Surrogate modeling for water reuse planning in complex water systems*

Sangiorgio, Matteo; Cananzi, Davide; Weber, Enrico; Salazar, Jazmin Zatarain; Castelletti, Andrea

DOI
10.1016/j.ifacol.2022.11.018

Publication date
2022

Document Version
Final published version

Published in
IFAC-PapersOnline

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-33 (2022) 111–116

Surrogate modeling for water reuse planning in complex water systems ⋆

Matteo Sangiorgio [*] **Davide Cananzi** [**] **Enrico Weber** [**]
Jazmin Zatarain Salazar [***] **Andrea Castelletti** [*]

[*] *Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy (e-mail: matteo.sangiorgio@polimi.it, andrea.castelletti@polimi.it)*
[**] *Fondazione Politecnico di Milano, Milan, Italy (e-mail: davide.cananzi@mail.polimi.it, enrico.weber@polimi.it)*
[***] *Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands (e-mail: J.ZatarainSalazar@tudelft.nl)*

Abstract: Integrated management of water reuse technologies and coordinated operations with other water system components is fundamental to fully exploiting reuse potential. Yet, these technologies are primarily designed considering their individual efficiency more than possible synergies with traditional water management practices. In this paper, we introduce a general-purpose framework that couples physical and surrogate modelling with optimal control methods to support policy-makers in selecting robust and efficient water planning portfolios, integrating traditional water management strategies and water loops. The framework is developed for the case study of the Apulia Region, Southern Italy, characterised by the presence of a complex water distribution network and multiple conflicting users across irrigation districts, industry, and urban water supply. In addition, the Apulia system shares strategic reservoirs in a drought-prone area. Numerical computations, here performed for the historical period 2010-2019, can be directly applied to consider future climatic scenarios (*i.e.*, modification in precipitation and temperature patterns), socio-economic changes (*i.e.*, variation in the water demand), and technological innovation (*i.e.*, different water reuse strategies). This work represents a first step towards enabling a circular water economy by integrating water management and treatment-reuse technologies.

Keywords: Model reduction and dynamic emulation, Optimal operation of water resources systems, Machine learning for environmental applications, Planning and management for participatory decision making, Integration of technology and environment, Complex system management

1. INTRODUCTION

Water reuse technologies are usually designed as stand-alone elements searching for a balance between costs and efficiency. To fully exploit the potential of these technologies, however, their integrated and coordinated use with the other components of the water system is fundamental to shift towards a circular economy for long-term, sustainable water management. In this paper, we contribute a general framework to design robust and efficient water planning portfolios, integrating traditional water management strategies and water reuse technologies.

We consider the water system of the Apulia region, Southern Italy, as a demo site. This system is extremely complex from both the structural and administrative perspectives. It is characterised by a vast and complex water distribution network and by the presence of multiple conflicting users (irrigation districts, industry, and the Apulian aqueduct, AQP in the Italian acronym) sharing five artificial reservoirs. In addition, the water-scarce hydrologic regime of the region is a suitable testing ground for innovative water reuse technologies to backup the traditional water supply system in particularly dry contingencies.

We first identified the model of the whole Apulian water system (strategic model) using the available observational data. Then, we developed a reduced-order model of the high-fidelity drinking water distribution network model Aquator [1] in the form of a non-linear response surface and coupled it with the strategic model. This allows solving the computational and practical issues related to the direct use of a stand-alone software like Aquator in

⋆ The work has been partially funded by the European Commission under the Project Ô project belonging to Horizon 2020 research and innovation programme (grant no. 776816).

[1] Aquator is a software developed by Oxford Scientific to build and simulate complex models of real-world water resource supply networks. Water utility companies and environment agencies widely use it to determine whether water networks can supply sufficient water to meet customer demand and develop policies on water abstraction.

10.1016/j.ifacol.2022.11.018

the simulation-optimisation procedure adopted to design optimal water planning portfolios. These are obtained by solving a multi-objective optimisation problem using an advanced evolutionary algorithm that generates a set of Pareto-optimal solutions via repeated simulations of the strategic model.

