Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Designing brackish-water desalination plants beyond the range of real data: AI methods and a physics-structured surrogate compared

Ayat S. Najjaa

ayatshawky821@gmail.com

Cairo University

Reem S. Ettouney

Cairo University

Mahmoud A. El-Rifai

Cairo University

Shadia Tewfik

National Research Centre

Marwa M. Elsayed

National Research Centre

Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11410615/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Designing brackish-water desalination plants beyond the range of real data: AI methods and a physics-structured surrogate compared

Ayat S. Najjaa[1,*], Reem S. Ettouney[2], Mahmoud A. El-Rifai[2], Shadia Tewfik[3] & Marwa M. Elsayed[4]

[1] Chemical Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt.

[2] Chemical Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt.

[3] Chemical Engineering and Pilot Plant Department, National Research Centre, El-Buhouth Street, Dokki, Giza, Egypt.

[4] National Research Centre, Giza, Egypt.

*email: ayatshawky821@gmail.com

Abstract

Brackish-water desalination plants of 100,000–600,000 $m^3\ d^{-1}$ are being planned, yet the largest operating brackish plant with published costs we found is about 68,000 $m^3\ d^{-1}$, so AI-assisted designs at this scale extrapolate. On a physics-based model with Egyptian feed-water presets, NSGA-II and multi-objective Bayesian optimisation gave equal cost–energy hypervolume, and Bayesian optimisation covered brine-pond area better than 40-member NSGA-II in all 26 runs. Reinforcement learning allocated capacity across four sites better than random but not better than a greedy rule. Data-driven surrogates failed silently beyond their training range; a hybrid that fits a capacity scaling law cut the error from 3.2% to 0.2% but not when the true scaling changed. Against few large-scale references, the model's capital cost was 51% low to 129% high; an element-by-element membrane simulation (unvalidated) gave 1–118% lower energy use, most at high recovery. Re-optimised with the simulated energy, the cheapest design needed 50–59% less evaporation-pond area. In a value-of-information analysis, once ponds were costed, none of eight uncertain cost inputs changed the best recovery, which was set by the feasible limit. In this study, model uncertainty and missing large-plant data, not the choice of algorithm, limited how far designs can be trusted.

Introduction

Brackish-groundwater desalination is expanding in water-scarce countries, and the plants being planned are growing: municipal facilities of 100,000–300,000 $m^3\ d^{-1}$ and regional schemes that serve several sites. Designing them involves decisions that single-objective, single-site tools handle poorly. Cost, energy use and the land needed for brine disposal compete with each other; capacities lie beyond the range in which simpler tools were built and checked; and meeting a regional demand means committing capacity to several sites in turn. We have previously reviewed the techno-economics of brackish-water desalting[1] and developed a physics-based design framework for Egyptian brackish groundwater and agricultural drainage water[2]; here we use the framework as common ground truth for comparing artificial-intelligence (AI) methods.

Brackish water is attractive because its salinity, typically about 1–10 $g\ L^{-1}$, is a fraction of that of seawater, so it can be desalted by reverse osmosis at lower pressure and energy. Of about 95 million $m^3\ d^{-1}$ of desalination capacity worldwide, almost half is in the Middle East and North Africa[3], and inland brackish aquifers are being assessed as a water resource well beyond that region[4]. Inland plants cannot discharge brine to the sea. Evaporation ponds are usually the cheapest disposal option where land is cheap and evaporation is high[5], so land becomes a design objective in its own right. Cost is the other driver. Unit capital cost is expected to fall as plants get larger[6], and capacity is the strongest single predictor of the capital cost of reverse-osmosis plants[7], but the size of this economy of scale is poorly constrained, and costs also fall over time as the industry gains experience[8]. A designer of a 100,000–600,000 $m^3\ d^{-1}$ brackish plant therefore has to optimise a model whose cost-scaling law has not been observed at that scale.

AI has been applied to reverse-osmosis (RO) design along four lines. Machine-learning surrogates of RO performance now commonly outperform classical response-surface models[9–11]. Multi-objective metaheuristics, such as NSGA-II[12], have been coupled with learned surrogates: a neural network with NSGA-II mapped energy–recovery trade-offs in brackish RO[13], particle-swarm search with surrogates was used for seawater RO design[14], and NSGA-II, particle-swarm and rabbit-optimiser variants sized renewable-powered seawater RO systems at nine sites in Sinai, Egypt[15]. Reinforcement learning[16] has been applied to RO control[17] and to the staged design of water-distribution networks[18]. Physics-informed neural networks, which embed governing equations in training, are an emerging alternative[19] and have been tested on seawater RO plant data[20].

A known weakness of data-driven surrogates is extrapolation. Tree ensembles return a constant beyond the range of their training data, and neural networks extrapolate in ways that depend on their architecture and initialisation; methods that estimate where a model can be trusted, its area of applicability, have therefore been proposed[21]. Process engineers have long addressed the same problem with hybrid, or grey-box, models that combine a mechanistic part with a data-driven part[22,23], and the idea has since been generalised as theory-guided data science[24] and physics-informed machine learning[25]. Such hybrids are expected to extrapolate better because the mechanistic part carries the known structure. In desalination design, however, the structure that matters most at large scale, the cost-scaling law, is itself uncertain, and it has not been tested how far a hybrid can be trusted when that structure is wrong.

Most of these studies apply one family of methods to one problem. What is missing is a like-for-like comparison of methods on the same model, at matched computational budgets, with repeated runs and statistics, for the problem classes that matter at large scale. Equally, an optimisation result is only as credible as the model being optimised, so checking that model against independent data belongs inside the comparison, not after it. A further gap is specific to scale. We found no operating brackish-water plant with published costs larger than about 68,000 $m^3\ d^{-1}$ (see Results), so both the model being optimised and any surrogate trained on it must extrapolate, and it is not known whether the algorithm, the surrogate or the model is the weaker link.

Here we compare four methods on three problems, using one physics-based model with Egyptian aquifer presets. First, a three-objective design of a single 100,000–300,000 $m^3\ d^{-1}$ plant (levelised cost of water, LCOW; specific energy consumption, SEC; brine-evaporation-pond area) with NSGA-II and multi-objective Bayesian optimisation (MOBO). Second, surrogate models for 100,000–600,000 $m^3\ d^{-1}$ plants, including a test of what happens when a surrogate is used outside its training range and a physics-structured hybrid surrogate built to extrapolate. Third, allocation of capacity across four sites with reinforcement learning, benchmarked against greedy and random policies. We then check the model against independent data: a global energy–salinity curve[26], a regression on 36 real plants[27], a published 300,000 $m^3\ d^{-1}$ design study[28] and two small seawater plants[29,30]. The study makes three contributions: a matched-budget, multi-seed comparison of the methods on large-scale problems; a measurement of silent surrogate failure outside the training range, together with a hybrid surrogate that combines a fitted capacity scaling law with a learned residual and is stress-tested against alternative large-scale cost scaling; and a check of the model against all the independent large-scale data we could find and against an element-by-element membrane simulation. Together they show where the methods tie, where one is better, and that, for this model and these problems, the uncertainty in the model mattered more than the choice of algorithm.

Methods

Physics-based design model

We used the framework described previously[2] (Python; released with this paper, see Code availability). For a given feed salinity, recovery r and permeate capacity Q_p it computes intake, feed and concentrate flows, sizes the treatment train, and estimates feed pressure from osmotic pressure, a net driving pressure and a friction allowance. Hydraulic power and SEC are

$$P_{hyd} = 0.02778 \cdot Q \cdot \Delta p \,/\, \eta_{pump}, \quad SEC = 24 \cdot P_{net} \,/\, Q_p \quad (\mathrm{kWh\ m^{-3}}) \quad (1)$$

with Q the feed flow ($m^3\ h^{-1}$), Δp the feed pressure (bar), $\eta_{pump} = 0.80$ and P_{net} the power after energy recovery. Capital and operating costs are scaled from benchmark values calibrated against an 18-plant set of operating and planned plants[2], and LCOW is annualised capital plus operating cost divided by annual output, with a capital recovery factor of 0.1175. Evaporation-pond area follows from concentrate flow and salinity and a net evaporation rate of 2.2 m yr^{-1}. The presets are illustrative composites of published hydrochemistry (Supplementary Table S1). Default recoveries follow the framework's rule (0.75–0.85 for brackish presets, 0.45 for seawater). The optional groundwater-lift term (Supplementary Section S1) was not used because pumping heads were not supplied.

Optimisation problems

Problem 1: minimise F(x) = [LCOW, SEC, pond area] over x = (recovery, capacity), with recovery in 0.50–0.90 (brackish) and capacity in 100,000–300,000 $m^3\ d^{-1}$. Problem 2: regression of LCOW on (feed salinity, recovery, capacity, water type). Problem 3: choose, increment by increment, which site receives the next 10,000 $m^3\ d^{-1}$, with at most 15 increments, subject to each site's capacity ceiling and pond budget (Supplementary Table S2), minimising the capacity-weighted LCOW of the resulting system.

NSGA-II

We implemented NSGA-II[12] with fast non-dominated sorting, crowding distance, binary tournament selection, simulated binary crossover (distribution index 15, probability 0.9) and polynomial mutation (index 20, probability 1/n). The default was a population of 40 for 25 generations (1,040 evaluations); populations of 20, 80 and 100 with 50, 12 and 10 generations were used for sensitivity. As a check we ran it on ZDT1 and ZDT2[31] (30 variables, population 100, 100 generations) and computed generational distance (mean distance from each found point to the true front) and inverted generational distance (mean distance from each true-front point to the found set).

Multi-objective Bayesian optimisation

For each of ten weight vectors w drawn from a uniform Dirichlet distribution, we minimised the Chebyshev scalarisation

$$g(x \mid w) = \max_k \; w_k \cdot \left| f_k(x) - z^*_k \right| \quad (2)$$

where z* is the ideal point estimated from eight warm-up evaluations. Each search used a Gaussian-process surrogate with a Matérn-5/2 kernel[32] and the expected-improvement acquisition function[33],

$$EI(x) = (y_{best} - \mu(x) - \xi) \cdot \Phi(z) + \sigma(x) \cdot \varphi(z), \quad z = (y_{best} - \mu(x) - \xi) \,/\, \sigma(x) \quad (3)$$

maximised over 1,500 random candidates per iteration ($\xi = 0.01$). The non-dominated set of all evaluated points formed the MOBO front. The budget was matched to NSGA-II (1,048 versus 1,040 evaluations).

Hypervolume

Hypervolume[34] was computed against a reference point 1.05 times the largest value of each objective over the fronts being compared. We report the two-objective value (LCOW, SEC) and the exact three-

objective value (LCOW, SEC, pond area), obtained by summing two-dimensional hypervolumes over slabs along the third objective. Fronts were first filtered for non-dominance in all three objectives, so the two-objective value ignores pond area only in the measure, not in the filtering.

