.e., using normalized variables, $\mathbf{c}_i \in [-1, 1]$ and $\mathbf{b}_i \in (0, 1]$, (Busoniu et al. 2011)). The parameter vector θ is therefore defined as $\theta = [c_{i,j}, b_{i,j}, w_{i,k}]$, with $i = 1, \ldots, N$, $j = 1, \ldots, M$, $k = 1, \ldots, n_u$, and belongs to $\mathbb{R}^{n_\theta}$, where $n_\theta = N(2M + n_u)$.

Multi-objective evolutionary algorithms

Multi-objective evolutionary algorithms (MOEAs) are iterative search algorithms that evolve a Pareto-approximate set of solutions by mimicking the randomized mating, selection, and mutation operations that occur in nature (Deb 2001; Coello Coello et al. 2007). These mechanisms allow MOEAs to deal with challenging multi-objective problems characterized by multi-modality, nonlinearity, stochasticity and discreteness, thus representing a promising alternative to gradient-based optimization methods in solving multi-objective water reservoirs problems (see Nicklow et al. (2010) and Maier et al. (2014) and references therein).

In this paper, we use the self-adaptive Borg MOEA (Hadka and Reed 2013), which employs multiple search operators that are adaptively selected during the optimization, based on their demonstrated probability of generating quality solutions. The Borg MOEA has been shown to be highly robust across a diverse suite of challenging multi-objective problems, where it met or exceeded the performance of other state-of-the-art MOEAs (Hadka and Reed 2012; Reed et al. 2013). In addition to adaptive operator selection, the Borg MOEA assimilates several other recent advances in the field of MOEAs, including an ε-dominance archiving with internal algorithmic operators to detect search stagnation, and randomized restarts to escape local optima. The flexibility of the Borg MOEA to adapt to challenging, diverse problems makes it particularly useful for addressing EMODPS problems, where the shape of the operating rule and its parameter values are problem-specific and completely unknown a priori.

Diagnostic framework

In this work, we apply a diagnostic framework developed from the one in Hadka and Reed (2012) to comparatively analyze the potential of the ANN and RBF policy parameterizations in solving EMODPS problems with no specific tuning of the policy design process. Since the presence of multiple objectives does not yield a unique optimal solution, but a set of Pareto optimal solutions, assessing the effectiveness of the policy design results (i.e., how close the solutions found are to the optimal ones) requires to evaluate multiple metrics, such as the distance of the final solutions from the Pareto optimal front or its best known approximation (i.e., reference set), the coverage of the non-dominated space, and the extent of the non-dominated front (Maier et al. 2014). In this work, we adopt three formal metrics, namely generational distance, additive ε-indicator, and hypervolume indicator, which respectively account for convergence, consistency, and diversity (Knowles and Corne 2002; Zitzler et al. 2003). In addition, due to the stochastic nature of the evolutionary algorithms (which can be affected by random effects in initial populations and runtime search operators), each optimization was run for multiple random generator seeds. The reliability of the ANN and RBF policy search is evaluated as the probability of finding a solution that is better or equal to a certain performance threshold in a single run, which measures the variability in the solutions' effectiveness for repeated optimization trials.

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

$$I_{GD}(S, \bar{S}) = \frac{\sqrt{\sum_{\mathbf{s} \in S} d_{\mathbf{s}}^2}}{n_S} \tag{8a}$$

with

$$d_{\mathbf{s}} = \min_{\bar{\mathbf{s}} \in \bar{S}} \sqrt{\sum_{i=1}^{k} [J^i(\mathbf{s}) - J^i(\bar{\mathbf{s}})]^2} \tag{8b}$$

where n_S is the number of points in S, and $d_{\mathbf{s}}$ the minimum Euclidean distance between each point in S and $\bar{S}$. I_{GD} is a pure measure of convergence and the easiest to satisfy, requiring 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, defined as

$$I_\varepsilon(S,\bar{S}) = \max_{\bar{\mathbf{s}}\in\bar{S}} \min_{\mathbf{s}\in S} \max_{1\le i\le k} (J^i(\mathbf{s}) - J^i(\bar{\mathbf{s}})) \tag{9}$$

This metric is very sensitive to gaps in tradeoff and can be viewed as a measure of an approximation set's consistency with the reference set, meaning that all portions of the tradeoff are present (Hadka and Reed 2012). Additionally, it captures diversity because of its focus on the worst case distance. If a Pareto approximate set S has gaps, then solutions from other regions must be translated much further distances to dominate its nearest neighbour in the reference set $\bar{S}$, dramatically increasing the I_ε value.

Finally, the hypervolume measures the volume of objective space dominated by an approximation set, capturing both convergence and diversity. It is the most challenging of the three metrics to satisfy. The hypervolume indicator I_H is calculated as the difference in hypervolume between the reference set $\bar{S}$, and an approximation set S, defined as

$$I_H(S,\bar{S}) = \frac{\int \alpha_S(\mathbf{s})ds}{\int \alpha_{\bar{S}}(\bar{\mathbf{s}})d\bar{s}} \tag{10a}$$

with

$$\alpha(\mathbf{s}) = \begin{cases} 1 & \text{if } \exists \mathbf{s}' \in S \text{ such that } \mathbf{s}' \preceq \mathbf{s} \\ 0 & \text{otherwise} \end{cases} \tag{10b}$$

Overall, a good set of Pareto approximate policies is characterized by low values of the first two metrics and a high value of the third one.

CASE STUDY DESCRIPTION

The Hoa Binh is a multi-purpose regulated reservoir in the Red River basin, Vietnam

(Figure 2). The Red River drains a catchment of 169,000 km^2 shared by China (48%), Vietnam (51%), and Laos (1%). Among the three main tributaries (i.e., Da, Thao, and Lo rivers), the Da River is the most important water source, contributing for 42% of the total discharge at Hanoi. Since 1989, the discharge from the Da River has been regulated by the operation of the Hoa Binh reservoir, which is one of the largest water reservoirs in Vietnam, characterized by a surface area of about 198 km^2 and an active storage capacity of about 6 billion m^3. The dam is connected to a power plant equipped with eight turbines, for a total design capacity of 1920 MW, which guarantees a large share of the national electricity production. Given the large storage capacity, the operation of Hoa Binh has also a key role for flood mitigation in Hanoi in the downstream part of the Red River catchment (Castelletti et al. 2012). In recent years, other reservoirs have been constructed on both the Da and Lo rivers (see the yellow triangles in Figure 2). However, given the limited data available since these reservoirs have started operating, they are not considered in this work.

Model and objectives formulation

The system is modeled by a combination of conceptual and data-driven models assuming a modeling and decision-making time-step of 24 hours. The Hoa Binh dynamics is described by the mass balance equation of the water volume s_t^{HB} stored in the reservoir, i.e.

$$s_{t+1}^{HB} = s_t^{HB} + q_{t+1}^{D} - r_{t+1} \tag{11}$$

where q_{t+1}^D is the net inflow to the reservoir in the interval $[t, t+1)$ (i.e., inflow minus evaporation losses) and r_{t+1} is the volume released in the same interval. The release is defined as $r_{t+1} = f(s_t^{HB}, u_t, q_{t+1}^D)$, where $f(\cdot)$ describes the nonlinear, stochastic relation between the decision u_t, and the actual release r_{t+1} (Piccardi and Soncini-Sessa 1991). The flow routing from the reservoir to the city of Hanoi is instead described by a data-driven feedforward neural network, providing the level in Hanoi given the Hoa Binh release (r_{t+1}) and the Thao (q_{t+1}^T) and Lo (q_{t+1}^L) discharges. The description of the Hoa Binh net inflows (q_{t+1}^D) and the flows in the Thao (q^T_{t+1}) and Lo (q^L_{t+1}) rivers depends on the approach adopted: with SDP, they are modeled as stochastic disturbances; with EMODPS, they are not explicitly modeled as this approach allows to directly embed exogenous information into the operating policies. Further details about the model of the Hoa Binh system can be found in Castelletti et al. (2012) and Castelletti et al. (2013).

The two conflicting interests affected by the Hoa Binh operation are modeled using the following objective formulations, evaluated over the simulation horizon H:

- *Hydropower production* (J^{hyd}): the daily average hydropower production (kWh/day) at the Hoa Binh hydropower plant, to be maximized, defined as

$$\begin{gathered} J^{hyd} = \frac{1}{H} \sum_{t=0}^{H-1} HP_{t+1} \\ \text{with } HP_{t+1} = \left(\eta g \gamma_w \bar{h}_t q^{Turb}_{t+1} \right) \cdot 10^{-6} \end{gathered} \tag{12}$$

  where η is the turbine efficiency, $g = 9.81$ (m/s^2) the gravitational acceleration, $\gamma_w = 1000$ (kg/m^3) the water density, $\bar{h}_t$ (m) the net hydraulic head (i.e., reservoir level minus tailwater level), q^{Turb}_{t+1} (m^3/s) the turbined flow;
- *Flooding* (J^{flo}): the daily average excess level h^{Hanoi}_{t+1}(cm^2/day) in Hanoi with respect to the flooding threshold $\bar{h} = 950$ cm, to be minimized, defined as

$$J^{flo} = \frac{1}{H} \sum_{t=0}^{H-1} max(h^{Hanoi}_{t+1} - \bar{h}, 0)^2 \tag{13}$$

  where h^{Hanoi}_{t+1} is the level in Hanoi estimated by the flow routing model, which depends on the Hoa Binh release (r_{t+1}) along with the Thao (q^T_{t+1}) and Lo (q^L_{t+1}) discharges.