2. MATERIALS AND METHODS

The identification of optimal management strategies or water planning portfolios for the Apulian water system is done by adopting the evolutionary multi-objective direct policy search (EMODPS) methodology (Giuliani et al., 2014b, 2016; Guariso and Sangiorgio, 2020c,b). As shown in Fig. 1, the traditional direct policy search (DPS) loop composed of a strategic simulation model and an optimisation engine (Giuliani et al., 2014a) is here combined with a surrogate component, emulating a complex physically-based implementation of the drinking water distribution system. The workflow is illustrated in the following subsections.

2.1 Strategic system implementation

The key components of the strategic simulation model adopted in this work are the five water reservoirs, Occhito, Conza, Locone, Pertusillo and Monte Cotugno (Fig. 2). The five associated watersheds are not explicitly modelled as we adopt a partial-data driven approach, which directly relies on observational data for the uncontrolled and control-independent components of the system. We derived net inflow time series by inverting equation (1) and using historical time series of storage and releases. In addition, two other water sources are implemented in the strategic model: the natural springs of Cassano and Caposele, and the Salento region well system that extracts groundwater flows from the aquifer. Downstream of each reservoir, a diversion dam distributes the release among different water users. With the only exception of the Monte Cotugno reservoir, which is also used for industrial purposes, all the dams serve both an irrigation district and the water distribution network.

We illustrate the mathematical model of the AQP topology considering a generic single-reservoir model connected through a river stretch with a diversion point (Fig. 3) serving three different types of users (agricultural, industrial, and drinking water supply). The reservoir dynamics are described by the mass balance equation of the water volume s_t stored in the reservoir at time t:

$$s_{t+1} = s_t + q_{t+1} - r_{t+1} \tag{1}$$

where q_{t+1} is the net inflow to the reservoir (*i.e.*, inflow minus evaporation losses and other minor phenomena such as the seepage losses) in the time interval $[t; t+1)$; r_{t+1} the volume of water released over the same time interval. In the adopted notation, the time subscript of a variable indicates the instant when its value is deterministically known. The storage s_t is observed at time t, whereas the inflow has subscript $t+1$, denoting the realisation of the stochastic inflow process in the time interval $[t; t+1)$.

The release is defined as $r_{t+1} = f(s_t, u_t, q_{t+1})$, where $f(\cdot)$ describes the non-linear, stochastic relation between the release decision and the actual release r_{t+1} (Piccardi and Soncini-Sessa, 1991). The release decision is determined by the closed loop operating policy $u_t = p(\cdot)$, based on the current state of the system (t, s_t) (*i.e.*, day of the year, reservoir storage). The actual release is generally equal to the release decision unless physical constraints rule it out, *e.g.*, if the prescribed release lies outside the minimum and maximum allowable releases, if there is insufficient water to meet the prescribed release, or if the prescribed release would result in the reservoir storage capacity being exceeded. Note that u_t also contains the total amount of water captured by the well fields and the fraction of water to be distributed between the different water users in the diversion nodes. The actual release is distributed to the different users through a regulated diversion mechanism (Celeste and Billib, 2009). The volume of water diverted to agricultural districts, industries, and the aqueduct are $rirr_{t+1}$, $rind_{t+1}$ and $rdri_{t+1}$ and should ideally fulfil the corresponding water demands. Each element of the system has to be individually defined and validated and then integrated to build the strategic model. For instance, each reservoir has its characteristics (*e.g.*, level-surface-storage and min-max release relationship) incorporated in the corresponding model component. The water distribution system in Fig. 2 is modelled with the Aquator tool, which returns the drinking water distribution deficit and cost. However, because of its real-to-run ratio (a simulation on a one-year time horizon requires approximately one hour of computing time), Aquator cannot be used in simulation-based optimisations involving millions of runs. To avoid these computational issues, we substitute Aquator for a surrogate model that will be presented below.