Surrogate models

Training sets were drawn at random over feed salinity (500–15,000 mg L^{-1} brackish; 28,000–45,000 mg L^{-1} seawater), recovery (0.50–0.90 brackish; 0.35–0.55 seawater), capacity and water type (30% seawater). The small-range set had 6,000 plants with capacity log-uniform over 2,000–150,000 $m^3\ d^{-1}$; the large-scale set had 4,000 plants with capacity uniform over 100,000–600,000 $m^3\ d^{-1}$. Models were implemented in scikit-learn[35]: random forest, gradient boosting (300 trees, depth 3, learning rate 0.05) and a multilayer perceptron (two hidden layers of 64 and 32 units, or 128, 64 and 32 for the large-scale set; ReLU, Adam, standardised inputs, early stopping). We report 5-fold cross-validated and held-out (20%) R^2, MAE and root-mean-square error. The extrapolation test compared the two gradient-boosting models with the physics-based model for a 5 g L^{-1} brackish feed at 80% recovery from 100,000 to 600,000 $m^3\ d^{-1}$.

Hybrid surrogate and scaling stress test

The hybrid first fits the prior LCOW_prior = a + a_sw s + (b + b_sw s)(Q/Q_ref)^c, where s is 1 for seawater and 0 otherwise and Q_ref = 20,000 $m^3\ d^{-1}$, to the training data by non-linear least squares (SciPy curve_fit), and then fits gradient boosting (300 trees, depth 3, learning rate 0.05) to the residual, using feed salinity, recovery, $\log_{10}$ capacity and water type. Its prediction is the sum of the two. This is the serial form of hybrid semi-parametric modelling, in which the data-driven part corrects a mechanistic prior[22,23]. The comparison models were gradient boosting on raw inputs and on $\log_{10}$ capacity, histogram gradient boosting with a monotone-decreasing constraint on capacity, and a multilayer perceptron (64 and 32 units) on standardised inputs. Each of five seeds drew 6,000 training plants (capacity log-uniform over 2,000–150,000 $m^3\ d^{-1}$), 1,000 in-range test plants and 2,000 test plants with capacity uniform over 150,000–600,000 $m^3\ d^{-1}$; the other inputs were sampled as above, and targets were unrounded LCOW. For the stress test, the model's capital-cost scale factor (Q/Q_ref)^−0.09 was kept below 150,000 $m^3\ d^{-1}$ and continued above it as (150,000/Q_ref)^−0.09 (Q/150,000)^e, with e = −0.20 or 0, so the training data were identical in all three scenarios. Both alternatives are plausible because the framework's exponent was fitted to plant data with a low coefficient of determination ($R^2 = 0.09$)[2].

Element-by-element RO simulation

Each 8-inch spiral-wound element (37 m^2) was solved in series along seven-element pressure vessels. The water flux was $J_w = A \cdot TCF \cdot FF \cdot (\bar{P} - \Delta\pi)$, where $\bar{P}$ is the feed-side pressure at mid-element and $\Delta\pi = \beta \cdot \pi(\bar{C}) - \pi(Cp)$; the salt flux was $J_s = B \cdot TCF \cdot (\beta \cdot \bar{C} - Cp)$, with $Cp = J_s/J_w$, $\bar{C}$ the mean of the feed and concentrate concentrations, and π = 0.0008 bar per mg L^{-1} of TDS, the model's own coefficient, so that the two calculations differ only in process structure. The concentration-polarisation factor $\beta = \exp(0.7r)$, with r the element recovery, the temperature correction factor $TCF = \exp[2640(1/298.15 - 1/T)]$ and the element pressure drop $\Delta P = 0.01 q^{1.7}$ (psi, with q the mean feed–concentrate flow in gpm) are the supplier's design equations[36]. A = 3.37 L $m^{-2}\ h^{-1}\ bar^{-1}$ and B = 0.187 L $m^{-2}\ h^{-1}$ were back-calculated from the datasheet test point of a standard brackish-water element (40 $m^3\ d^{-1}$ of permeate and 99.5% rejection at 2,000 mg L^{-1} NaCl, 15.5 bar, 25 °C and 15% recovery)[37], and a fouling factor FF = 0.85 was applied to A. For each recovery the simulation used the lowest-energy array among one stage (recovery up to 55%), two stages (2:1 vessels, 50–85%) and three stages (4:2:1, 70% and above), preferring arrays within the element's pressure rating; all had an average flux of 22 L $m^{-2}\ h^{-1}$ as in the model, and the presets' feed temperatures. The feed pressure was solved so that the first stage produced its share of the permeate, and each later stage received an interstage booster only when needed for its share; a variant without boosters solved the feed pressure

for the total recovery alone. SEC was the hydraulic work of all pumps divided by the permeate volume, with a pump efficiency of 0.80 and no energy recovery, as in the model. Varying the fouling factor from 0.70 to 1.0 or the pressure-drop correlation by ±50% changed SEC by less than 8%. Designs were flagged when any element pressure exceeded the element's 41-bar rating[37].

Re-optimisation with the simulated energy

Problem 1 was solved again with NSGA-II (population 80, 12 generations, about 1,040 evaluations; seeds 42, 1 and 2) for the five brackish presets, once with the model's lumped SEC and once with the simulated SEC at each recovery, everything else unchanged. Designs whose simulated pressure exceeded the 41-bar element rating were penalised. From each front we report the cheapest design and the smallest pond area reachable within 1% of the cheapest LCOW.

Value of information

The decision was the design recovery (0.50–0.90 in steps of 0.01) of a 200,000 $m^3\ d^{-1}$ plant with evaporation ponds, and its cost was either the LCOW alone or the LCOW plus the annualised capital cost of the ponds. Eight inputs were uncertain, with independent uniform distributions: the capacity exponent of unit capital cost (−0.20 to 0); a multiplier on the capital-cost benchmark (0.6–1.4); a weight w on the model's lumped SEC against the simulated SEC, SEC = (1 − w)·SEC_sim + w·SEC_model, with w from 0 to 1; the electricity tariff (0.03–0.10 USD kWh^{-1}); the non-energy operating cost (0.30–0.50 USD m^{-3}); the net evaporation rate (1.5–3.0 m yr^{-1}); the capital cost of lined ponds (5–30 USD m^{-2}, bracketing roughly 12–17 USD m^{-2} for a lined pond built to prescriptive rules[38]); and the discount rate (8–14%). For 20,000 Monte Carlo samples we found the cost-optimal recovery and the minimum cost. The expected value of perfect information (EVPI) is the expected cost of the best fixed recovery minus the expected minimum cost. The expected value of partial perfect information (EVPPI) of each input was estimated by grouping the samples into 40 quantile bins of that input and taking the best recovery within each bin[39]. First-order sensitivity indices, the share of an output's variance explained by one input[40], were estimated with the same bins, and a dummy input with no effect gave the noise floor.

Reinforcement learning

The state was each site's allocation as a fraction of its ceiling plus the remaining-demand fraction. The policy had one hidden layer of 16 tanh units and a softmax over the four sites, with infeasible sites masked. It was trained with REINFORCE (analytic policy gradient, moving-average baseline with factor 0.9, learning rate 0.08, 600 episodes) on the reward −LCOW − 2×10^{-5} × (unmet demand in $m^3\ d^{-1}$):

$$\nabla_\theta J = E[\ (R - b) \cdot \Sigma_t \nabla_\theta \log \pi_\theta(a_t \mid s_t)\ ] \quad (4)$$

The greedy policy chose the feasible site with the lowest marginal LCOW; the random policy chose uniformly among feasible sites. Recovery was fixed at each preset's default. We used 20 training seeds, and one random episode per seed. In the coupled version, the sum of the four sites' required pond areas was limited to a fraction f of the sum of their budgets, infeasible actions were masked, the state gained the fraction of the shared budget used, and f was set to 1, 0.55, 0.50, 0.49 and 0.48.

Verification of the NSGA-II and REINFORCE implementations

Because NSGA-II and REINFORCE were implemented from scratch, we checked NSGA-II on the ZDT1 and ZDT2 benchmarks[31], whose Pareto fronts are known analytically. It reached generational distances of 0.0031 and 0.0024 (inverted distances 0.0047 and 0.0048), below the 0.01 usually taken as converged (Supplementary Fig. S3). Among the REINFORCE settings we tried (one seed), 600 episodes at a learning rate of 0.08 performed best (LCOW 0.7287, against 0.7378 for 150 episodes, 0.7392 for a learning rate of 0.02 and 0.7313 for 0.15; Supplementary Table S9).

Statistics

Paired per-seed comparisons used the Wilcoxon signed-rank test[41] in SciPy[42]. MOBO took about 100 s per run in our environment, so the two-objective hypervolume used ten seeds (El Moghra), and the three-objective hypervolume ten seeds for El Moghra and four for each of four other presets (Nile Delta shallow, Western Desert, Sinai coastal, Nile Delta drainage); the Wilcoxon test was applied to the 26 pooled ratios, which treats runs as independent although runs share presets. All tests were two-sided with the exact p value reported (for example, 3×10^{-8} is $2/2^{26}$, the exact two-sided value when all 26 ratios fall on the same side). Values written as mean ± s are sample standard deviations across seeds (n = 20 training seeds for the reinforcement-learning results).

Independent data

Rosa et al.[26]: we used the source-data file (sheet "Figure 1": salinity and energy, 51 points, 2–52 ppt; 1 ppt taken as 1 g L^{-1}). Its cost column equals energy multiplied by a constant 0.0843 USD kWh^{-1} (checked to six decimals), so it is not a full LCOW and was not compared with ours. Below 2 g L^{-1} the curve was extrapolated linearly and flagged. Stillwell & Webber[27]: their municipal-scale regression of SEC on year of operation (YR), raw-water TDS (c_{rw}) and product-water TDS (c_{pw}) for 36 plants ($R^2 = 0.86$) is printed as $SEC = 260 - 0.13\ YR + 8.3 \times 10^{-5}\ c_{rw} - 2.4 \times 10^{-3}\ c_{pw}$, with coefficients "rounded to two significant figures". The year term (0.13 × about 2,000) is then uncertain by about ±10 kWh m^{-3} through rounding alone, and the printed equation does not reproduce the authors' own Carlsbad example (3.5 kWh m^{-3}): it gives 0.1. We therefore used only the partial slope for raw-water TDS, 8.3×10^{-5} kWh m^{-3} per mg L^{-1} (0.083 per g L^{-1}). The 300,000 $m^3\ d^{-1}$ design study[28] is a conceptual system sized and costed with design software. We computed its RO-train SEC from the pump powers in its Table 6 (12,622, 626 and 2,502 kW) and its RO permeate (270,000 + 18,000 $m^3\ d^{-1}$), its whole-system electricity as 1.47×10^8 kWh y^{-1} (its Table 8) over 300,000 $m^3\ d^{-1}$ × 365 d × 0.9, or equivalently from its listed powers (18,188 kW) over 294,000 $m^3\ d^{-1}$ of product, and its costs from its Tables 7–9 (capital cost 116,405 thousand USD; operating cost 47,955 thousand USD y^{-1}). Recoveries are the design values (90% and 60% in the two RO stages). The Hurghada values are from two vendor case studies[29,30]; the 2.2–2.5 kWh m^{-3} figure is the operator's reported average, and neither source is an independent measurement.