It is worth noting that the proposed model and objective formulations are defined as Markov Decision Processes (Soncini-Sessa et al. 2007) to allow comparing the results of EMODPS with traditional DP-based solutions. A more realistic representation would require the development of hydrological models describing the rivers catchments and the use of a flooding objective function that is not time-separable (e.g., the duration of the flood event, which may induce dykes breaks when exceeding critical thresholds). Yet, these alternatives would enlarge the state vector dimensionality beyond the SDP limits. Moreover, the curse of multiple objectives narrows the number of water-related interests that can be considered, preventing a better understanding of the full set of tradeoffs (e.g., flood peaks vs flood duration) and ignoring less critical sectors (e.g., irrigation and environment). The adopted formulations therefore represent a relatively simplified system representation which, in principle, should maximize the potential of SDP. Given the heuristic nature of EMODPS, which has no guarantee of optimality, we use SDP as a benchmark to evaluate the quality of the approximation attained by the EMODPS operating policies. If EMODPS met or exceeded the SDP performance, the general value of the proposed EMODPS approach would increase by including additional model/objective complexities.

Computational Experiment

The Hoa Binh operating policies are parameterized by means of three-layered nonlinear approximating networks, where different numbers of neurons and basis functions are tested. According to Bertsekas (1976), the minimum set of policy inputs required to produce the best possible performance is the state of the system $\mathbf{x}_t$, possibly coupled with a time index (e.g., the day of the year) to take into account the time-dependency and cyclostationarity of the system and, consequently, of the operating policy. However, according to previous works (Pianosi et al. 2011; Giuliani et al. 2014), the operating policy of the Hoa Binh reservoir benefits from the consideration of additional variables, which cannot be employed in DP methods without enlarging the state-vector dimensionality. In particular, the best operation of the Hoa Binh reservoir is obtained by conditioning the operating policies upon the following input variables $\mathcal{I}_t = [sin(2\pi t)/365, cos(2\pi t)/365, s_t^{HB}, q_t^D, q_t^{lat}]$, where $q_t^{lat} = q_t^T + q_t^L$ is the lateral inflow accounting for the Thao and Lo discharges. The role of the previous day inflow observations q_t^D and q_t^{lat} is key in enlarging the information on the current system condition, particularly with respect to the flooding objective in Hanoi, which depends on both the Hoa Binh releases as well as on the lateral flows of Thao and Lo rivers.

The EMODPS optimization of the parameterized operating policies employs the Borg MOEA. Since it has been demonstrated to be relatively insensitive to the choice of parameters, we use the default algorithm parameterization suggested by Hadka and Reed (2013). Epsilon-dominance values equal to 5000 for J^{hyd} and 5 for J^{flo} are used to set the resolution of the two operating objectives. Each optimization was run for 500,000 function evaluations. To improve solution diversity and avoid dependence on randomness, the solution set from each formulation is the result of 20 random optimization trials. In the analysis of the runtime search dynamics, the number of function evaluations (NFE) was extended to 2 millions. Each optimization was run over the horizon 1962-1969, which has been selected as it comprises normal, wet, and dry years. The final set of Pareto approximate policies for each policy structure is defined as the set of non-dominated solutions from the results of all the 20 optimization trials. The three metrics (i.e., generational distance, additive ε-indicator, and hypervolume indicator) are computed with respect to the overall best known approximation of the Pareto front, obtained as the set of non-dominated solutions from the results of all the 280 optimization runs (i.e., 2 approximating networks times 7 structures times 20 seeds). In total, the comparative analysis comprises 220 million simulations and requires approximately 1,220 computing hours on an Intel Xeon E5-2660 2.20 GHz with 32 processing cores and 96 GB Ram. However, it should be noted that our computational experiment is more rigorous than would be necessary in practice and it was performed to support a rigorous diagnostic assessment of the ANN and RBF policy parameterizations. The EMODPS policy design reliably attained very high fidelity approximations of the Pareto front in each optimization run with approximately 150,000 NFE, corresponding to only 50 computing minutes.

The SDP solutions were designed by computing the value function (eq. 3) over the 2-dimensional state vector $\mathbf{x}_t = [t, s_t]$ and the Hoa Binh release decision u_t. The two objectives are aggregated trough a convex combination as the ε-constraint method would violate the

SDP requirement of time-separability. The policy performance are then evaluated via simulation of the same model used in the EMODPS experiments. The stochastic external drivers are represented as follows:

$$
\begin{aligned}
q_{t+1}^{D} &\sim \mathcal{L}_t \\
q_{t+1}^{T} &= \alpha^T q_{t+1}^{D} + \varepsilon_{t+1}^{T} \\
q_{t+1}^{L} &= \alpha^L q_{t+1}^{D} + \varepsilon_{t+1}^{L}
\end{aligned}
\tag{14}
$$

where $\mathcal{L}_t$ is a log-normal probability distribution and the coefficients (α^T, α^L) describe the spatial correlation of the inflow processes, with normally distributed residuals $\varepsilon_{t+1}^{T} \sim \mathcal{N}^T$ and $\varepsilon_{t+1}^{L} \sim \mathcal{N}^L$. The models of the inflows, namely the three probability distributions $\mathcal{L}_t, \mathcal{N}^T, \mathcal{N}^L$ as well as the coefficients (α^T, α^L), were calibrated over the horizon 1962-1969 to provide the same information employed in the EMODPS approach.

The SDP problem formulation hence comprises two state variables, one decision variable, and three stochastic disturbances. Preliminary experiments allow calibrating the discretization of state, decision, and disturbance vectors as well as the number of weights combinations for aggregating the two competing objectives to identify a compromise between modeling accuracy and computational requirements. Each solution designed via SDP required around 45 computing minutes. In order to obtain an equivalent exploration of the Pareto front as in the EMODPS approach, in principle the SDP should be run for 40 different combinations of the objectives, corresponding to 30 computing hours. Yet, the non-linear relationships between the values of the weights and the corresponding objectives value does not guarantee to obtain 40 different solutions as most of them are likely to be equivalent or Pareto dominated. Despite a very accurate tuning of the objectives' weights, we obtained only four Pareto approximate solutions. Finally, the cost of developing the inflows models should be also considered in the estimation of the overall effort required by the SDP, whereas in the EMODPS case such cost is null given the possibility of directly employing the exogenous information.

RESULTS

In this section, we first use our EMODPS diagnostic framework to identify the most effective and reliable policy approximation scheme for the Hoa Binh water reservoir problem. Secondly, we validate the EMODPS Pareto approximate policies by contrasting them with SDP-based solutions. Finally, we analyze one potentially interesting compromise solution to provide effective recommendation supporting the operation of the Hoa Binh reservoir.

Identification of the operating policy structure

The first step of the EMODPS diagnostic framework aims to identify the best parameterized operating policy's structure in terms of number of neurons (for ANN policies) or basis functions (for RBF policies), for a given number $M = 5$ of policy input variables. Figure 3 shows the results for seven different policy structures with the number of neurons and basis functions increasing from $n = 4$ to $n = 16$. The performance of the resulting Pareto approximate operating policies, computed over the optimization horizon 1962-1969, are illustrated in Figure 3a, with the arrows identifying the direction of preference for each objective. The ideal solution would be a point in the top-left corner of the figure. The figure shows the reference set identified for each policy structure, obtained as the set of non-dominated solutions across the 20 optimization trials performed. The overall reference set, obtained as the set of non-dominated solutions from the results of all the 280 optimization runs (i.e., 2 approximating networks times 7 structures times 20 seeds), is represented by a black dotted line. Comparison of the best Pareto approximate sets attained across all random seed trials changing the structures of both ANNs and RBFs, namely the Pareto approximate solutions represented by different shapes, does not show a clear trend of policy performance improvement with increasing numbers of neurons or basis functions. The results in Figure 3a attest a general superiority of the RBF policies over the ANN ones, particularly in the exploration of the tradeoff region with the maximum curvature of the Pareto front (i.e., for J^{flo} values between 100 and 200, RBFs allows attaining higher hydropower production). The ANN policies do outperform the RBF ones in terms of maximum hydropower production, although this small difference is concentrated in a restricted range of J^{flo}, and, likely, not decision-relevant.