The simulation of the system requires the definition of the planning scenario, *i.e.*, the set of input variables that we cannot control in the context of our optimal management problem but have an influence on the system's dynamics. Typical examples of these variables are inflows, water demands, and treatment costs. In this work, the scenario is defined as the set $\omega = \{q_1^h, wirr_0^{h-1}, wind_0^{h-1}, wdri_0^{h-1}, cdri_0^{h-1}\}$, where q_1^h is the vector of all inflows, $wirr_0^{h-1}$, $wind_0^{h-1}$ and $wdri_0^{h-1}$ are the specific water demands, and $cdri_0^{h-1}$ represents the vector of all supply, treatment, and distribution costs depending on the structural configuration and the technologies adopted. Specifically, we focus on the baseline scenario, constituted by observational inflows and structural factors. However, the framework can also be used to design water planing portfolios against future hydrologic, structural and socio-economic scenarios.

The interest of the stakeholders is formally represented by a four-dimensional vector $J = [J^{irr}, J^{ind}, J^{dri,d}, J^{dri,c}]$ of operating objectives to be minimised, whose components are the following:

- Squared irrigation deficit

$$J^{irr} = \sum_{i=1}^{5} \frac{1}{h} \sum_{t=0}^{h-1} (\max(wirr_t(i) - rirr_{t+1}(i), 0))^2 \tag{2}$$

  where $wirr_t(i)$ and $rirr_{t+1}(i)$ are the irrigation water demand and supply for the i-th irrigation district.
- Squared industrial deficit

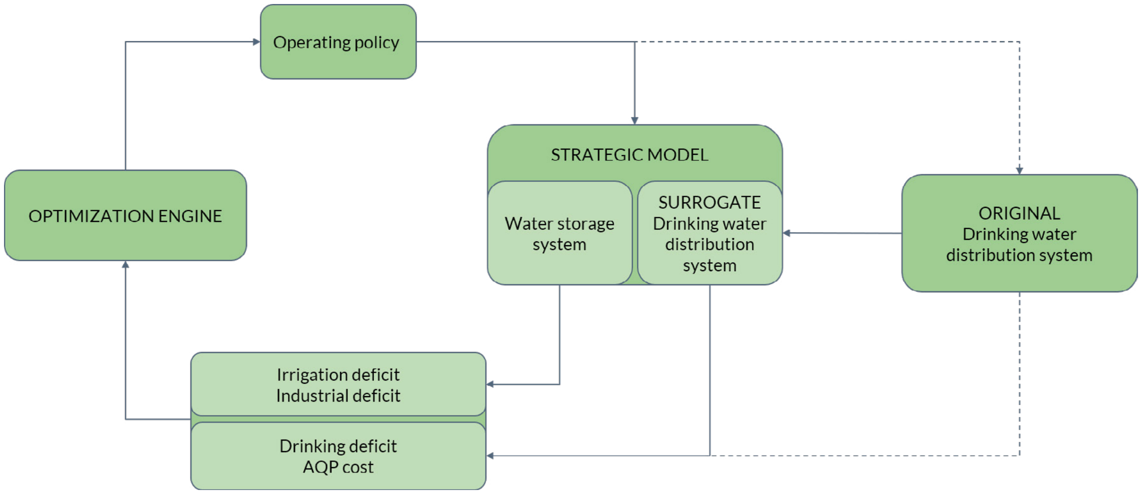

Fig. 1. The DPS workflow with simulation and optimisation steps. The surrogate model replaces the original Aquator implementation in the strategic model.

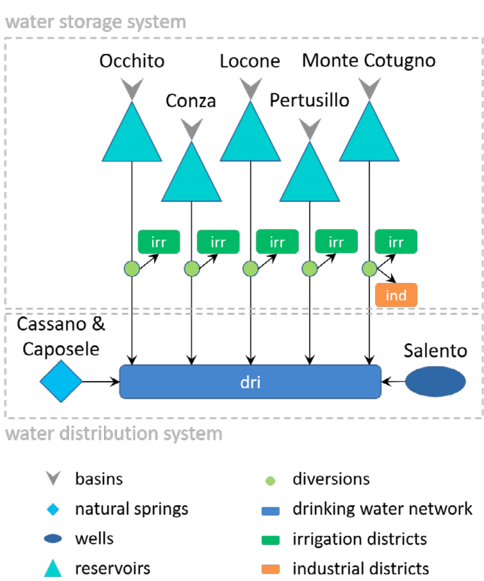

Fig. 2. Topological scheme of the strategic simulation model of the Apulia water system.