Scale test on operating plants: the model's non-energy operating cost constant was set to the median over the eight Texas plants of 25,000 $m^3\ d^{-1}$ or less with a reported tariff, and capital and operating cost power laws were fitted to the small plants by least squares in log–log space. Predictions were made for the five operating plants of 40,000 $m^3\ d^{-1}$ or more (Florida tariff assumed 0.08 USD kWh^{-1}; 0.06–0.125 shifted the predicted operating cost by −0.01 to +0.03 USD m^{-3}). Only one large plant has a capital cost, so that comparison is a single case. Capital costs are in nominal USD of different years across sources.

Results and discussion

The physics-based model sizes pretreatment, membrane stages, post-treatment and brine management and returns SEC, capital and operating cost, LCOW and pond area for a given feed salinity, recovery and capacity (Methods). We use five Egyptian brackish presets (feed salinity 1.8–8 g L^{-1}) and a Red Sea seawater preset (Supplementary Table S1). Unless stated, SEC is the membrane-train value and excludes groundwater lift.

NSGA-II and MOBO find the same cost–energy trade-off but differ in coverage of pond area

We optimised recovery and capacity (100,000–300,000 $m^3\ d^{-1}$) for the El Moghra preset (5 g L^{-1}) against LCOW, SEC and pond area, using NSGA-II (population 40, 25 generations, 1,040 model evaluations) and MOBO (ten Chebyshev-scalarised Bayesian searches, 1,048 evaluations). Both

reached LCOW of 0.689–0.751 USD m^{-3} and SEC of 1.00–1.83 kWh m^{-3} (Fig. 1a). In the LCOW–SEC plane the hypervolume[34] of the two fronts was equal to within 10^{-5} in all ten seeds (0.091683 ± 0.000165 for both; paired Wilcoxon signed-rank test[41], p = 1.00; Supplementary Table S3). In the model, land rather than cost is the main trade-off: along a front on which LCOW varies by only 0.062 USD m^{-3} (9%), pond area spans 185–2,563 ha. The hypervolume is the volume of objective space that a front dominates up to a reference point; it rewards both closeness to the true front and spread along it, and is the standard single-number measure for comparing multi-objective optimisers[43].

The methods differ in how they describe the front. NSGA-II returns 40 points by construction and MOBO 269, many of which are non-dominated only because they differ in pond area. The difference follows from how the methods work. NSGA-II keeps a population of fixed size and, each generation, discards solutions that are dominated or crowded. MOBO, like ParEGO[44], turns the objectives into a single one with a different weight vector in each search and keeps every point it evaluates, so its archive fills in parts of the front, here large pond areas, that a small population thins out. When pond area is included, the three-objective hypervolume of NSGA-II with a population of 40 was smaller than that of MOBO in every one of 26 runs (ten seeds for El Moghra and four seeds for each of four other presets; median ratio 0.950, range 0.917–0.985; Wilcoxon signed-rank $p = 3 \times 10^{-8}$; Fig. 1b; Supplementary Tables S4 and S11). The gap depended on the preset, with a median ratio of 0.923 for the shallow Nile Delta aquifer and 0.978 for the Sinai coastal aquifer. Population size appears to account for most of it: at about the same budget (1,040–1,100 evaluations), a population of 100 reached a median of 0.985 of MOBO's value (range 0.970–0.997), which is closer but still below 1 in all 26 runs. In the single-seed El Moghra test, populations of 80 and 100 reached 0.992 and 0.993, whereas a population of 20 reached 0.928. MOBO's front extended to larger pond areas (up to 2,724 versus 2,563 ha). One NSGA-II run took about 2 s and one MOBO run about 100 s in our environment, because the Gaussian-process fits dominate MOBO's cost.

NSGA-II and MOBO found the same cost–energy trade-off, and their difference lay in the third objective: MOBO's archive-based front covered pond area more fully than a 40-member NSGA-II population in all 26 runs, and a population of 80–100 removed most, but not all, of that gap (median 1.5%). In practice, NSGA-II with a population of at least 80 is the cheaper default here. Our problem has only two decision variables and smooth objectives, so it does not test the regime in which Bayesian optimisation is expected to help most, namely many variables and expensive evaluations. Bayesian optimisation was developed for that regime, where each evaluation takes minutes to hours and the overhead of fitting the surrogate is negligible[45]; coupling the framework to element-by-element membrane simulation, or adding membrane type, staging and pretreatment as decision variables, would create it.

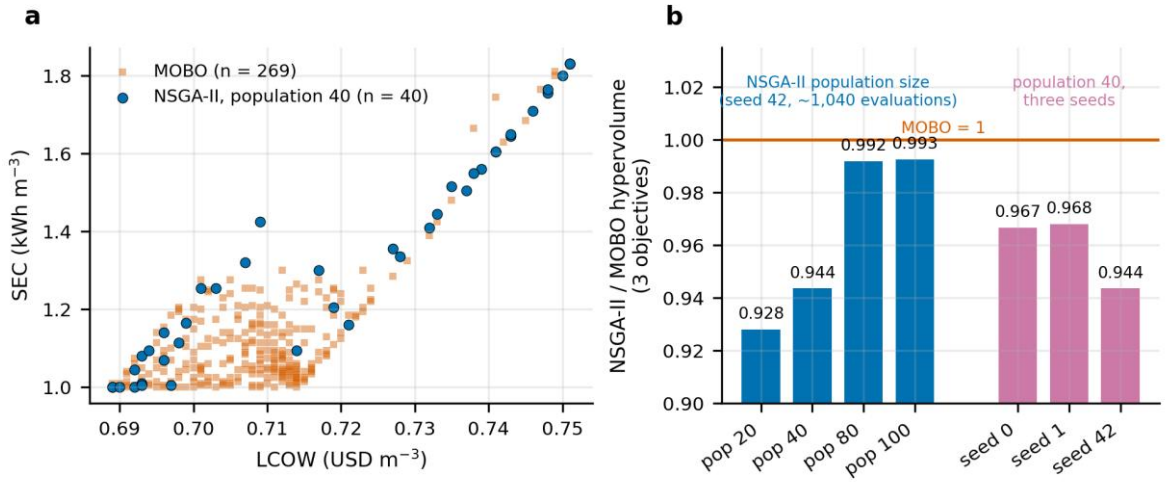


Figure 1 | Multi-objective design of a 100,000–300,000 m³ d⁻¹ plant (El Moghra preset). a, Points in the LCOW–SEC plane found by NSGA-II (population 40, 1,040 evaluations; blue) and MOBO (1,048 evaluations; orange). MOBO points that appear dominated differ in pond area. b, Ratio of NSGA-II to MOBO hypervolume in the three-objective space (LCOW, SEC, pond area) for four NSGA-II population sizes at about 1,040 evaluations (seed 42, blue) and for population 40 in the first three El Moghra seeds (purple; all 26 runs are in Supplementary Table S11). A ratio below 1 means MOBO covers more of the objective space.

A surrogate trained on a smaller range fails silently outside it

Surrogates are used to speed up design-space search, but tree-based models cannot extrapolate. A regression tree predicts a constant in each leaf, and the leaf that holds the largest training capacities also holds every larger capacity, so beyond the training range the prediction stops changing. We trained gradient-boosting regressors[46] for LCOW as a function of feed salinity, recovery, capacity and water type, either on 6,000 plants of 2,000–150,000 m³ d⁻¹ or on 4,000 plants of 100,000–600,000 m³ d⁻¹. Within its range each was accurate. On held-out data, gradient boosting reached R^2 = 0.9992 (mean absolute error, MAE, 0.0017 USD m⁻³) and a deeper neural network 0.9990 (MAE 0.0020) for the large-scale set; for the small-range set, random forest[47], gradient boosting and a neural network gave R^2 = 0.9990, 0.9991 and 0.9978 (Supplementary Table S5).

Beyond its training range the small-range surrogate gave no sign of failure. It returned a constant 0.714 USD m⁻³ for every capacity above 150,000 m³ d⁻¹, while the physics-based LCOW kept falling, from 0.712 at 150,000 m³ d⁻¹ to 0.682 at 600,000 (Fig. 2a). Its error therefore grew from 0.3% at 150,000 m³ d⁻¹ to 2.5%, 3.8% and 4.8% at 300,000, 450,000 and 600,000 m³ d⁻¹ (Fig. 2b). The surrogate trained on 100,000–600,000 m³ d⁻¹ stayed within 0.4% above 150,000 m³ d⁻¹ and within 0.7% over the whole range. A purely data-driven surrogate is therefore valid only for the range it was trained on. We next asked whether building engineering structure into the surrogate removes this limit.

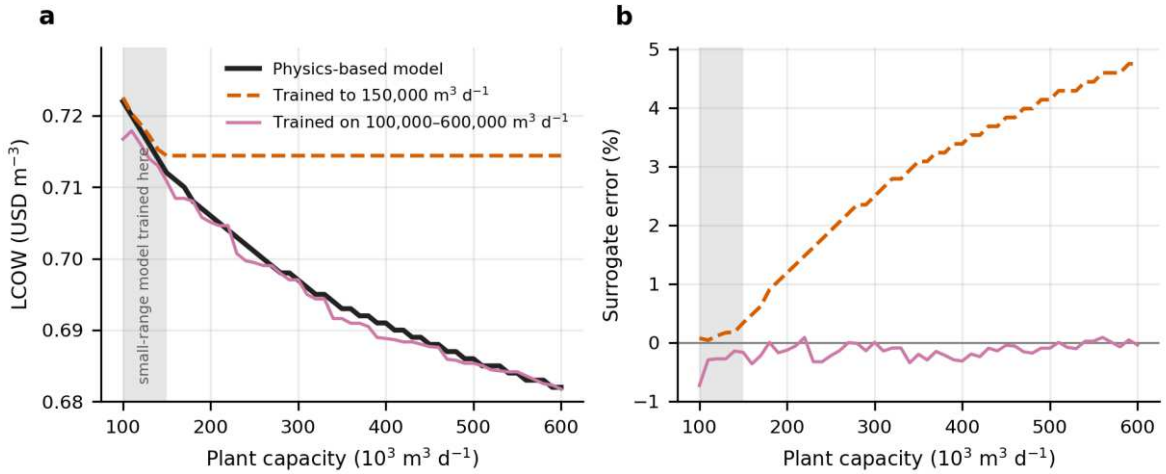


Figure 2 | A surrogate trained on a smaller capacity range fails silently outside it. a, LCOW from the physics-based model (black) and from gradient-boosting surrogates trained on 2,000–150,000 m³ d⁻¹ (orange, dashed) or on 100,000–600,000 m³ d⁻¹ (purple), for a 5 g L⁻¹ brackish feed at 80% recovery. The grey band marks the part of the plotted range on which the small-range surrogate was trained. b, Surrogate error relative to the physics-based model.