In order to better analyze the effectiveness and the reliability in attaining good approximations of the Pareto optimal set using different ANN/RBF structures, we computed the three metrics of our diagnostic framework on the solutions obtained in each optimization run. The metrics are evaluated with respect to the best known approximation of the Pareto front, namely the overall reference set (i.e., the black dotted line in Figures 3a). Figures 3b-d report the best (solid bars) and average (transparent bars) performance in terms of generational distance I_{GD}, additive ε-indicator I_ε, and hypervolume indicator I_H, respectively. Effective policy parameterizations are characterized by low values of I_{GD} and I_ε, and high values of I_H. The deviations between the best and the average metric values reflect the reliability of the policy design, with large deviations identifying low reliable structures. In contrast with the results in Figure 3a, the values of the metrics show substantial differences between ANNs and RBFs as well as their dependency on the number of neurons and basis functions. The average metrics of RBF policies are consistently better than the ones of ANN policies. Moreover, the average performance of ANN policies degrade when the number of neurons increases (except for $n = 4$, where the number of ANN inputs is larger than the number of neurons) probably because ANNs are overfitting the data, while the RBF policies seem to be less sensitive to the number of basis. It is worth noting that the gap between RBFs and ANNs decreases when looking at the best optimization run. This result suggests that the ANN policy parameterization is very sensitive to the initialization and the sequence of random operators employed during the Borg MOEA search, probably due to the larger domain of the ANN parameters with respect to the RBF ones. In the case of RBFs, indeed, the parameter space is the Cartesian product of the subsets $[-1, 1]$ for each center $c_{j,i}$ and $(0, 1]$ for each radius $b_{j,i}$ and weight $w_{i,k}$. In the case of ANNs, instead, parameters have no direct relationship with the policy inputs. In this work, the domain $-10,000 < a_k, b_{i,k}, \mathbf{c}_{i,k}, d_{i,k} < 10,000$ is used as in Castelletti et al. (2013) to guarantee flexibility to the ANN structure and prevents that any Pareto approximate solution is excluded a priori.

To further compare the performance of RBFs and ANNs, in the second step of the analysis we perform a more detailed assessment of the reliability of attaining high quality Pareto approximations for alternative operating policy structures. To this purpose, we define the reliability of the ANN and RBF policy search as the probability of finding a solution that is better or equal to a certain performance threshold (i.e., 75% or 95%) in a single optimization run, which measures the variability in the solutions' effectiveness for repeated optimization trials. Figure 4 illustrates the probability of attainment with a 75% (panel a) and 95% (panel b) threshold, along with a representative example of these thresholds in the objective space (panel c). Figure 4a shows that the ANN policies are not able to consistently meet the 75% threshold, even in terms of I_{GD} which is generally considered the easiest metric to meet requiring only a single solution close to the reference set. As shown in Figure 4c, not attaining 75% in I_{GD} means to have a very poor understanding of the 2-objective tradeoff, with almost no information on the left half of the Pareto front. The thresholds on I_ε are instead fairly strict, as this metric strongly penalizes the distance from the knee region of the reference set. The results in Figure 4a demonstrates the superiority of the RBF policy parameterizations, which attain 75% of the best metric value with a reliability of 100% independently from the number of basis functions. Assuming that the 75% approximation can be an acceptable approximation level of the Pareto optimal set, these results imply that the Hoa Binh policy design problem can likely be solved by a single optimization run with an RBF policy. However, Figure 4b shows that if the 95% level was required, it would be necessary to run multiple random seeds and to accumulate the best solutions across them.

The results in Figure 4 also allow the identification of the most reliable structure of the operating policies in terms of number of neurons and basis functions. Results in Figure 4a show that the most reliable ANN policy relies on 6 neurons, which attains the highest reliability in I_ε and I_H, while all the RBF policies are equally reliable. By considering a stricter threshold (i.e., 95%), results in Figure 4b show that the most reliable RBF policy, particularly in terms of convergence and diversity (i.e., hypervolume indicator), requires 6 or 8 basis functions. Note that attaining 95% in terms of I_ε resulted to be particularly challenging (i.e., probabilities around 10-15%) and, as illustrated in Figure 4c this threshold is almost equivalent to require the identification of the best known approximation of the Pareto front in a single run. In the following, we select the 6-basis structure because it depends on a lower number of parameters and allows a better comparison with the 6 neurons ANNs.

The last step of the analysis looks at the runtime evolution of the Borg MOEA search to ensure that the algorithm's search is at convergence. To this purpose, we run a longer optimization with 2 millions function evaluations for a 6 neurons ANN policy and a 6 basis RBF policy, with 20 optimization trials for each approximating network. In each run, we track the search progress by computing the values of I_{GD}, I_ε, and I_H every 1,000 function evaluations until the first 50,000 evaluations and, then, every 50,000 until 2 millions. The runtime search performance are reported in Figure 5 as a function of the number of function evaluations used. The values of I_{GD} in Figure 5a show that few function evaluations (i.e., around 250,000) allows the identification of solutions close to the reference set identified from the results obtained at the end of the optimization (i.e., after 2 million function evaluations) across the 20 random optimization trials performed for each approximating network (i.e., 6 neurons ANN and 6 basis RBF). The performance in terms of I_{GD} of both ANN and RBF policies are then almost equivalent from 250,000 to 2 millions function evaluations.

A higher number of function evaluations is instead necessary to reach full convergence in the other two metrics, namely I_ε and I_H illustrated in Figures 5b-c, respectively. In general, the runtime analysis of these two metrics further confirm the superiority of the RBF operating policies over the ANN ones, both in terms of consistency (i.e., I_ε) as well as convergence and diversity (i.e., I_H). Such a superiority of RBFs is evident from the beginning of the search, when it is probably due the larger dimensionality of the ANN parameters' domain, which increases the probability of having a poor performing initial population. However, the Borg MOEA successfully identifies improved solutions for both ANN and RBF policies in few runs, with diminishing returns between 100,000 and 200,000 function evaluations. The search progress stops around 400,000 function evaluations, with the RBF policies that consistently outperform the ANN ones. The limited improvements in the performance of each solution from 400,000 to 2 millions demonstrate the convergence of the Borg MOEA search for both ANNs and RBFs, guaranteeing the robustness of the results previously discussed, which were obtained with 500,000 functions evaluations.

Validation of EMODPS policy performance

The performance of the operating policies discussed in the previous section is computed over the optimization horizon 1962-1969. To validate this performance, the designed operating policies are re-evaluated via simulation over a different horizon, namely 1995-2004, to estimate their effectiveness under different hydroclimatic conditions. We focus the analysis on the most reliable policy structures resulting from the previous section, using a 6 neurons ANN and a 6 basis RBF parameterization. The comparison between the performance over the optimization and the validation horizons is illustrated in Figure 6a, which reports the reference set obtained in the two cases across the 20 optimization trials. It is not surprising that the performance attained over the optimization horizon (transparent solutions) degrade when evaluated over the validation horizon (opaque solutions) since the two sets are independently used in the analysis. Although both ANNs and RBFs successfully explore different tradeoffs between J^{hyd} and J^{flo} over the optimization horizon, the difference in performance between optimization and validation clearly demonstrate that RBF operating policies outperform the ANN ones. This can be explained as a consequence of the ANNs over-fitting during the optimization. Indeed, although a subset of ANN policies is Pareto dominating some RBF solutions over the optimization horizon (i.e., for J^{flo} values between 220 and 300), the ANN Pareto approximate front is completely dominated in validation by the RBF solutions. The designed ANN policies seem to be over fit on the hydroclimatic conditions on which they were trained and suffering from too much parametric complexity. Consequently, the ANN policies fail to manage unforeseen situations. Conversely RBFs maintains good performance over the validation horizon, with the corresponding Pareto front that presents less gaps and with a more consistent exploration of the tradeoff between the two objectives.

Figure 6b contrasts the performance of the RBF policies with solutions designed via Stochastic Dynamic Programming (represented by black circles) over the validation horizon 1995-2004. To provide a fair comparison, we illustrate both the RBF solutions conditioned upon $\mathcal{I}_t = [sin(2\pi t)/365, cos(2\pi t)/365, s_t^{HB}, q_t^D, q_t^{lat}]$ (red crosses) and, those obtained by conditioning the decisions on the same variables employed by SDP, namely the day of the year t and the Hoa Binh storage s_t^{HB} (magenta crosses). Results demonstrate that, despite the theoretical guarantee of optimality, SDP solutions produce a significantly lower performance than EMODPS even with basic information. The two main reasons for this are that SPD uses a simplified representation of the spatial and temporal correlation of the inflows and a discretization of state, decision, and disturbance domains. Optimization experiments with SDP using finer discretization grids (not shown for brevity) demonstrate that improvements enabled by finer resolution would be marginal. In contrast, we expect that SDP performance would likely increase by improving the model of the inflows, either by using an autoregressive model to characterize their autocorrelation in time or by extending the time-series to better estimate their pdf and their spatial correlation. However, this refinement would further increase the computational requirements of SDP. In addition, the difficulty of balancing the two objectives when aggregated through a convex combination produces multiple Pareto dominated or overlapping solutions, ultimately limiting the exploration of the tradeoff between J^{hyd} and J^{flo}. Moreover, this objectives' aggregation provides a convex approximation of the Pareto front and prevents the design of solutions in concave regions, resulting in large gaps among the SDP solutions. This limitation does not affect the EMODPS approach, which indeed identifies Pareto approximate sets with concave region in correspondence to the gaps in the SDP solutions. Finally, the possibility of directly employing exogenous information in conditioning the decisions successfully enhances the resulting policy performance, with the red solutions that completely dominate the magenta and black ones.