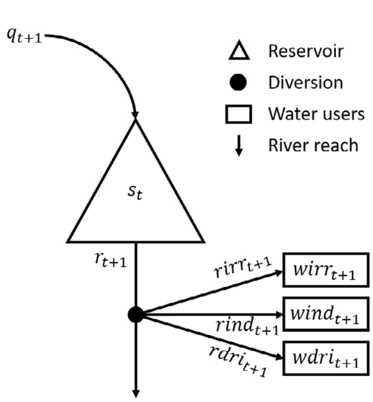

Fig. 3. An illustrative example of a single-reservoir subsystem.

$$J^{ind} = \frac{1}{h}\sum_{t=0}^{h-1}(\max(wind_t - rind_{t+1}, 0))^2 \qquad (3)$$

where $wind_t$ and $rind_{t+1}$ are the industrial water demand and supply.

- Potable water deficit, $J^{dri,d}$. Unlike the two objectives defined above, we cannot directly compute $J^{dri,d}$ as the deficit between the drinking water demand and the corresponding supply. $J^{dri,d}$ depends on how the water drawn from different sources (reservoirs, wells and natural springs) is allocated to hundreds of distribution points through the complex water network implemented in Aquator.
- Drinking water distribution cost, $J^{dri,c}$. As $J^{dri,d}$, also the distribution cost cannot be easily expressed with a single mathematical formulation but requires a specific model (again, Aquator).

In this work, we limit the analysis to operating strategies only. Planning action are not considered here but the proposed framework can be easily take them into account. The system management's optimisation requires operating the reservoir dams, the wells and the diversions. The decisions u_t is determined at each time step by a closed-loop operating policy that is optimised with respect to the four objectives.

2.2 Surrogate model identification

The drinking water distribution system managed by AQP is currently modeled through the Aquator Toolbox. In principle, Aquator could be nested in the strategic model to compute the drinking water deficit and the AQP distribution costs, the two objectives representing the aqueduct management's performance. However, this would lead to unfeasible computational costs (i.e. more than 4000 days for a 10-year horizon and a basic setting for the optimisation algorithm), making the problem practically intractable due to the high number of simulations required to find reasonable results in the optimisation procedure. A widely adopted method to solve this issue consists in building a surrogate model, i.e. a lower order, computationally faster replica of the high-resolution model (Castelletti et al., 2010a,b, 2011; Guariso and Sangiorgio, 2018, 2019, 2020a). The surrogate model is not supposed to describe physical processes with the same level of detail as the original high-fidelity model, though simply reproducing the relationship between a set of input variables (*i.e.*, the water flows released into the aqueduct and the drinking water demand) and a corresponding set of outputs (drinking water deficit and the distribution costs) in a straightforward and computationally efficient way.

The implementation of a surrogate model usually follows a 3-step procedure:

(1) Input/output samples generation via design of computer experiments (sampling of the input space) and simulation of the original high-fidelity model.
(2) Selection of the surrogate structure and optimisation of its parameters.
(3) Testing phase and evaluation of the accuracy of the surrogate by comparison with the high-fidelity simulations.

In the first phase, we generate the input/output pairs running multiple parallel simulations of Aquator. We sample the input space following two different sets of experiments. The first consists of a grid-search approach with an exhaustive evaluation of a rectangular grid. The specific values for each input are selected based on a statistical analysis of the range of values covered by the historical data. Minimum and maximum water flow into the aqueduct are selected as extreme values. In addition, we consider two intermediate values (selected looking at the density functions of the data). Using this rectangular grid, we cover the whole input hyperspace and the resulting surrogate is an interpolation model. The second set of experiments consists of a clustering procedure (performed by the mean-shift algorithm); we identify clusters of historical data and use the centroids as inputs for the surrogate identification. This improves the accuracy of the surrogate model in the input hyperspace areas more likely to occur in the future as they occurred in the past.