A hybrid surrogate extrapolates only as far as its scaling law holds

Capital cost per unit of capacity is commonly described by a power law in capacity. We therefore built a hybrid surrogate that fits LCOW = a + b(Q/Q_ref)^c, with separate a and b for seawater, to the small-range training data and trains gradient boosting only on the residual (Methods). The exponent c is fitted, not copied from the model: over five training sets it was −0.082 ± 0.022, against −0.09 in the model. We compared the hybrid with gradient boosting on raw inputs, on log capacity and with a monotone constraint on capacity, and with a neural network, all trained on 6,000 plants of 2,000–150,000 m³ d⁻¹ and tested on 2,000 plants of 150,000–600,000 m³ d⁻¹ of all feed types (five seeds; Table 1).

Within the training range all surrogates were accurate (mean absolute percentage error 0.12% for the hybrid and 0.25–0.37% for the others). Beyond it, the three gradient-boosting variants all extrapolated flat, with a mean error of 3.2% (4.5% at 450,000–600,000 m³ d⁻¹). Neither log-transformed capacity nor a monotone constraint helped, because a tree ensemble cannot predict below the smallest value it has seen. The neural network gave 3.1 ± 1.1%, with a shape that changed from seed to seed (Fig. 3a). The hybrid gave 0.20 ± 0.10% (95th percentile 0.5%), a 16-fold reduction.

This result holds only if the scaling law continues beyond the training range. Because no plant above 150,000 m³ d⁻¹ was in the training data, we tested two alternative ‘truths’ that are identical to the model below 150,000 m³ d⁻¹ and differ above it: steeper economies of scale (exponent −0.20) and none at all (exponent 0, as for a plant built from identical modular trains). Both are compatible with the real data, which constrain the exponent only weakly (Methods). With steeper scaling the hybrid’s error was 3.2%, half that of gradient boosting (6.6%) and the neural network (6.0%). With no further economies of scale it was 2.8%, whereas the neural network was 2.4% and gradient boosting, whose flat extrapolation happens to match this case, 0.4% (Fig. 3b). No surrogate trained on small plants can tell these cases apart, because their training data are identical. Building engineering structure into the surrogate therefore removes the extrapolation error caused by the learning method, but not the error caused by not knowing how costs scale; only data from large plants can remove that.

The surrogate experiment turns a familiar warning into a measurement. A model that was within 0.3% at the upper edge of its training range was 4.8% wrong at four times that capacity, with no indication in its output. Building a fitted capacity scaling law into the surrogate removed 94% of the mean extrapolation error (3.2% to 0.2%) when the scaling held, a simpler route than physics-informed

networks[19], which we did not test. The stress test shows the limit of both: a structured surrogate is right beyond its training range only if costs keep scaling there as they did within it, and no surrogate trained on small plants can know whether they do.

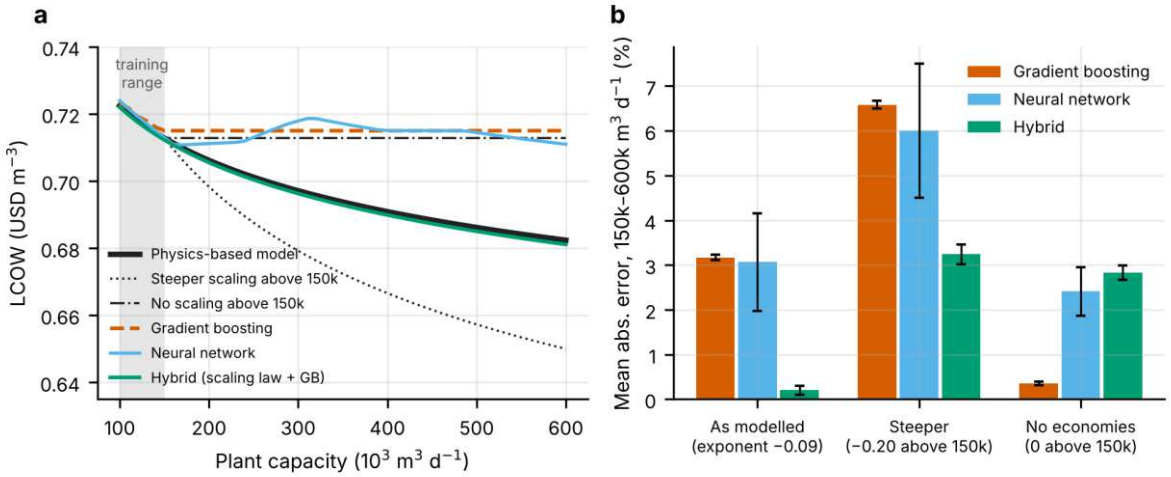


Figure 3 | A physics-structured hybrid surrogate extrapolates only as far as its scaling law holds. a, LCOW for a 5 g L^{-1} brackish feed at 80% recovery from the physics-based model (thick black) and from surrogates trained on 2,000–150,000 m^3 d^{-1} (seed 42): gradient boosting (orange, dashed), neural network (light blue) and hybrid (green). Thin black lines show the two alternative scalings above 150,000 m^3 d^{-1}. b, Mean absolute percentage error over 2,000 test plants of 150,000–600,000 m^3 d^{-1} of all feed types, for the modelled scaling and the two alternatives; bars show means and error bars standard deviations over five training seeds.

Table 1 | Surrogate error within and beyond the training range. Mean absolute percentage error against the physics-based model (mean ± s.d. over five training seeds) for surrogates trained on 2,000–150,000 m^3 d^{-1}, tested on 1,000 plants within that range and on 2,000 plants of 150,000–600,000 m^3 d^{-1} of all feed types. Beyond the training range the true LCOW follows the modelled capacity scaling (exponent −0.09) or one of two alternatives above 150,000 m^3 d^{-1}.

Surrogate	Within range (%)	Beyond, scaling as modelled (%)	Beyond, steeper scaling (%)	Beyond, no further economies (%)
Gradient boosting, raw inputs	0.25 ± 0.01	3.17 ± 0.06	6.58 ± 0.09	0.36 ± 0.04
Gradient boosting, log capacity	0.25 ± 0.01	3.17 ± 0.06	6.58 ± 0.09	0.36 ± 0.04
Gradient boosting, monotone in capacity	0.26 ± 0.00	3.16 ± 0.07	6.57 ± 0.08	0.36 ± 0.05
Neural network (64–32)	0.37 ± 0.09	3.06 ± 1.10	6.00 ± 1.50	2.41 ± 0.54
Hybrid: scaling law + gradient-boosting residual	0.12 ± 0.00	0.20 ± 0.10	3.24 ± 0.22	2.83 ± 0.16

Reinforcement learning allocates capacity well but does not beat a greedy rule

We then asked how to build out 150,000 m^3 d^{-1} of a 250,000 m^3 d^{-1} regional demand in 10,000 m^3 d^{-1} increments across four sites (El Moghra, Nile Delta, Western Desert, Sinai). Each site has an aquifer-yield ceiling (60,000–120,000 m^3 d^{-1}) and a brine-pond land budget (165–350 ha); both are illustrative (Supplementary Table S2). Together the sites can supply 230,000 m^3 d^{-1}, so it is the build-out budget that forces choices, and every policy leaves 100,000 m^3 d^{-1} of demand unmet by design. We trained a REINFORCE policy[16] for 600 episodes and compared it with a greedy policy (always add the increment with the lowest marginal LCOW) and a random feasible policy. REINFORCE is a policy-gradient method: the agent picks sites at random according to a learned policy, and after each complete build-out it makes the choices that led to a lower system cost more likely[48].

Across 20 training seeds, the REINFORCE policy gave a system LCOW of 0.7305 ± 0.0049 USD m^{-3} (range 0.7245–0.7393), against 0.7385 ± 0.0026 for random allocation (Wilcoxon signed-rank $p = 1 \times$

10^{-5}; REINFORCE better in 18 of 20 seeds) and 0.7245 for the deterministic greedy policy (Fig. 4a; Supplementary Table S6). On average REINFORCE therefore ended 0.8% above greedy, and matched it in the best seed. The learned allocation differed from the greedy one (Sinai rather than El Moghra; Fig. 4b) at almost the same cost, so several allocations are nearly equally good. Greedy is expected to do well here because each site's marginal cost rises independently of the others. For allocation problems with separable, convex costs, adding each increment where its marginal cost is lowest is provably optimal[49], so here the greedy policy is close to the best possible and is a demanding benchmark rather than a weak one. We therefore tested a coupled version of the problem.

In the coupled version the sites share a regional brine-land budget equal to a fraction f of the four site budgets combined (925 ha at f = 1), and the policy's state includes the fraction used. We scored policies by reward, which is −LCOW − 2 × 10^{-5} × unmet demand (the 100,000 m³ d⁻¹ left unmet by design contributes −2). At f = 1 the result was unchanged (REINFORCE −2.731 ± 0.005, greedy −2.724). As f fell, funding the cheapest site first began to use up land that later increments needed. At f = 0.55 (509 ha) greedy still built the full 150,000 m³ d⁻¹ (−2.724), whereas REINFORCE left 6,500 m³ d⁻¹ unbuilt on average (−2.857 ± 0.202). At f = 0.50 (463 ha) greedy left 10,000 m³ d⁻¹ unbuilt and REINFORCE 18,500 (rewards −2.925 and −3.097 ± 0.200), and at f = 0.48 (444 ha) 20,000 and 25,500 (−3.124 and −3.238 ± 0.182). REINFORCE beat random allocation at every f we tried but never exceeded greedy on average, and no more than half of its 20 seeds matched or beat greedy at any f (Supplementary Table S12). The 600-episode, one-hidden-layer policy therefore did not learn to save land for later increments. Other forms of coupling (shared pipelines, uncertain demand, multi-year finance), longer training or other algorithms might change this, and we did not test them.

Reinforcement learning matched a strong heuristic on a separable problem, did not exceed it when the sites shared a land budget, and clearly beat random allocation, as a deep reinforcement-learning agent also performed comparably to heuristics in water-network design[18]. We read this as evidence that it learns a sensible policy, not that it should replace greedy allocation here.