Analysis of the EMODPS operating policy

In order to provide effective recommendation supporting the operation of the Hoa Binh reservoir, we select a potential compromise solution (see Figure 6b) and we analyze the corresponding operating policy. Figure 7a provides a multivariate representation of the multi-input single-output RBF policy, approximated with an ensemble of 5,000 elements obtained via Latin Hypercube Sampling of the policy inputs domains. The parallel-axes plot represents each release decision u_t (reported on the first axis and highlighted by the green color ramp) as a line crossing the other axes at the values of the corresponding policy inputs (i.e., the day of the year t, the Hoa Binh storage s_t, and the previous day flow observations of the Da River q_t^D and of the lateral contribution of Thao and Lo Rivers q_t^{lat}, respectively). The figure shows that the highest release decisions (dark green lines) are concentrated at the beginning of the monsoon season (i.e., May and June), when it is necessary to drawdown the reservoir storage to make space for the flood peak, while are less dependent on the Hoa Binh storage or the flow in the Da river. As expected, since the policy under consideration is a compromise between the two objectives, it ensures flood protection by suggesting high releases when the flows in the Thao and Lo rivers are small. Focusing on the second axis, representing the day of the year, it is possible to appreciate the cyclostationary behavior of the operating policy, which provides similar release decisions (i.e., mid-tone green lines) at the beginning (bottom) and at the end (top) of the year.

Further details are provided by Figure 7b-d, which represents the release decision projected as a function of the reservoir storage, with the colors illustrating how the release decision changes depending on the day of the year (panel b), the flow in the Da River (panels c-d), and the lateral flow in Thao and Lo Rivers (panel e). Figure 7b confirms the cyclostationary behavior of the operating policy throughout the year (for fixed, intermediate values of flow in the Da River as well as in the Thao and Lo Rivers). The release decision is indeed increasing to make room for the incoming flood before and during the monsoon season, from May (green lines) to August (blue lines). Then, after the monsoon, it decreases and the operation at the end of the year is equivalent to the one at the beginning of the year (red lines). Figure 7c shows the release decision as a function of the Hoa Binh storage on January the 1^{st} for different values of flow in the Da River (and a fixed intermediate value of flow in the Thao and Lo Rivers). In this case, according to the value of the inflow (i.e., moving from light to dark green) the release decision increases to maximize the hydropower production, while maintaining a high and constant water level in the Hoa Binh reservoir. Although we are considering a compromise policy, such increasing releases are acceptable also in terms of flood protection because the monsoon season is far in the future. The modification of the policy during the monsoon season is evident in Figure 7d, which shows again the release decision as a function of the Hoa Binh storage for different values of flow in the Da River (and a fixed intermediate value of flow in the Thao and Lo Rivers) but on May the 1^{st}. In this case the release decision is first increasing with the inflow but, when this latter exceeds 9,000 m^3/s, it starts decreasing to reduce the flood costs in Hanoi. Finally, Figure 7e represents the dual situation, namely the release decision as a function of the Hoa Binh storage on May the 1^{st} for different values of flow in the Thao and Lo Rivers (and a fixed intermediate value of flow in the Da River). In this case, effective flood protection is obtained by decreasing the release decision when the lateral flow increases (i.e., moving from light to dark green lines).

CONCLUSIONS

The paper formalizes and demonstrates the potential of the evolutionary multi-objective direct policy search approach in advancing water reservoirs operations. The method combines direct policy search method, nonlinear approximating networks, and multi-objective evolutionary algorithms to design Pareto approximate operating policies for multi-purpose water reservoirs. The regulation of the Hoa Binh water reservoir in Vietnam is used as a case study.

The comparative analysis of two widely used nonlinear approximating networks (i.e., Artificial Neural Networks and Gaussian Radial Basis Functions) for the parameterization of the operating policy suggests the general superiority of RBFs over ANNs. Results show that RBF solutions are more effective that ANN ones in designing Pareto approximate policies for the Hoa Binh reservoir, with better performance attained by the associated Pareto fronts in terms of convergence, consistency, and diversity. Moreover, the adopted EMODPS diagnostic framework demonstrates that the search of RBF policies is more reliable than using ANNs, thus guaranteeing an high probability of designing high quality solutions. Finally, the performance of RBF policies consistently outperform the ANN ones also when simulated on a different horizon with respect to the one used for the optimization. Although accurate calibration and preconditioning of ANN policies have been shown to improve their performance (Castelletti et al. 2013), they require a priori information about the shape of the optimal policy. On the contrary, RBF operating policies successfully attain high quality results without any tuning or preconditioning of the policy design process, thus representing a potentially effective, case study-independent option for solving EMODPS problems. In addition, although the Hoa Binh policy design problem formulation as a 2-objective Markov Decision Process should maximize the potential of Stochastic Dynamic Programming, our results demonstrate that EMODPS successfully improves the SDP solutions, showing the potential to overcome most of the limitations of DP family methods. The general value of the proposed EMODPS approach would further increase when transitioning to more complex problems. Finally, the analysis of the RBF policy shows physically sound interpretations, favoring its acceptability for the reservoir operators and contributing quantitative practical recommendation to improve the Hoa Binh regulation.

Future research efforts will focus on testing the scalability of EMODPS with respect to the dimensionality of the state and decision vectors as well as to the number of objectives, particularly to support the use of EMODPS in multireservoir systems (Biglarbeigi et al. 2014), possibly including robustness criteria to face global change (Herman et al. 2015). Moreover, the scope of the comparative analysis might be enlarged by including other approximators, such as fuzzy systems or support vector machine. Finally, a diagnostic assessment on different state-of-the-art MOEAs in EMODPS problems will be developed.

ACKNOWLEDGEMENT

This work was partially supported by the *IMRR - Integrated and sustainable water Management of the Red-Thai Binh Rivers System in changing climate* research project funded by the Italian Ministry of Foreign Affair as part of its development cooperation program. Francesca Pianosi was supported by the Natural Environment Research Council (Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1).

REFERENCES

Ansar, A., Flyvbjerg, B., Budzier, A., and Lunn, D. (2014). "Should we build more large dams? The actual costs of hydropower megaproject development." *Energy Policy*, 69, 43–56.

Baxter, J., Bartlett, P., and Weaver, L. (2001). "Experiments with infinite-horizon, policy-gradient estimation." *J. Artif. Intell. Res.(JAIR)*, 15, 351–381.

Bellman, R. (1957). *Dynamic programming*. Princeton University Press, Princeton.

Bertsekas, D. (1976). *Dynamic programming and stochastic control*. Academic Press, New York.

Bertsekas, D. (2005). "Dynamic programming and suboptimal control: a survey from ADP to MPC." *European Journal of Control*, 11(4-5).

Bertsekas, D. and Tsitsiklis, J. (1996). *Neuro-dynamic programming*. Athena Scientific, Belmont, MA.

Biglarbeigi, P., Giuliani, M., and Castelletti, A. (2014). "Many-objective direct policy search in the Dez and Karoun multireservoir system, Iran." *Proceedings of the World Environmental & Water Resources Congress (ASCE EWRI 2014)*, Portland (Oregon).

Brill., E., Flach, J., Hopkins, L., and Ranjithan, S. (1990). "MGA: A Decision Support System for Complex, Incompletely Defined Problems." *IEEE Transactions on Systems, Man, and Cybernetics*, 20(4), 745–757.

Buhmann, M. (2003). *Radial basis functions: theory and implementations*. Cambridge university press Cambridge.

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

Busoniu, L., Ernst, D., De Schutter, B., and Babuska, R. (2011). "Cross–Entropy Optimization of Control Policies With Adaptive Basis Functions." *IEEE Transactions on systems, man and cybernetics–Part B: cybernetics*, 41(1), 196–209.

Castelletti, A., de Rigo, D., Rizzoli, A., Soncini-Sessa, R., and Weber, E. (2007). "Neuro-dynamic programming for designing water reservoir network management policies." *Control Engineering Practice*, 15(8), 1031–1038.

Castelletti, A., Galelli, S., Restelli, M., and Soncini-Sessa, R. (2010). "Tree-based reinforcement learning for optimal water reservoir operation." *Water Resources Research*, 46(W09507).

Castelletti, A., Pianosi, F., Quach, X., and Soncini-Sessa, R. (2012). "Assessing water reservoirs management and development in Northern Vietnam." *Hydrology and Earth System Sciences*, 16(1), 189–199.