The second step requires selecting a suitable architecture for the reduced model and training it. Machine learning models are usually adopted for this aim for their flexibility and high performance in terms of accuracy and computing time. Here, we adopt Artificial Neural Networks (ANNs) with a feed-forward structure. The model receives as input the water flow discharged into the aqueduct (from the reservoirs, the natural springs and the wells) and the water demand pattern and computes the two AQP objectives: drinking water deficit and the associated distribution costs. The neural network is implemented in Python using Keras (Chollet et al., 2015), a high-level open-source library that provides built-in functions for building neural architectures and auto-differentiation algorithms for their training (Sangiorgio et al., 2021). The neural architecture is then rebuilt in native C++, according to the main routine for simulation and optimisation, to delete any possible source of inefficiency in the program execution. As it is common practice in machine learning applications, we split the available dataset generated by Aquator simulation into training, validation and test set. The training set is used to compute the gradients of the model's parameters (*i.e.*, weights and biases) during the optimisation. The validation set is used to monitor the risk of overfitting and set the hyperparameters' value. Once the neural surrogate model has been identified, the testing phase is performed to assess the model accuracy and to check the model's appropriateness for our task (*e.g.*, absence of under/overfitting on the training or validation datasets).

2.3 Optimisation

Direct Policy Search is a suitable method to design policies in the context of complex water systems' operations (Giuliani et al., 2016). DPS is a simulation-based optimisation that directly operates in the policy space. The operating policy p_θ is first parameterised within a given family of functions, then the algorithm explores the parameter space θ seeking for the best parameterisation with respect to the objectives of the problem, *i.e.*:

$$p_\theta^* = \arg\min_{p_\theta} J_{p_\theta}, \tag{4}$$

subject to the constraints set by the strategic model.

DPS solves the problem of parameter optimisation for a given policy structure. It can therefore find, at most, the best possible solution for the chosen class of functions. The choice of a function with limited flexibility can thus restrict the search to a subspace of policies that is not likely to contain the optimal one. Hence, it is advisable to select a very flexible class of functions to ensure the possibility of approximating the unknown optimal solution to the problem. The two most widely used nonlinear approximators in policy identification problems are ANNs and radial basis functions (RBFs) (Zoppoli et al., 2002; Sangiorgio and Guariso, 2018). Although ANNs are more popular in water management than RBFs, a comparative analysis Giuliani et al. (2014b) shows the general superiority of RBF over ANN in reservoir operation problems.

Once the policy structure is defined, a suitable optimisation algorithm has to be chosen. Due to many objectives, discreteness, stochasticity and non-derivable relations, global optimisation algorithms, such as multi-objective evolutionary algorithms (MOEAs), are more convenient than gradient-based methods in solving complex multi-objective water reservoir problems (Maier et al., 2014; Busa-Fekete et al., 2014). Specifically, we use the self-adaptive Borg MOEA Hadka and Reed (2013), which is highly robust in solving multi-objective optimal control problems (Salazar et al., 2016).

3. RESULTS

The first result is relative to the surrogate model identification. We performed a grid search on the following hyperparameters: training algorithm, non-linear activation function, batch size, learning rate and the number of neurons in the hidden layer. This is a crucial hyperparameter since it defines the surrogate model's architecture. We obtain high accuracy with a limited number of neurons. With 8 neurons, the R^2 score is higher than 0.99 and does not improve significantly when adding more hidden units, meaning that it can be considered an appropriate model complexity. We do not test deep architectures because of the high performance obtained with a single-layer neural network. This relatively simple architecture allows limiting the computational cost.

Table 1 reports the R^2 score between the outputs obtained simulating Aquator and the corresponding values (*i.e.*, computed with the same inputs) estimated by the emulator. The surrogate ensures high accuracy and generalisation capability (the same performance on the train, validation and test set proves the absence of overfitting). It can thus substitute the original aqueduct simulation model in the strategic modelling framework.

Table 1. Surrogate accuracy in terms of R^2 score for potable water deficit and AQP cost computed on training, validation and test set.