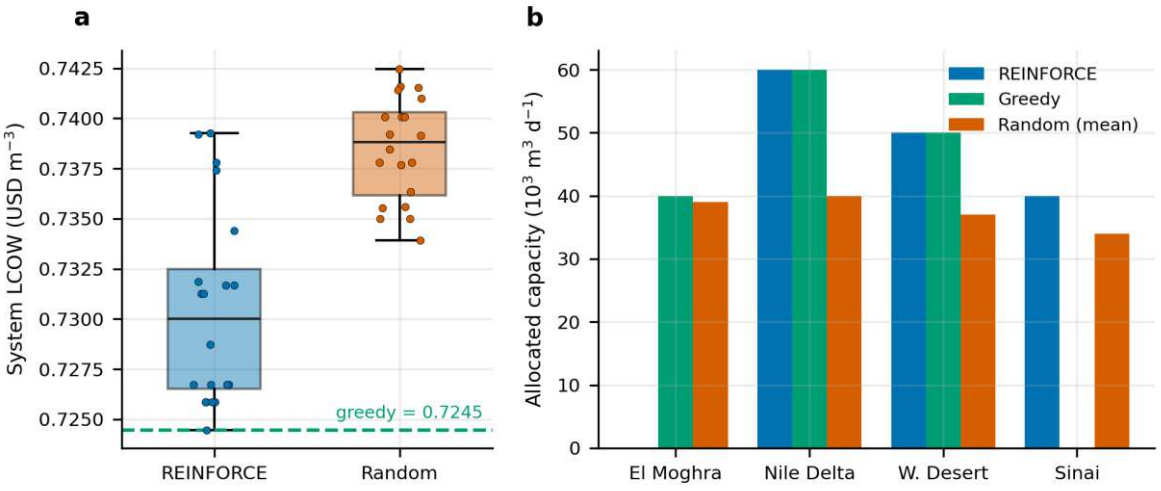


Figure 4 | Reinforcement learning for four-site capacity allocation. a, System LCOW of the REINFORCE policy (20 training seeds) and of a random feasible policy (20 episodes); dashed line, deterministic greedy policy. Boxes show quartiles and points show individual seeds. b, Final allocation of 150,000 m³ d⁻¹ across the four sites by the REINFORCE policy (seed 42), the greedy policy and the mean of ten random episodes.

The model's energy use depends strongly on recovery, and its capital cost is uncertain at large scale

Optimal designs are only as reliable as the model they optimise, so we compared the model with four independent sources (Fig. 5, Table 2). The only source in the large-scale range is not an operating plant but a published design study by some of the authors: a conceptual 300,000 m³ d⁻¹ zero-liquid-

discharge system for Nile Delta drainage water (2.7 g L^{-1}), sized and costed with design software[28]. We treat it as a design benchmark, not as validation against measurements. It uses two RO stages (recovery 90% and 60%; 96% overall for RO), a thermal brine concentrator and evaporation ponds (98% overall water recovery). From its pump powers, the RO train needs 1.31 kWh m^{-3} of RO permeate. The whole system, including pretreatment and the thermal unit's electricity, needs about 1.49 kWh m^{-3} of product, the figure carried in our 18-plant validation set[2].

The model's SEC for this feed depends strongly on the recovery assumed (Fig. 5b). At 2.9 g L^{-1}, the salinity entering the RO stages, it is 0.77, 0.93 and 1.18 kWh m^{-3} at 75%, 85% and 90% recovery. For a single-stage process this is expected: as recovery rises, the concentrate leaving the last membrane element becomes more saline and the applied pressure must exceed its osmotic pressure, so specific energy rises at high recovery, and the cost-optimal recovery depends on pretreatment and membrane costs as well as on energy[50]. Staged arrays with interstage pumping weaken this dependence, which the model does not represent (see below). At 90%, the design study's first-stage recovery, the model is 10% below the study's RO-train value and 4% above the global energy–salinity curve[26] (1.13 kWh m^{-3} at that salinity); at 85% it is 29% and 18% below, and at 75% it is 42% and 32% below. The whole-system figure of 1.49 is not comparable, because it includes pretreatment and thermal-unit electricity that the model's SEC excludes. A reference that does not state recovery therefore cannot by itself reveal a bias in the model. At the presets' default recoveries (75–85%) the model was 21–36% below the curve at 1.8–2.7 g L^{-1}, within 10% at 5–8 g L^{-1} (−4.6% and +9.5%) and 13% below at 41 g L^{-1} (Red Sea; Fig. 5a; Supplementary Table S7). The low-salinity differences largely reflect using 85% rather than 90% recovery for those presets. Across the brackish presets the model's SEC rises by about 0.13 kWh m^{-3} per g L^{-1} of salinity, 2.6 times the slope of the curve (0.052) and 1.6 times the partial coefficient for raw-water TDS in a regression on 36 real plants (0.083)[27]. We used only that coefficient: the regression's printed coefficients are rounded to two significant figures, so its absolute level cannot be reproduced (Methods).

For seawater, two small plants in Hurghada, on Egypt's Red Sea coast, give a range. One (600 m^3 d^{-1} per train) reported 2.7 kWh m^{-3} after retrofitting with a pressure-exchanger energy-recovery device[29]; a newer plant of three 600 m^3 d^{-1} trains reports an average of 2.2–2.5 kWh m^{-3}[30]. Both sources are vendor case studies, not independent measurements. The model gives 2.70, equal to the first and 8–23% above the second, so it sits at the upper end of the range for modern small plants. This tests the model's energy level for seawater, not its scale dependence, because its SEC is independent of capacity by construction.

For cost, the design study estimates a capital cost of 116 million USD, or 396 USD per m^3 d^{-1} of product for the whole system including the thermal concentrator and ponds (298 USD per m^3 d^{-1} for pumping, pretreatment and RO alone), and an operating cost of 0.487 USD per m^3 of feed. The model's capital cost at 300,000 m^3 d^{-1} is 682 USD per m^3 d^{-1}, 72% and 129% above these figures. Its operating cost is 0.433 USD m^{-3} with its own SEC (75% recovery) and 0.467 with the study's 1.49 kWh m^{-3} (−11% and −4%). The two capital estimates differ in scope and neither is a measured cost, so this shows that the model's capital cost is high relative to a bottom-up design estimate, not that it is wrong. This design study is one of the 18 cases used to calibrate the model's cost scaling, so these cost differences are in-sample residuals, not independent tests. Because LCOW is dominated by capital cost, the capital difference matters more for the economics than the SEC difference.

The strongest scale test the open data allow uses operating plants only. We calibrated the non-energy operating cost on the eight Texas plants of 25,000 m^3 d^{-1} or less that report their tariff (a constant of 0.439 USD m^{-3}, the median of reported operating cost minus energy at the reported tariff) and predicted the operating cost of the five largest operating plants in the Pearson et al. compilation[51]: four Florida plants (43,906–68,130 m^3 d^{-1}, tariff assumed 0.08 USD kWh^{-1}) and Kay Bailey Hutchison, Texas (56,818 m^3 d^{-1}). The model's mean absolute percentage error was 16.9% (mean absolute error

0.066 USD m⁻³, bias +0.064). A constant equal to the small-plant median gave 18.5%, and a power law fitted to the small plants and extrapolated gave 11.8% (Supplementary Table S13). The model was therefore no better than a naive baseline. Its operating cost is, in effect, a constant plus a small energy term, and the real operating costs of large plants (0.36–0.50 USD m⁻³) varied more with site than with capacity. For capital cost, only Kay Bailey Hutchison has a figure: 1,602 USD per m³ d⁻¹ against the model's 792, which is 51% low (a power law fitted to small plants gave 859, 46% low). Taken with the 72–129% overestimate relative to the 300,000 m³ d⁻¹ design study, capital cost errs in both directions, so no single correction would fix it. We found no operating brackish-water plant above about 68,000 m³ d⁻¹ with published costs. Egyptian brackish plants are small, so the 100,000–600,000 m³ d⁻¹ range used here remains an extrapolation beyond real data.

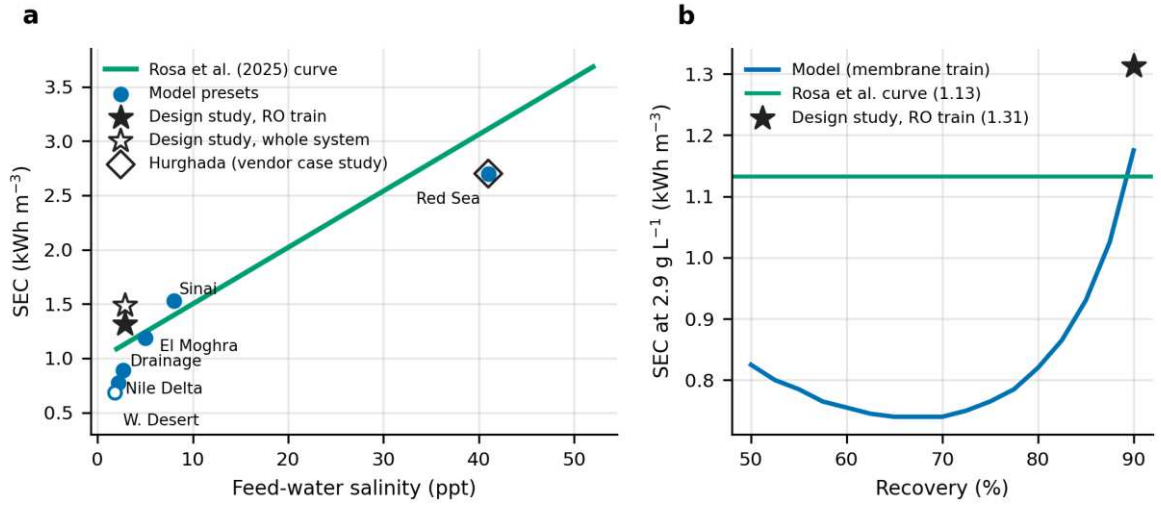


Figure 5 | Comparison of the model with independent data. a, SEC against feed salinity: model presets at their default recoveries (blue; the open circle, Western Desert, lies below the 2 g L⁻¹ lower limit of the curve and was extrapolated linearly), the global curve of Rosa et al.[26] (green), the RO train (filled star) and whole system (open star) of the 300,000 m³ d⁻¹ design study[28], and the Hurghada vendor case study[29] (open diamond, coinciding with the Red Sea preset). b, Model SEC against recovery at 2.9 g L⁻¹, the feed to the design study's RO stages, with the Rosa et al. curve value at that salinity (green line) and the design study's RO-train SEC at its first-stage recovery (star).

Table 2 | Model compared with a published design study and vendor case studies. Model SEC is the membrane-train value, given at 2.9 g L⁻¹ (the design study's RO feed). Operating cost is shown with the model's own SEC (75% recovery) and with the study's whole-system SEC. No operating plant above 100,000 m³ d⁻¹ was available. The design study was one of the 18 cases used to calibrate the model's costs.