Castelletti, A., Pianosi, F., and Restelli, M. (2013). "A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run." *Water Resources Research*, 49.

Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2008). "Water reservoir control under economic, social and environmental constraints." *Automatica*, 44(6), 1595–1607.

Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2012). "Stochastic and robust control of water resource systems: Concepts, methods and applications." *System Identification, Environmental Modelling, and Control System Design*, Springer, 383–401.

Celeste, A. and Billib, M. (2009). "Evaluation of stochastic reservoir operation optimization models." *Advances in Water Resources*, 32(9), 1429–1443.

Chankong, V. and Haimes, Y. (1983). *Multiobjective decision making: theory and methodology*. North-Holland, New York, NY.

Chen, T. and Chen, H. (1995). "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems." *IEEE Transactions on Neural Networks*, 6(4), 911–917.

Clark, E. (1950). "New York control curves." *Journal of the American Water Works Association*, 42(9), 823–827.

Clark, E. (1956). "Impounding reservoirs." *Journal of the American Water Works Association*, 48(4), 349–354.

Coello Coello, C., Lamont, G., and Veldhuizen, D. V. (2007). *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic Algorithms and Evolutionary Computation)*. Springer, New York, 2 edition.

Cohon, J. L. and Marks, D. (1975). "A review and evaluation of multiobjective programing techniques." *Water Resources Research*, 11(2), 208–220.

Cui, L. and Kuczera, G. (2005). "Optimizing water supply headworks operating rules under stochastic inputs: Assessment of genetic algorithm performance." *Water Resources Research*, 41.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function." *Mathematics of control, signals and systems*, 2(4), 303–314.

Dariane, A. and Momtahen, S. (2009). "Optimization of Multireservoir Systems Operation Using Modified Direct Search Genetic Algorithm." *Journal of Water Resources Planning and Management*, 135(3), 141–148.

de Rigo, D., Castelletti, A., Rizzoli, A., Soncini-Sessa, R., and Weber, E. (2005). "A selective improvement technique for fastening neuro-dynamic programming in water resources network management." *Proceedings of the 16th IFAC World Congress*, Prague (Czech Republic).

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons.

Deisenroth, M., Neumann, G., and Peters, J. (2011). "A Survey on Policy Search for Robotics." *Foundations and Trends in Robotics*, Vol. 2, 1–142.

Desreumaux, Q., Côté, P., and Leconte, R. (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.

Draper, A. and Lund, J. (2004). "Optimal Hedging and Carryover Storage Value." *Journal of Water Resources Planning and Management*, 130(1), 83–87.

Esogbue, A. (1989). "Dynamic programming and water resources: Origins and interconnections." *Dynamic Programming for Optimal Water Resources Systems Analysis*, Prentice-Hall, Englewood Cliffs.

Faber, B. and Stedinger, J. (2001). "Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts." *Journal of Hydrology*, 249(1), 113–133.

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

Funahashi, K. (1989). "On the approximate realization of continuous mappings by neural networks." *Neural networks*, 2(3), 183–192.

Gaggero, M., Gnecco, G., and Sanguineti, M. (2014). "Suboptimal Policies for Stochastic N-Stage Optimization: Accuracy Analysis and a Case Study from Optimal Consumption." *Models and Methods in Economics and Management Science*, F. E. Ouardighi and K. Kogan, eds., number 198 in International Series in Operations Research & Management Science, Springer International Publishing, 27–50.

Gass, S. and Saaty, T. (1955). "Parametric objective function - Part II." *Operations Research*, 3, 316–319.

Giuliani, M., Galelli, S., and Soncini-Sessa, R. (2014). "A dimensionality reduction approach for Many-Objective Markov Decision Processes: application to a water reservoir operation problem." *Environmental Modeling & Software*, 57, 101–114.

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

Giuliani, M., Mason, E., Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2014). "Universal approximators for direct policy search in multi-purpose water reservoir management: A comparative analysis." *Proceedings of the 19th IFAC World Congress*, Cape Town (South Africa).

Gleick, P. and Palaniappan, M. (2010). "Peak water limits to freshwater withdrawal and use." *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), 11155–11162.

Guariso, G., Rinaldi, S., and Soncini-Sessa, R. (1986). "The Management of Lake Como: A Multiobjective Analysis." *Water Resources Research*, 22(2), 109–120.

Guo, X., Hu, T., Zeng, X., and Li, X. (2013). "Extension of parametric rule with the hedging rule for managing multireservoir system during droughts." *Journal of Water Resources Planning and Management*, 139(2), 139–148.

Hadka, D. and Reed, P. (2012). "Diagnostic assessment of search controls and failure modes in many–objective evolutionary optimization." *Evolutionary Computation*, 20(3), 423–452.

Hadka, D. and Reed, P. (2013). "Borg: An Auto–Adaptive Many–Objective Evolutionary Computing Framework." *Evolutionary Computation*, 21(2), 231–259.

Haimes, Y., Lasdon, L., and 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.

Hall, W. and Buras, N. (1961). "The dynamic programming approach to water-resources development." *Journal of Geophysical Research*, 66(2), 517–520.

Heidrich-Meisner, V. and Igel, C. (2008). "Variable metric reinforcement learning methods applied to the noisy mountain car problem." *Recent Advances in Reinforcement Learning*, Springer, 136–150.

Hejazi, M. and Cai, X. (2009). "Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm." *Advances in Water Resources*, 32(4), 582–593.

Herman, J. D., Reed, P. M., Zeff, H. B., and Characklis, G. W. (2015). "How Should Robustness Be Defined for Water Systems Planning under Change?." *Journal of Water Resources Planning and Management*.

Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer feedforward networks are universal approximators." *Neural networks*, 2(5), 359–366.

Kasprzyk, J., Reed, P., Kirsch, B., and Characklis, G. (2009). "Managing population and drought risks using many-objective water portfolio planning under uncertainty." *Water Resources Research*, 45(12).

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

Koutsoyiannis, D. and Economou, A. (2003). "Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems." *Water Resources Research*, 39(6), 1170–1187.

Labadie, J. (2004). "Optimal operation of multireservoir systems: State-of-the-art review." *Journal of Water Resources Planning and Management*, 130(2), 93–111.

Loucks, D. and Sigvaldason, O. (1982). "Multiple-reservoir operation in North America.." *The operation of multiple reservoir systems*, Z. Kaczmarck and J. Kindler, eds., IIASA Collab. Proc. Ser., 1–103.

Loucks, D., van Beek, E., Stedinger, J., Dijkman, J., and Villars, M. (2005). *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*. UNESCO, Paris, France.

Lund, J. and Guzman, J. (1999). “Derived operating rules for reservoirs in series or in parallel.” *Journal of Water Resources Planning and Management*, 125(3), 143–153.

Maass, A., Hufschmidt, M., Dorfman, R., Thomas Jr, H., Marglin, S., and Fair, G. (1962). *Design of water–resource systems*. Harvard University Press Cambridge, Mass.

Maier, H. and Dandy, G. (2000). “Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications.” *Environmental modelling & software*, 15(1), 101–124.

Maier, H., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L., Cunha, M., Dandy, G., Gibbs, M., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D., Vrugt, J., Zecchin, A., Minsker, B., Barbour, E., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., and Reed, P. (2014). “Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions .” *Environmental Modelling & Software*, 62(0), 271–299.

Marbach, P. and Tsitsiklis, J. (2001). “Simulation-based optimization of Markov reward processes.” *IEEE Transactions on Automatic Control*, 46(2), 191–209.

Mayne, D. Q., Rawlings, J. B., Rao, C. V., and Scokaert, P. O. (2000). “Constrained model predictive control: Stability and optimality.” *Automatica*, 36(6), 789–814.

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., Todd, M., and Montgomery, M. (2011). “Urban growth, climate change, and freshwater availability.” *Proceedings of the National Academy of Sciences*, 108(15), 6312–6317.

Mhaskar, H. and Micchelli, C. (1992). “Approximation by superposition of sigmoidal and radial basis functions.” *Advances in Applied mathematics*, 13(3), 350–373.

Momtahen, S. and Dariane, A. (2007). “Direct search approaches using genetic algorithms for optimization of water reservoir operating policies.” *Journal of Water Resources Planning* *and Management*, 133(3), 202–209.

Moriarty, D., Schultz, A., and Grefenstette, J. (1999). "Evolutionary Algorithms for Reinforcement Learning." *Journal of Artificial Intelligence Research*, 11, 199–229.

Nalbantis, I. and Koutsoyiannis, D. (1997). "A parametric rule for planning and management of multiple-reservoir systems." *Water Resources Research*, 33(9), 2165–2177.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., and Zechman, E. (2010). "State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management." *Journal of Water Resources Planning and Management*, 136(4), 412–432.

Oliveira, R. and Loucks, D. P. (1997). "Operating rules for multi reservoir systems." *Water Resources Research*, 33, 839–852.