Objective	Train	Validation	Test
Potable water deficit	0.9950	0.9949	0.9947
AQP cost	0.9877	0.9875	0.9874

Once the surrogate component is identified, we can enter the traditional simulation-optimisation loop reported in Fig. 1. The 10-year simulation horizon covers the period from 01/01/2010 to 31/12/2019 at a daily time step. The core result of the analysis is the Pareto front of Fig. 4 using a parallel axes plot representation. It allows to visualise the trade-off between the four considered objectives, all to be minimised, and considering relative values between 0 (best) to 1 (worst): irrigation, industrial, drinking deficits (first, second and third vertical axes), and potable water distribution cost (fourth axis). The utopia point, whose coordinates are the best value that can be obtained for each objective individually, would be a horizontal line linking the lower value of each axis (0 in our re-scaled visualisation) in the parallel plot. As suggested by its name, this point is not practically reachable due to the trade-off between the objectives. Still, it serves as a reference point to guide the negotiation between the stakeholders to select a suitable compromise policy.

As expected, the parallel plot shows that the four objectives are highly competing. The trade-off between the three deficits (agricultural, industrial, and drinking) is due to the consumptive nature of these water uses. Due to the water scarcity regime in the region, the water volume available is not sufficient to satisfy all the water demands simultaneously, leading to a competitive situation. The trade-off is particularly critical between the two primary users regarding the amount of water demanded: the agricultural districts and the aqueduct. The management policies in green give low values for the drinking deficit but high values for the agricultural deficit (the opposite situation occurs for the policies in red). The figure shows the same pattern also for the last two axes: drinking deficit can be reduced only by increasing the economic cost related to the water distribution (including pumping, water treatment, ...) because, within certain limits, the lack of surface water can be compensated by withdrawals from the aquifer, a generally more expensive solution. As a compromise policy, we selected the one represented by the black line in Fig. 4. In order to reflect the stakeholders' concerns, the three water deficits have been prioritised over the drinking distribution cost, that is covered by the bills paid by the final users (*e.g.*, the citizens), an element that is not considered in our analysis.

4. CONCLUSIONS

We implemented the Apulian water storage and distribution system by coupling physically-based and surrogate modelling techniques and set up an EMODPS optimisation engine that allowed us to identify efficient operating policies for the system. Despite the complexity of the system and the high number of actors involved (institutions, authorities and stakeholders), the results obtained show that efficient compromise strategies exist. The proposed framework thus represents a consistent improvement with respect to the current management at least in two directions: it extends the mathematical model of the system considering industry and agriculture other than the drinking water distribution and adopts a more advanced and robust optimisation tool.

Future work will allow us to evaluate the effect of adopting innovative water-saving strategies by partially replacing the traditional sources considered in the scenario here considered with unconventional ones closing local water loops. In particular, we will assess the effectiveness of these solutions within the aqueduct, considering the reactivation of wells for emergency use, and in the irrigation districts, using refinement water, which could partially decrease the water demands. Structural improvements in the drinking water distribution system can be implemented as well. This framework thus represents a practical tool to support the aqueduct modernisation that will occur in the next years, in line with the investment plan of AQP. Technological and structural changes can also be easily evaluated considering future climate scenarios, affecting water availability and, consequently, the inflow to the reservoirs.

REFERENCES

Busa-Fekete, R., Szörényi, B., Weng, P., Cheng, W., and Hüllermeier, E. (2014). Preference-based reinforcement learning: evolutionary direct policy search using a preference-based racing algorithm. *Machine learning*, 97(3), 327–351.

Castelletti, A., Pianosi, F., Soncini-Sessa, R., and Antenucci, J. (2010a). A multiobjective response surface approach for improved water quality planning in lakes and reservoirs. *Water Resources Research*, 46(6).

Castelletti, A., Antenucci, J., Limosani, D., Quach Thi, X., and Soncini-Sessa, R. (2011). Interactive response surface approaches using computationally intensive models for multiobjective planning of lake water quality remediation. *Water Resources Research*, 47(9).

Castelletti, A., Lotov, A., and Soncini-Sessa, R. (2010b). Visualization-based multi-objective improvement of environmental decision-making using linearization of response surfaces. *Environmental Modelling & Software*, 25(12), 1552–1564.