Quantity	Reference value	Model	Difference
RO-train SEC, 300,000 m³ d⁻¹ design study	1.31 kWh m⁻³ of permeate, from pump powers[28]	1.18 (90% recovery); 0.93 (85%); 0.77 (75%)	−10%; −29%; −42%
SEC at the same salinity, global curve	1.13 kWh m⁻³[26]	1.18; 0.93; 0.77	+4%; −18%; −32%
Capital cost, design study	396 USD per m³ d⁻¹ (whole system); 298 (without thermal unit and ponds)[28]	682 USD per m³ d⁻¹	+72%; +129%
Operating cost, design study	0.487 USD per m³ of feed[28]	0.433 (model SEC); 0.467 (study SEC)	−11%; −4%
SEC, Hurghada plant after retrofit (600 m³ d⁻¹ per train)	2.7 kWh m⁻³, vendor case study[29]	2.70 kWh m⁻³	0%
SEC, second Hurghada plant (three 600 m³ d⁻¹ trains)	2.2–2.5 kWh m⁻³, vendor case study[30]	2.70 kWh m⁻³	+8% to +23%

An element-by-element membrane simulation shows that the model overstates energy at high recovery

The comparison above could not separate errors in the model from differences in scope, so we also checked the model's membrane-train energy against an element-by-element simulation of the kind that membrane suppliers' design software performs (Methods). Each 8-inch element was solved in series with the solution–diffusion equations, concentration polarisation and pressure drop, using the supplier's design equations[36] and water and salt permeabilities back-calculated from the datasheet of a standard brackish-water element[37] (A = 3.37 L m^{-2} h^{-1} bar^{-1}, B = 0.19 L m^{-2} h^{-1}). For each recovery the simulation took the lowest-energy array of seven-element vessels in one, two or three stages (Methods), run at the model's average flux of 22 L m^{-2} h^{-1}, with interstage booster pumps where needed to share the flux between stages, the same pump efficiency and no energy recovery, so that the two calculations differ only in how they obtain the pressure.

The model's SEC was higher than the simulated SEC in all 20 cases (four presets and the design study's RO feed, each at 75–90% recovery), by 1–118% (mean 43%; Fig. 6b; Table 3), and the gap grew with recovery (Fig. 6a). Between 75% and 90% recovery the model's SEC rose by 39–81%, whereas the simulated SEC varied by only 8–13%. The reason is structural: a staged array raises only the shrinking concentrate stream to the higher pressure needed near its tail, whereas the model applies the concentrate-end osmotic pressure to the whole feed. Without interstage boosters the simulated SEC rose by 6–57%, still less than the model's. The simulated SEC also rose more gently with salinity, by 0.095 kWh m^{-3} per g L^{-1} at 80% recovery against 0.175 for the model, close to the 0.083 found in a regression on real plants[27]. The simulation also flagged designs that the model accepts: for the Sinai preset (8 g L^{-1}) at 85% and 90% recovery the required pressure, 41.2 and 61.1 bar, exceeds the 41-bar rating of brackish-water elements[37].

Both calculations describe the membrane train only, and both fall below the whole-plant references above, which also include intake, pretreatment and product pumping, motor and drive losses and membrane ageing. The simulation therefore does not close the gap to real plants. What it suggests is that the model's energy figures are higher than an element-level calculation gives for the membrane train itself, and that the model's strong dependence of energy on recovery comes from its lumped pressure rule rather than from the process. The simulation is itself a model, built from datasheet parameters and not validated against measured data or supplier design software, so this conclusion is provisional.

The comparison with independent data limits how far the absolute numbers can be trusted. The model's SEC for brackish water depends strongly on recovery: it agrees with a global energy–salinity curve and with the RO train of a published design study to within about 10% at 90% recovery, but is 20–40% lower at 75–85%. The element-by-element simulation suggests that this recovery dependence is largely an artefact of the model's lumped pressure rule: the model's energy was 1–118% above that of the simulated membrane train, which barely changed between 75% and 90% recovery. Because the energy penalty of high recovery appears small in a staged design, the land–energy trade-off in Fig. 1 may be weaker than the model implies, and raising recovery to shrink the evaporation ponds may be limited mainly by mineral scaling rather than by energy. Its rise with salinity across the presets is 1.6–2.6 times that of the references, and its capital cost at 300,000 m^3 d^{-1} is 72% above the design study's estimate (an in-sample residual, because the study was one of the cases used to calibrate costs). No operating plant above 100,000 m^3 d^{-1} was available, so none of this is validation against measured large-scale performance. On the largest operating plants we could find (up to 68,000 m^3 d^{-1}) the model's operating cost was no better than a constant, and its capital cost was 51% low at the one plant with data. All four methods optimise the same model, so the comparisons among them are less affected than the absolute LCOW, SEC and payback values, which should not be used as predictions. Two further caveats apply: the design study's whole-system energy includes pretreatment and thermal-unit electricity that the model excludes, and the global curve is itself a model result, not plant

data. The hybrid-surrogate stress test points the same way: improving the learning method removes the error that the method causes, but the remaining uncertainty, in how costs scale above about 150,000 $m^3\ d^{-1}$, can be removed only by data. The next step is calibration, for example of the SEC–salinity–recovery relation against pilot-scale measurements of the kind used previously[13], and of capital cost against operating large plants. Replacing the lumped pressure rule with the element-level calculation used here, and checking both against suppliers' design software for the same feeds, is a direct first step.

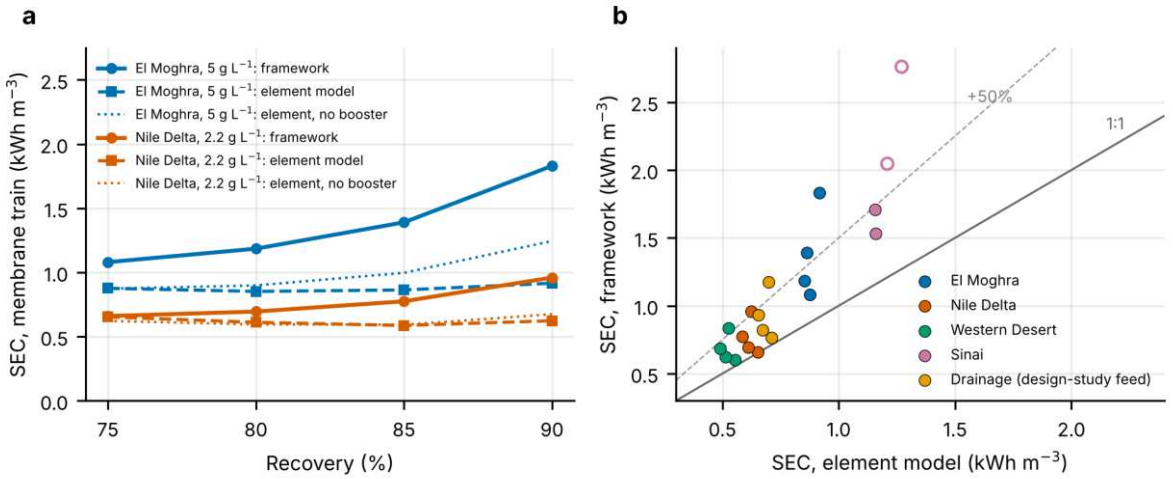


Figure 6 | The model's membrane-train energy compared with an element-by-element simulation. a, SEC against recovery for the El Moghra (5 g L^{-1}) and Nile Delta (2.2 g L^{-1}) presets from the model (solid), and from the simulation with interstage boosters (dashed) and without (dotted). b, Model SEC against simulated SEC for four presets and the design study's RO feed at 75, 80, 85 and 90% recovery; open symbols mark designs whose simulated pressure exceeds the 41-bar element rating. Both calculations use a pump efficiency of 0.80 and no energy recovery.

Table 3 | Membrane-train SEC from the element-by-element simulation and from the model. Each cell gives the simulated SEC with interstage boosters / the model's SEC (kWh m^{-3}) at the stated recovery; both use a pump efficiency of 0.80 and no energy recovery. † The simulated pressure exceeds the 41-bar element rating.

Feed	TDS (g L^{-1})	75%	80%	85%	90%
Western Desert	1.8	0.55 / 0.60	0.51 / 0.62	0.49 / 0.69	0.53 / 0.83
Nile Delta	2.2	0.65 / 0.66	0.61 / 0.70	0.59 / 0.78	0.62 / 0.96
Drainage water (design-study RO feed)	2.9	0.71 / 0.77	0.67 / 0.82	0.66 / 0.93	0.70 / 1.18
El Moghra	5.0	0.88 / 1.08	0.85 / 1.19	0.86 / 1.39	0.92 / 1.83
Sinai	8.0	1.16 / 1.53	1.16 / 1.71	1.21 / 2.05 †	1.27 / 2.77 †

With the simulated energy, the cheapest design needs about half the pond area

Because the model overstates the energy of high recovery, we asked whether that changes the design it recommends. We repeated the three-objective optimisation for the five brackish presets with the simulated SEC in place of the model's (Methods). The cheapest design moved from 62–74% to 78–85% recovery, its LCOW fell by only 0.4–1.3%, and its evaporation-pond area fell by 50–59%, from 1,741–2,939 ha to 866–1,408 ha (Fig. 7; means over three seeds). The smallest pond area reachable within 1% of the cheapest LCOW fell from 730–1,742 ha to 448–763 ha. For El Moghra the whole front moved towards smaller ponds and lower cost (Fig. 7a), so much of the land–energy trade-off in the original front (Fig. 1) appears to come from the lumped energy model rather than from the process, to the extent that the element-level simulation is accurate.

Two cautions apply. The simulated optimum lies at or near the upper end of the two-stage arrays allowed in the simulation, so its exact recovery depends on how arrays are chosen. And neither the model nor the simulation includes mineral scaling, which may cap recovery below these values for some feeds.

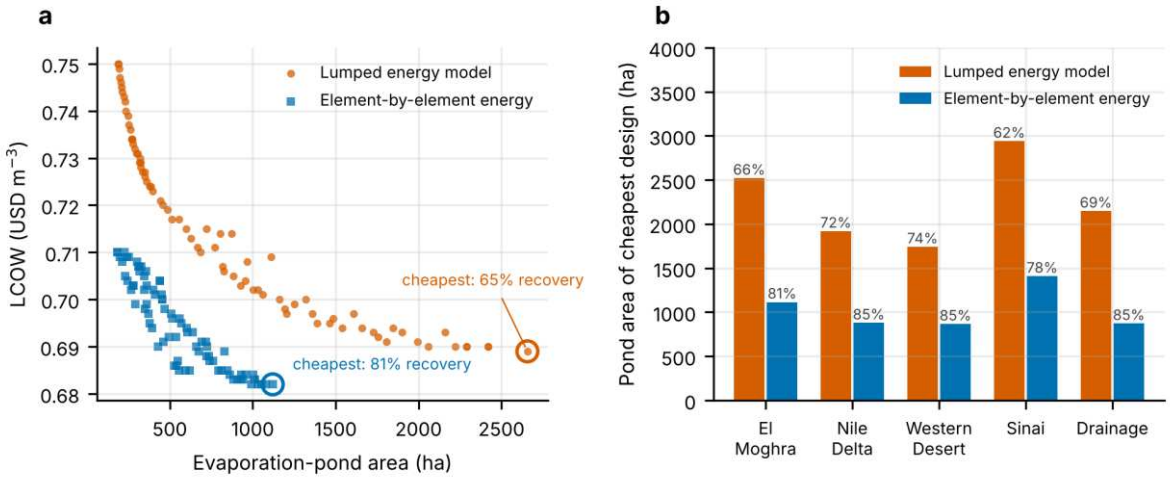


Figure 7 | Re-optimising with the simulated energy. a, Three-objective NSGA-II fronts for the El Moghra preset (seed 42) projected on the plane of pond area and LCOW, with the model's lumped SEC (orange) and the simulated SEC (blue); circles mark the cheapest designs. b, Pond area of the cheapest design for the five brackish presets (means over three seeds), with its recovery above each bar.