Park, J. and Sandberg, I. (1991). "Universal approximation using radial-basis-function networks." *Neural computation*, 3(2), 246–257.

Peters, J. and Schaal, S. (2008). "Reinforcement learning of motor skills with policy gradients." *Neural networks*, 21(4), 682–697.

Pianosi, F., Quach, X., and Soncini-Sessa, R. (2011). "Artificial Neural Networks and Multi Objective Genetic Algorithms for water resources management: an application to the Hoa Binh reservoir in Vietnam." *Proceedings of the 18th IFAC World Congress*, Milan, Italy.

Piccardi, C. and Soncini-Sessa, R. (1991). "Stochastic dynamic programming for reservoir optimal control: Dense discretization and inflow correlation assumption made possible by parallel computing." *Water Resources Research*, 27(5), 729–741.

Powell, W. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*. Wiley, NJ.

Reed, P., Hadka, D., Herman, J., Kasprzyk, J., and Kollat, J. (2013). "Evolutionary Multiobjective Optimization in Water Resources: The Past, Present, and Future." *Advances in Water Resources*, 51, 438–456.

Reed, P. M. and Kollat, J. B. (2013). "Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example." *Advances in Water Resources*, 56, 1–13.

ReVelle, C. and McGarity, A. E. (1997). *Design and operation of civil and environmental engineering systems*. John Wiley & Sons.

Rippl, W. (1883). "The capacity of storage reservoirs for water supply." *Minutes of the Proceedings, Institution of Civil Engineers*, Vol. 71, Thomas Telford. 270–278.

Rosenstein, M. and Barto, A. (2001). "Robot weightlifting by direct policy search." *International Joint Conference on Artificial Intelligence*, Vol. 17, Citeseer. 839–846.

Sehnke, F., Osendorfer, C., Rückstieß, T., Graves, A., Peters, J., and Schmidhuber, J. (2010). "Parameter-exploring policy gradients." *Neural Networks*, 23(4), 551–559.

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

Sutton, R., McAllester, D., Singh, S., and Mansour, Y. (2000). "Policy Gradient Methods for Reinforcement Learning with Function Approximation." *Advances in Neural Information Processing Systems*, 12, 1057–1063.

Tejada-Guibert, J., Johnson, S., and Stedinger, J. (1995). "The value of hydrologic information in stochastic dynamic programming models of a multireservoir system." *Water Resources Research*, 31(10), 2571–2579.

Tikk, D., Kóczy, L., and Gedeon, T. (2003). "A survey on universal approximation and its limits in soft computing techniques." *International Journal of Approximate Reasoning*, 33(2), 185–202.

Tsitsiklis, J. and Van Roy, B. (1996). "Feature-Based Methods for Large Scale Dynamic Programming." *Machine Learning*, 22, 59–94.

Tu, M., Hsu, N., and Yeh, W. (2003). "Optimization of reservoir management and operation with hedging rules." *Journal of Water Resources Planning and Management*, 129(2), 86–97.

U.S. Army Corps of Engineers (1977). *Reservoir System Analysis for Conservation, Hy-*

drologic Engineering Methods for Water Resources Development. Hydrologic Engineering Center, Davis, CA.

Whiteson, S. and Stone, P. (2006). “Evolutionary function approximation for reinforcement learning.” *The Journal of Machine Learning Research*, 7, 877–917.

Whitley, D., Dominic, S., Das, R., and Anderson, C. (1994). *Genetic reinforcement learning for neurocontrol problems*. Springer.

Woodruff, M., Reed, P., and Simpson, T. (2013). “Many objective visual analytics: rethinking the design of complex engineered systems.” *Structural and Multidisciplinary Optimization*, 1–19.

Yeh, W. (1985). “Reservoir management and operations models: a state of the art review.” *Water Resources Research*, 21 (12), 1797–1818.

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

Zoppoli, R., Sanguineti, M., and Parisini, T. (2002). “Approximating networks and extended ritz method for the solution of functional optimization problems.” *Journal of Optimization Theory and Applications*, 112(2), 403–440.

List of Figures

FIG. 1. Schematization of the evolutionary multi-objective direct policy search (EMODPS) approach. The dashed line represents the model of the system, the gray box the MOEA algorithm.

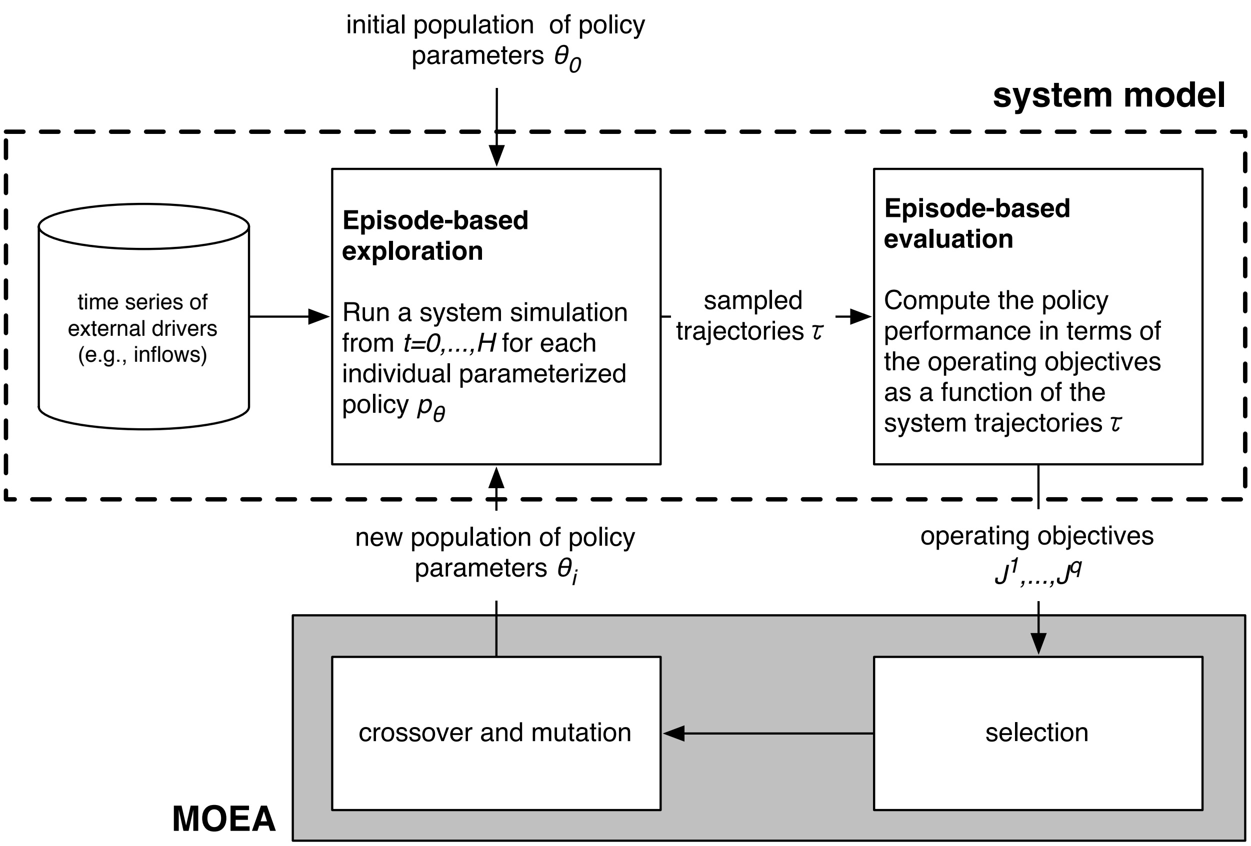


FIG. 2. (a) Map of the Red River basin and (b) schematic representation of the main components described in the model.