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

Chollet, F. et al. (2015). Keras. URL `https://github.com/fchollet/keras`.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., and Reed, P.M. (2016). Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations.

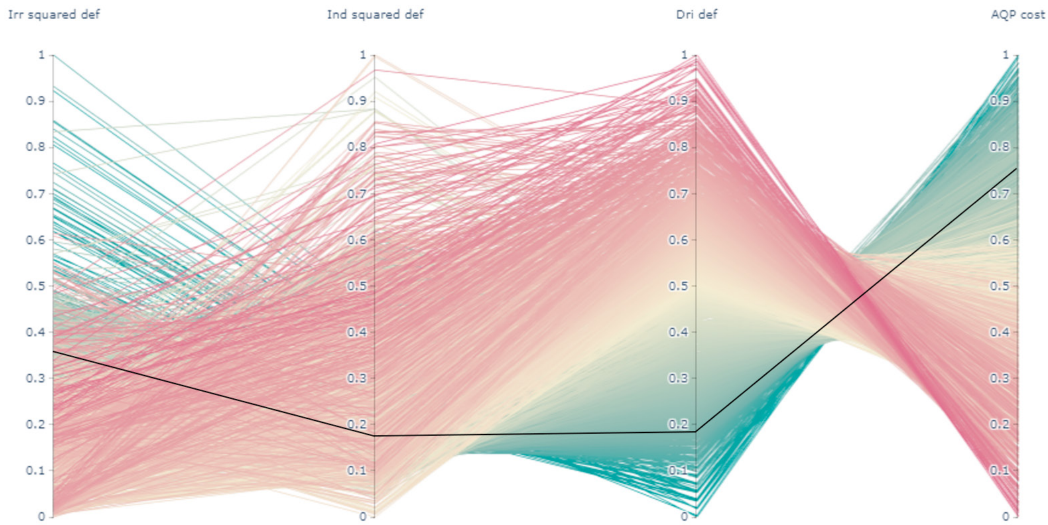

Fig. 4. Graphical representation of the trade-off between the four considered objectives. The numerical values of the objectives reported in the parallel plot are re-scaled between 0 and 1, corresponding to the extreme values (minimum and maximum) found in the optimisation process. All the objectives have to be minimised. The black line indicates the compromise policy.

Journal of Water Resources Planning and Management, 142(2), 04015050.

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

Giuliani, M., Mason, E., Castelletti, A., Pianosi, F., and Soncini-Sessa, R. (2014b). Universal approximators for direct policy search in multi-purpose water reservoir management: A comparative analysis. *IFAC Proceedings Volumes*, 47(3), 6234–6239.

Guariso, G. and Sangiorgio, M. (2020a). Valuing the cost of delayed energy actions. *IFAC-PapersOnLine*, 53(2), 16575–16580.

Guariso, G. and Sangiorgio, M. (2018). Integrating economy, energy, air pollution in building renovation plans. *IFAC-PapersOnLine*, 51(5), 102–107.

Guariso, G. and Sangiorgio, M. (2019). Multi-objective planning of building stock renovation. *Energy Policy*, 130, 101–110.

Guariso, G. and Sangiorgio, M. (2020b). Improving the performance of multiobjective genetic algorithms: An elitism-based approach. *Information*, 11(12), 587.

Guariso, G. and Sangiorgio, M. (2020c). Performance of implicit stochastic approaches to the synthesis of multireservoir operating rules. *Journal of Water Resources Planning and Management*, 146(6), 04020034.

Hadka, D. and Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary computation*, 21(2), 231–259.

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

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.

Salazar, J.Z., Reed, P.M., Herman, J.D., Giuliani, M., and Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in water resources*, 92, 172–185.

Sangiorgio, M., Dercole, F., and Guariso, G. (2021). Neural approaches for time series forecasting. In *Deep Learning in Multi-step Prediction of Chaotic Dynamics*, 43–57. Springer.

Sangiorgio, M. and Guariso, G. (2018). NN-based implicit stochastic optimization of multi-reservoir systems management. *Water*, 10(3), 303.

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.