In this model, cost data set the price of water but not the recovery decision

The check against independent data showed that the model's costs are uncertain at large scale, and the argument so far has been that more data are needed. To ask which data, we treated eight inputs as uncertain and measured how much knowing each would improve the choice of recovery for a 200,000 $m^3\ d^{-1}$ plant (Methods). When cost meant the LCOW alone, as in the optimisation above, the cost-optimal recovery for El Moghra ranged from 65% to 80% across the samples (5th–95th percentiles), and the choice of energy model explained almost all of that variation (first-order index 0.98–1.00 for every preset; Fig. 8a). At the middle of the ranges, the lumped model chose 57–69% recovery and 1,458–2,448 ha of ponds, and the simulation 78–85% and 573–915 ha. The monetary stakes of that choice were small, an EVPI of 0.03–0.06 million USD per year, because the LCOW changes little with recovery; the stakes in land were large.

Once the ponds were costed, the picture changed. Ponds then made up 14–21% of the cost of water at the best design, and the highest recovery the array could reach was optimal in 94–100% of samples for every preset (Fig. 8a; 84% recovery for Sinai, where the 41-bar rating binds). No input changed that decision enough to matter: the EVPI was at most 0.02 million USD per year, and no input's EVPPI rose above the noise floor (Table 4). The uncertain inputs did matter for the cost of water itself, which spanned 0.69–1.09 USD m^{-3} for El Moghra (5th–95th percentiles). Its variance was explained mainly by the capital-cost level (first-order index 0.25), the non-energy operating cost (0.23), the pond cost (0.19) and the discount rate (0.13), with the capacity exponent at 0.08 at this capacity (Fig. 8b). In this model, the data that would most change the design are therefore not cost data but the feed-water chemistry and scaling tests that set the highest feasible recovery, which neither model represents; because the best recovery sits at that limit, this conclusion is conditional on how the limit is set. Cost data, above all large-plant capital and operating costs, are what is needed to price that design.

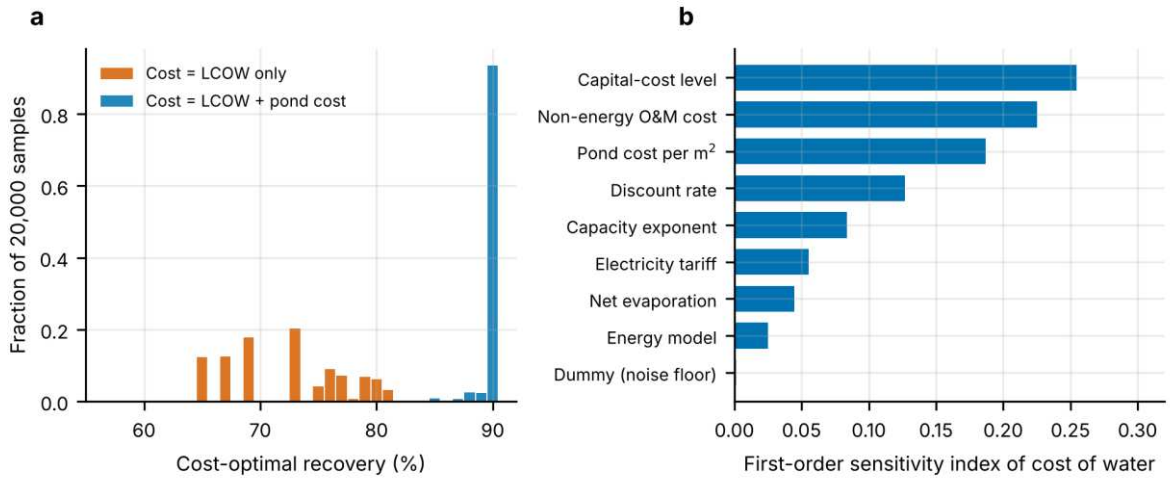


Figure 8 | Value of information for the recovery decision (El Moghra, 200,000 m³ d⁻¹, 20,000 samples). a, Distribution of the cost-optimal recovery when cost is the LCOW alone (orange) and when the annualised cost of evaporation ponds is added (blue). b, First-order sensitivity indices of the minimum cost of water (LCOW plus ponds) for the eight uncertain inputs and a dummy input with no effect.

Table 4 | Value of information for the recovery decision of a 200,000 m³ d⁻¹ plant (20,000 samples). Best recovery at the middle of the input ranges with the model's lumped SEC and with the simulated SEC when cost is the LCOW alone; expected value of perfect information (EVPI, million USD per year) for the two cost definitions; share of samples in which the highest feasible recovery is optimal when ponds are costed; and the ponds' share of the cost of water at the best design. With ponds costed, no input had an EVPPI above the noise floor.

Preset	Best recovery, LCOW only: lumped / simulated	EVPI, LCOW only	Highest feasible recovery optimal, ponds costed	EVPI, ponds costed	Pond share of cost
El Moghra	60% / 81%	0.064	94% (at 90%)	0.015	14%
Nile Delta	66% / 85%	0.049	100% (at 90%)	0.000	14%
Western Desert	69% / 85%	0.040	100% (at 90%)	0.000	15%
Sinai	57% / 78%	0.034	99% (at 84%)	0.001	21%
Drainage water	66% / 85%	0.051	99% (at 90%)	0.000	14%

Cost, payback and brine disposal of the selected designs

For the cheapest design on the NSGA-II front (recovery 65.6%, 294,443 m³ d⁻¹) the model gives a capital cost of 201 million USD, an operating cost of 0.445 USD m⁻³ and an LCOW of 0.689 USD m⁻³ (Supplementary Table S10). If the water displaces trucked supply, as at remote sites, the simple payback is 1.3 years. At the subsidised municipal tariff assumed in the framework (0.045 USD m⁻³), revenue would not cover operating cost, so such a plant depends on public finance. The allocation found by REINFORCE (seed 42), which uses three of the four sites, has a capital cost of 120 million USD and site LCOW of 0.717–0.761 USD m⁻³. Brine disposal is the main land–energy trade-off. The cheapest design needs 2,563 ha of evaporation ponds; zero-liquid-discharge treatment (a brine concentrator at 95% recovery plus a crystalliser, yielding 108,561 kg d⁻¹ of dry salt) would remove the ponds but, at the framework's assumed 40–60 kWh per m³ of concentrate, would use more than 20 times the energy of the membrane train. Because the model's capital cost is 72–129% above the design-study estimate but 51% below the one large operating plant with data, the sign of the error in these capital figures, and so in the payback, is uncertain; they illustrate the trade-offs and are not predictions.

Implications for design practice

On one model and at matched budgets, the four methods behaved differently from what a single-method study would suggest. Taken together, the results suggest a division of labour (Table 5). For multi-objective design with cheap evaluations, NSGA-II with a population of 80 or more is a good default and gives a compact shortlist of designs. We expect MOBO to be preferable when each evaluation is expensive enough to justify its overhead, but our problem does not test that, and its archive must be filtered to a usable shortlist. For surrogates, a physics-structured hybrid was more accurate than purely data-driven surrogates when its assumed scaling held, but any use beyond the training range should still be flagged, because its accuracy there rests on that assumption. For sequential allocation with independent site costs, a greedy rule is hard to beat; reinforcement learning did not become attractive in our test of coupling through a shared land budget, although other forms of coupling remain untested.

For planners in Egypt and in similar settings, these results suggest a tentative workflow, to be confirmed on calibrated models. First, screen designs with NSGA-II (population of 80 or more) directly on the physics-based model, which is cheap enough not to need a surrogate. Second, use a surrogate only when it sits inside a larger loop, such as uncertainty analysis or regional planning, and then use a hybrid with an explicit flag for queries outside the training range[21]. Third, allocate capacity between sites with a greedy rule and check the result against the shared brine-land budget. Fourth, report every cost figure above about 70,000 $m^3\ d^{-1}$, the size of the largest operating plant with published costs, as a scenario that depends on an assumed scaling law, ideally bounded by steeper and flatter alternatives such as those used here. Fifth, cost the brine ponds explicitly, calculate energy element by element, and set recovery from scaling tests on the actual feed water, which in this analysis is the information that most changes the design.

Table 5 | Summary of the method comparisons and model checks.

Method	Problem	Budget and repeats	Main result
NSGA-II versus MOBO	Three-objective design, 100,000–300,000 $m^3\ d^{-1}$ (El Moghra)	About 1,040 evaluations each; 10 seeds (two objectives), 26 runs over five presets (three objectives)	Two-objective hypervolume equal ($p = 1.00$). Three-objective: MOBO larger in all 26 runs: median 5% at population 40, 1.5% at population 100
Gradient boosting and neural-network surrogates	LCOW, 100,000–600,000 $m^3\ d^{-1}$	4,000 plants; 5-fold cross-validation	$R^2 = 0.9992$ and 0.9990. A surrogate trained to 150,000 $m^3\ d^{-1}$ was 4.8% off at 600,000 $m^3\ d^{-1}$
Hybrid surrogate (scaling law plus gradient-boosting residual)	LCOW, trained on 2,000–150,000 $m^3\ d^{-1}$, tested on 150,000–600,000 $m^3\ d^{-1}$	6,000 training and 2,000 test plants; 5 seeds; three scaling scenarios	Error 0.2% against 3.2% for gradient boosting if scaling continues; 3.2% against 6.6% with steeper scaling; 2.8% against 0.4% with no further economies of scale
REINFORCE	Four-site allocation of 150,000 $m^3\ d^{-1}$	600 episodes; 20 seeds	LCOW 0.7305 ± 0.0049 versus random 0.7385 ± 0.0026 ($p = 1 \times 10^{-5}$); greedy 0.7245
Element-by-element RO simulation	Membrane-train SEC, five feeds, 75–90% recovery	20 cases; fouling factor 0.70–1.0 and pressure drop ±50%	Model SEC 1–118% higher; its rise with recovery appears largely to be an artefact of the lumped pressure rule
Re-optimisation with simulated energy	Three-objective design, five presets	NSGA-II, about 1,040 evaluations, 3 seeds	Cheapest design moves from 62–74% to 78–85% recovery; its pond area falls by 50–59%
Value of information	Recovery of a 200,000 $m^3\ d^{-1}$ plant, eight uncertain inputs	20,000 samples; binned EVPPI	With ponds costed, no input changes the decision (EVPI ≤ 0.02 million USD per year)

Limitations

Beyond the model checks above, the study has further limits. The feed presets are illustrative composites for Egypt; the pumping heads and evaporation rate in the model are placeholders, and groundwater lift was excluded from all results. Neither the model nor the element-level simulation checks mineral scaling, which in practice caps the recovery of brackish-water plants. The two-objective hypervolume used one preset and ten seeds, and the three-objective comparison 26 runs over five presets (four seeds for four of them); MOBO uses Chebyshev scalarisation rather than a hypervolume-based acquisition function. The allocation problem has four sites and illustrative limits. No operating plant in the large-scale range was available to us. The element-by-element simulation uses permeabilities back-calculated from a single datasheet test point and has not been checked against measured plant data or supplier design software, so the energy results that depend on it are provisional. All comparisons use one model, two decision variables for the optimisation problems and one set of illustrative presets, so the ranking of methods may differ for other problems, and the conclusion that model uncertainty outweighs the choice of algorithm applies to this study only. Future work should combine recovery and allocation in one problem, test reinforcement learning with other forms of coupling between sites, and compare the hybrid surrogate with physics-informed neural networks once cost data for large plants become available.