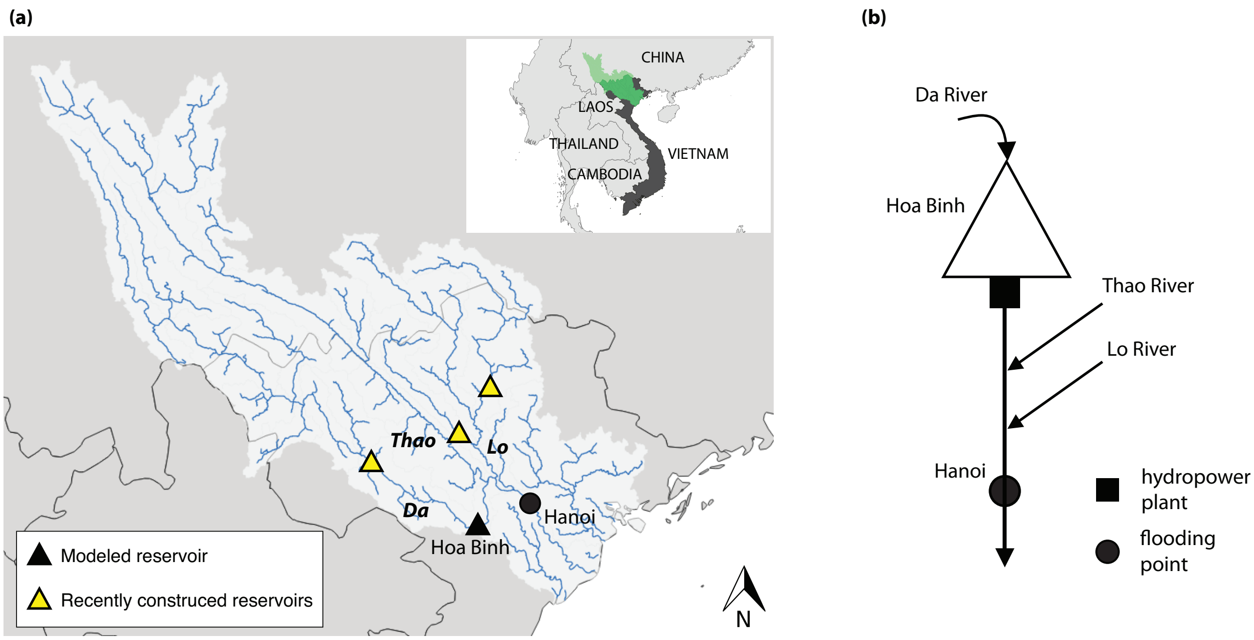


FIG. 3. Policy performance obtained with different structures of ANNs and RBFs over the optimization horizon 1962-1969 (a), and evaluation of the associated Pareto fronts in terms of generational distance (b), additive ε-indicator (c), and hypervolume indicator (d).

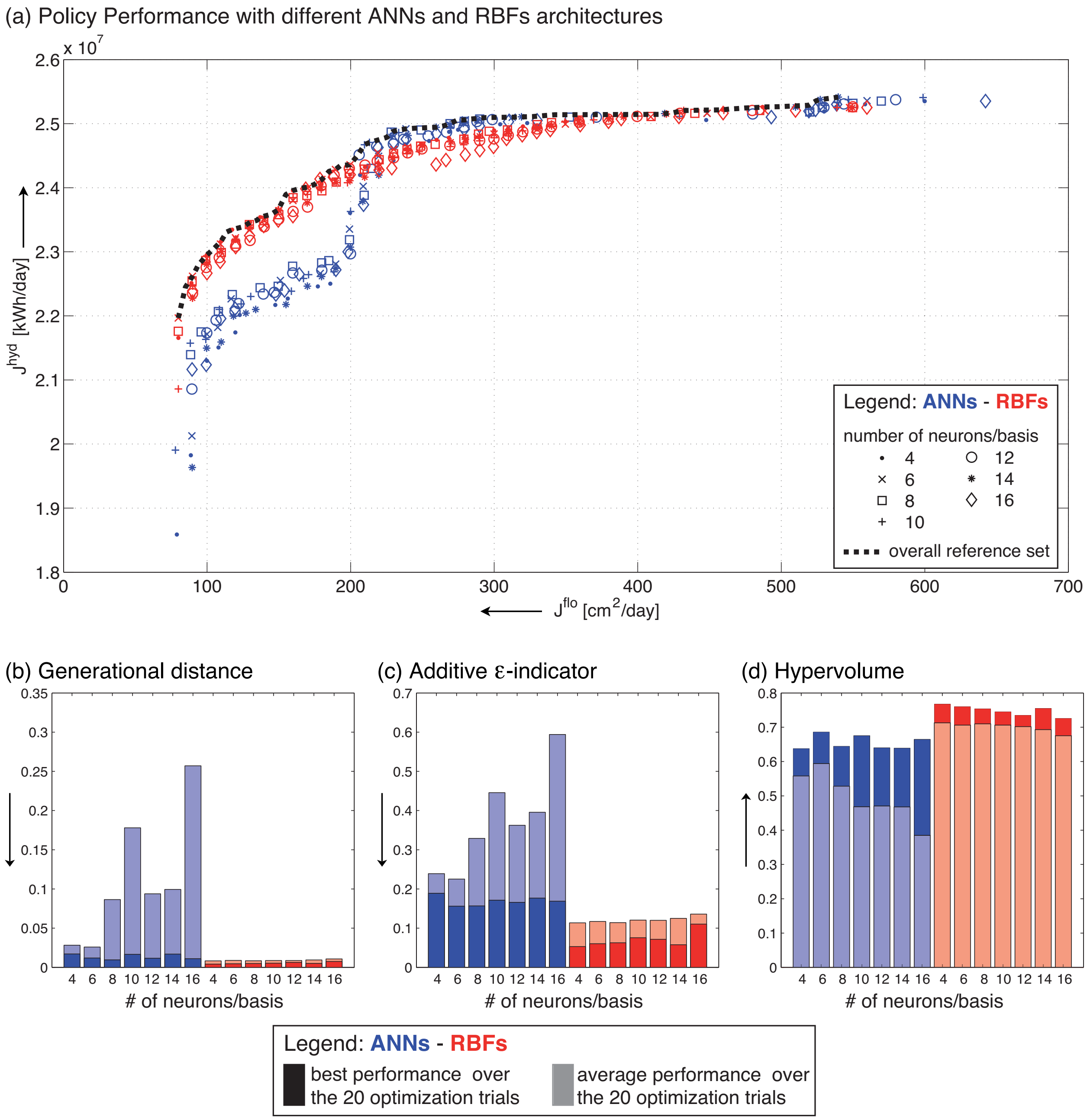


FIG. 4. Probability of attainment with a threshold equal to 75% (a) and to 95% (b) of the best metric values for different ANN and RBF architectures.

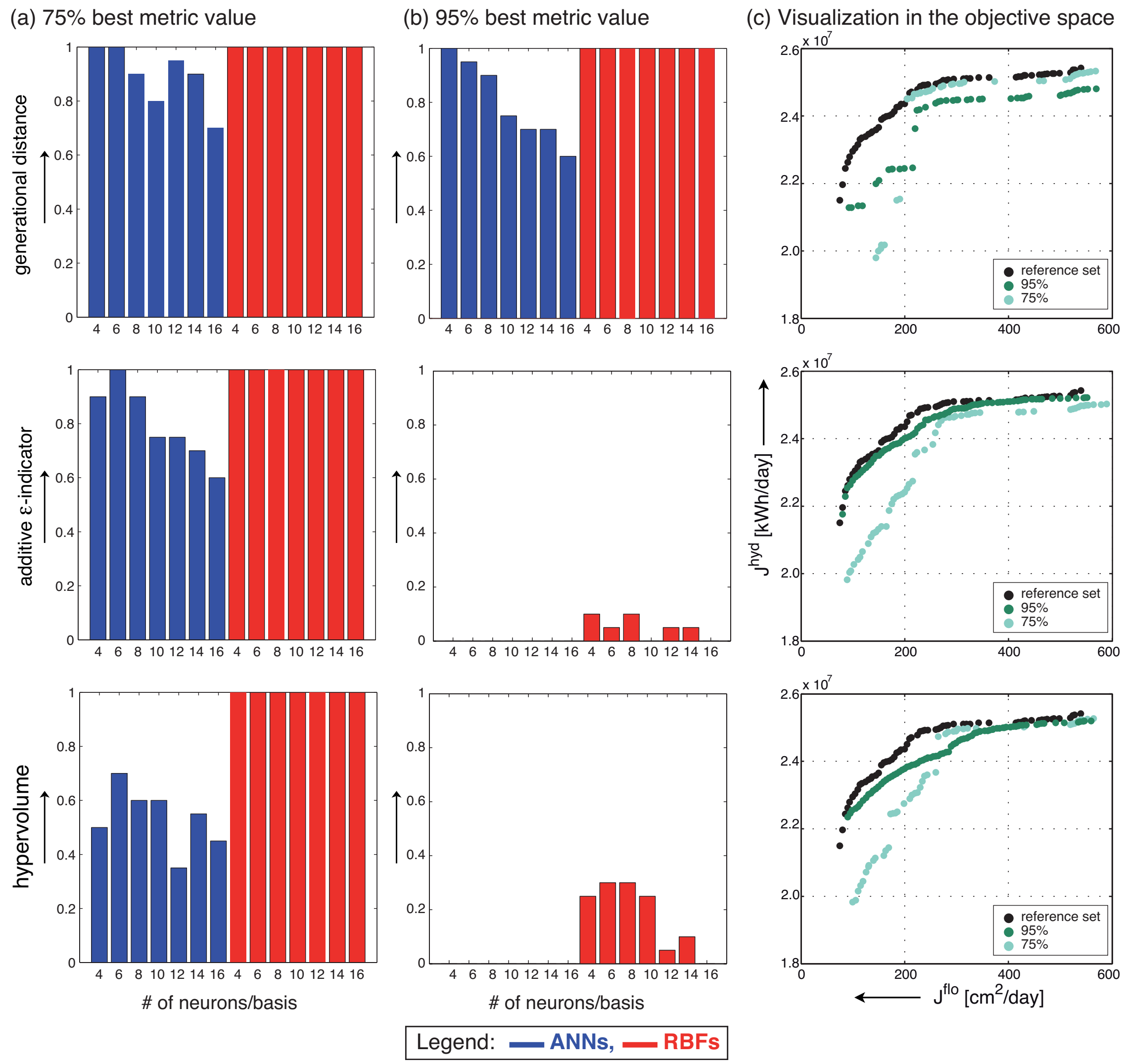


FIG. 5. Analysis of runtime search dynamics for ANN and RBF operating policy optimization in terms of generational distance (a), additive ε-indicator (b), and hypervolume (c).

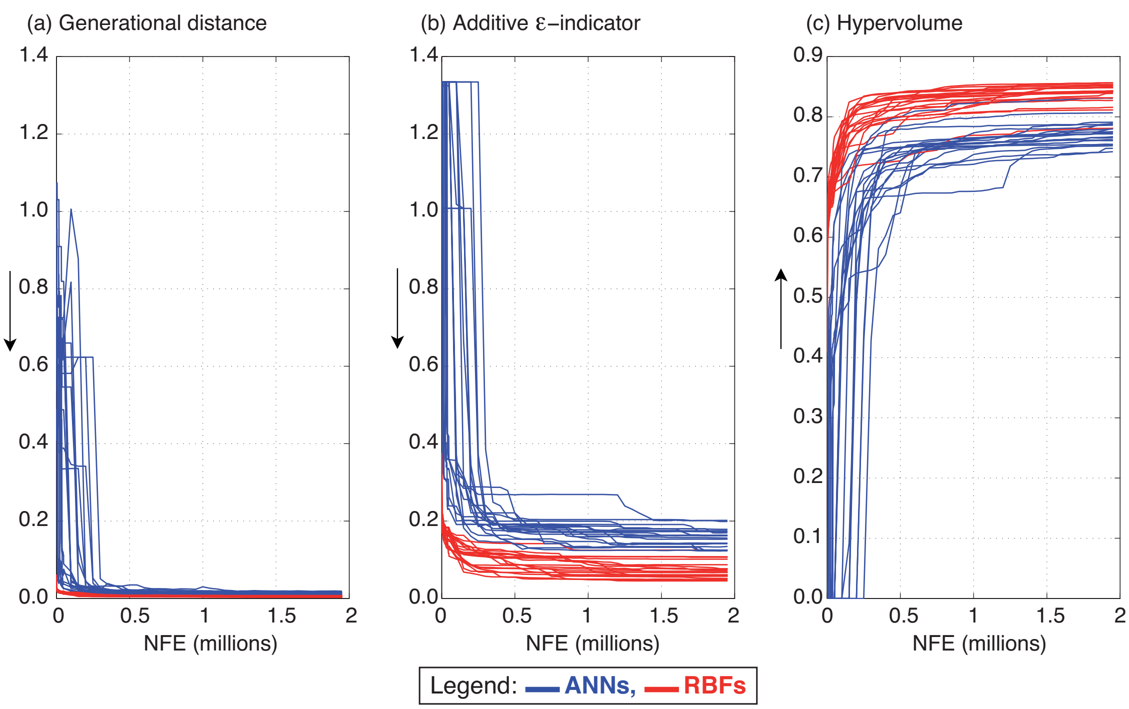
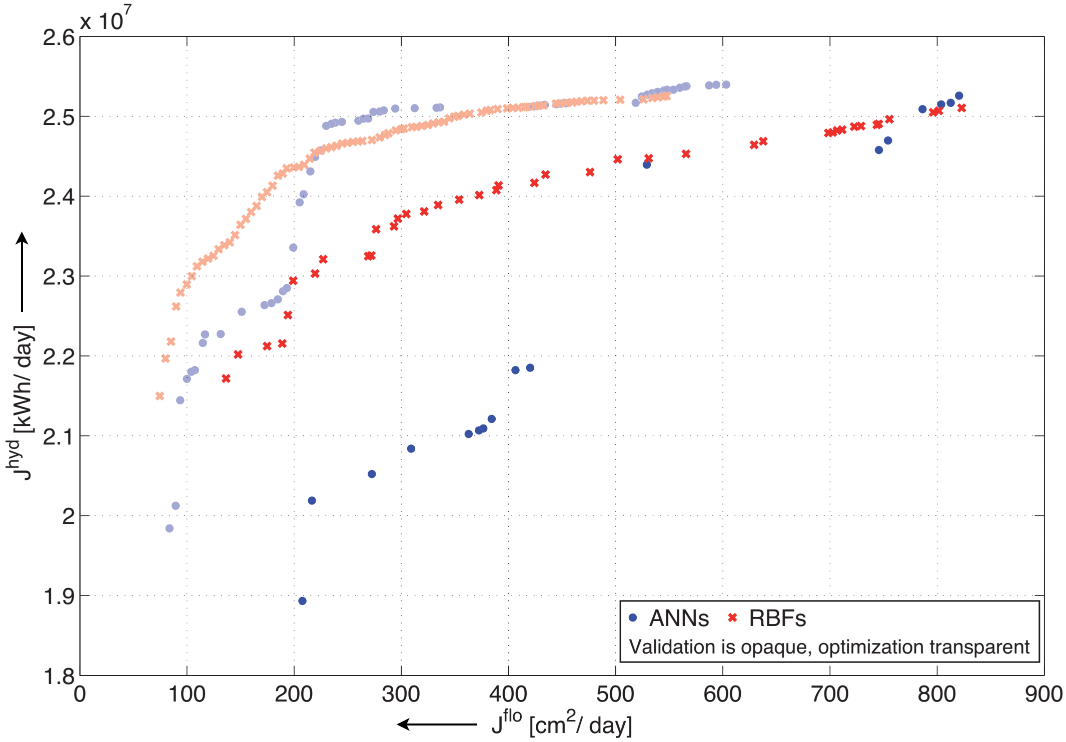


FIG. 6. Validation of EMODPS operating policies via comparison of ANN and RBF performance over the optimization and the validation horizons (a) and comparison with SDP solutions (b).

(a) ANNs and RBFs policy performance over the optimization and validation horizons



(b) RBFs policy performance constrasted with SDP solutions

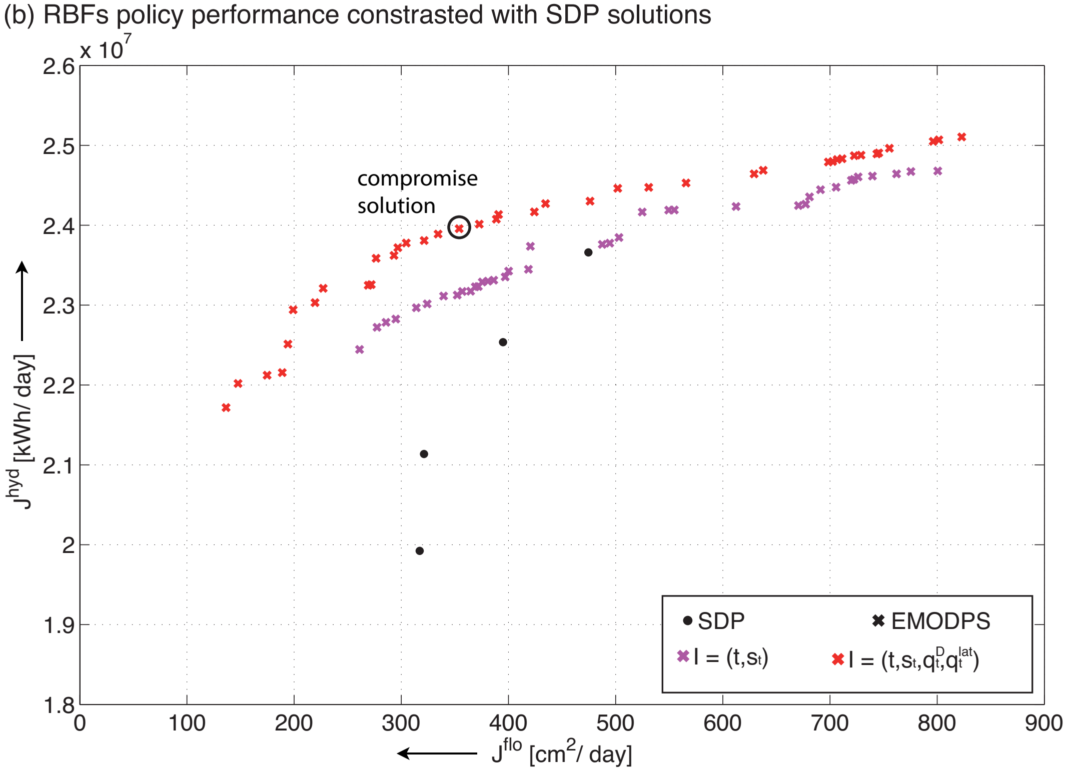


FIG. 7. Visualization of the compromise operating policy selected in Figure 6b.