The value-of-information analysis used independent uniform ranges chosen by us, a single capacity and the model's cost structure, in which capital cost does not depend on recovery. In practice wells, intake and pretreatment scale with the feed flow, which favours high recovery further, whereas antiscalant and cleaning costs rise with recovery.

Conclusions

On a physics-based model of large brackish-water desalination plants with Egyptian feed waters, in this study the choice among AI methods mattered less than the trustworthiness of the model they optimise, although the comparison covers one model and problems with few decision variables. NSGA-II and multi-objective Bayesian optimisation found the same cost–energy trade-off, and an NSGA-II population of 80 or more removed most of the difference in how fully they covered pond area. A greedy rule matched or beat reinforcement learning for allocating capacity across sites, including under a shared brine-land budget. Data-driven surrogates failed silently beyond their training range; a hybrid surrogate that fits a capacity scaling law cut that error from 3.2% to 0.2%, but only for as long as the assumed scaling held. Checks against independent data showed that the model's capital cost errs in both directions at large scale, and an element-by-element membrane simulation suggested that its energy use is too high, increasingly so at high recovery, so the energy penalty of high recovery, and with it the land–energy trade-off, is probably smaller than the model suggests: re-optimised with the simulated energy, the cheapest design needed about half the evaporation-pond area. A value-of-information analysis showed that once the ponds are costed, none of eight uncertain cost inputs changes the best recovery, which is set by the feasible limit; in this model, cost data determine the price of water, not the design. For plants larger than any operating brackish plant with published costs that we found, design results should be reported as scenarios bounded by alternative cost-scaling laws, and the most valuable next steps appear to be not a better algorithm but scaling tests on the actual feed waters, which set the design, and cost and performance data from large plants, which price it.

Use of AI tools

Large language model assistance (Claude Sonnet 5.5 and Claude Opus 5.5, Anthropic, 2026) was used throughout this work: to write the Python code for the algorithms and analyses, to run the computational experiments, to generate figures, to search for and summarise the literature, and to draft

the text. The authors set the study questions and take full responsibility for the content. The key scripts (NSGA-II benchmark validation, external validation against the Rosa et al. data, and the reinforcement-learning allocation) were re-executed from the archived release by the AI assistant and reproduced the reported values. During preparation the AI assistant produced an incorrect "correction" to the published Stillwell & Webber equation; this was detected against the source paper and removed, and no result in this paper depends on it. The assistant also initially described the 300,000 m^3 d^{-1} design study as an operating plant; this was found and corrected when the source paper was read.

Data availability

The raw per-seed results, the derived result files behind all tables and figures, the analysis and figure scripts, and the unmodified Rosa et al. source-data file (CC BY 4.0) are available in the GitHub repository https://github.com/ayatshawky821-beep/egypt-desalination-framework and archived on Zenodo as release V1.2.0 (DOI: https://doi.org/10.5281/zenodo.23590553). Plant values are from the cited publications.

Code availability

All code (MIT licence) is in the same repository and Zenodo release (V1.2.0); the physics-based framework alone is release v1.0.0 (https://doi.org/10.5281/zenodo.23101824).

Author contributions

A.S.N. conceived the study, developed the software, performed the computational experiments and analysis, and wrote the original draft. R.S.E., M.A.E., S.T. and M.M.E. supervised the work and reviewed and edited the manuscript. All authors approved the final manuscript.

Funding

This research received no external funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information accompanies this paper. Correspondence and requests for materials should be addressed to A.S.N. ayatshawky821@gmail.com. ORCID iD: 0009-0002-6209-2732.

References

1. Najjaa, A. S., Ettouney, R. S., El-Rifai, M. A., Tewfik, S. & Elsayed, M. M. Review of techno-economic aspects of brackish water desalting. Egypt. J. Chem. (2025). https://doi.org/10.21608/EJCHEM.2025.396722.12292

2. Najjaa, A. S., Ettouney, R. S., El-Rifai, M. A., Tewfik, S. & Elsayed, M. M. MF/UF/NF/RO desalination design framework: a case study on brackish groundwater and agricultural drainage water desalination in Egypt. Manuscript in preparation (2026). Code: https://doi.org/10.5281/zenodo.23101824 (v1.0.0).

3. Jones, E., Qadir, M., van Vliet, M. T. H., Smakhtin, V. & Kang, S.-M. The state of desalination and brine production: a global outlook. Sci. Total Environ. 657, 1343–1356 (2019). https://doi.org/10.1016/j.scitotenv.2018.12.076

4. Stanton, J. S. et al. Brackish groundwater in the United States. U.S. Geological Survey Professional Paper 1833 (2017). https://doi.org/10.3133/pp1833

5. Ahmed, M. et al. Use of evaporation ponds for brine disposal in desalination plants. Desalination 130, 155–168 (2000).

6. Ghaffour, N., Missimer, T. M. & Amy, G. L. Technical review and evaluation of the economics of water desalination: current and future challenges for better water supply sustainability. Desalination 309, 197–207 (2013). https://doi.org/10.1016/j.desal.2012.10.015

7. Loutatidou, S., Chalermthai, B., Marpu, P. R. & Arafat, H. A. Capital cost estimation of RO plants: GCC countries versus southern Europe. Desalination 347, 103–111 (2014). https://doi.org/10.1016/j.desal.2014.05.033

8. Caldera, U. & Breyer, C. Learning curve for seawater reverse osmosis desalination plants: capital cost trend of the past, present, and future. Water Resour. Res. 53, 10523–10538 (2017). https://doi.org/10.1002/2017WR021402

9. Ahmad, U., Abdala, A., Ng, K. C. & Akhtar, F. H. Machine learning-driven design of membranes for saline and produced water treatment across scales. Environ. Sci.: Water Res. Technol. 11 (2025). https://doi.org/10.1039/D5EW00417A

10. Habieeb, A. R., Kabeel, A. E., Sultan, G. I. & Abdelsalam, M. M. Advancements in water desalination through artificial intelligence: a comprehensive review of AI-based methods for reverse osmosis membrane processes. Water Conserv. Sci. Eng. 8, 53 (2023). https://doi.org/10.1007/s41101-023-00227-7

11. Mahadeva, R., Kumar, M., Gupta, V., Manik, G. & Patole, S. P. Modified whale optimization algorithm based ANN: a novel predictive model for RO desalination plant. Sci. Rep. 13, 2901 (2023). https://doi.org/10.1038/s41598-023-30099-9

12. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197 (2002).

13. Abkar, L. et al. From fixed points to optimum regions: AI–NSGA-II framework for high-recovery, low-energy brackish water RO. Water Res. 289, 124934 (2026). https://doi.org/10.1016/j.watres.2025.124934

14. Chen, K., Luo, J., Chen, J., Lu, Y. & Heng, Y. A rapid-convergent particle swarm optimization approach for multiscale design of high-permeance seawater reverse osmosis systems. Commun. Eng. 3 (2024). https://doi.org/10.1038/s44172-024-00289-y

15. Kaoud, O. G., Elbassoussi, M. H., Abido, M. A. & Zubair, S. M. Renewable–diesel-powered multi-element series SWRO for off-grid Sinai resorts: a multi-objective optimization for cost–carbon minimization. Desalination 623, 119823 (2026). https://doi.org/10.1016/j.desal.2025.119823

16. Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. 8, 229–256 (1992).

17. Bonny, T., Kashkash, M. & Ahmed, F. An efficient deep reinforcement machine learning-based control reverse osmosis system for water desalination. Desalination 522, 115443 (2022). https://doi.org/10.1016/j.desal.2021.115443

18. Tsiami, L., Makropoulos, C. & Savic, D. A. Rethinking urban water network design: a reinforcement learning framework for long-term flexible planning. Water Resour. Manag. (2025). https://doi.org/10.1007/s11269-025-04290-8

19. Yang, S., Kim, H., Hong, Y., Yee, K., Maulik, R. & Kang, N. Data-driven physics-informed neural networks: a digital twin perspective. Comput. Methods Appl. Mech. Eng. 428, 117075 (2024).

20. Helali, S., Albalawi, S. & Bel Hadj Ali, N. Harnessing physics-informed neural networks for performance monitoring in SWRO desalination. Water 17, 297 (2025). https://doi.org/10.3390/w17030297

21. Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633 (2021). https://doi.org/10.1111/2041-210X.13650

22. Psichogios, D. C. & Ungar, L. H. A hybrid neural network–first principles approach to process modeling. AIChE J. 38, 1499–1511 (1992). https://doi.org/10.1002/aic.690381003

23. von Stosch, M., Oliveira, R., Peres, J. & Feyo de Azevedo, S. Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput. Chem. Eng. 60, 86–101 (2014). https://doi.org/10.1016/j.compchemeng.2013.08.008

24. Karpatne, A. et al. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29, 2318–2331 (2017). https://doi.org/10.1109/TKDE.2017.2720168

25. Karniadakis, G. E. et al. Physics-informed machine learning. Nat. Rev. Phys. 3, 422–440 (2021). https://doi.org/10.1038/s42254-021-00314-5

26. Rosa, L., Gabrielli, P. & Sangiorgio, M. Supplementary dataset for “Global multi-model assessment of energy, costs, and emissions trade-offs associated with reverse osmosis desalination under future water scarcity”. Zenodo https://doi.org/10.5281/zenodo.15569107 (2025).

27. Stillwell, A. S. & Webber, M. E. Predicting the specific energy consumption of reverse osmosis desalination. Water 8, 601 (2016). https://doi.org/10.3390/w8120601

28. El Sayed, M. M., Abulnour, A. M. G., Tewfik, S. R., Sorour, M. H. & Shaalan, H. F. Reverse osmosis membrane zero liquid discharge for agriculture drainage water desalination. Membranes 12, 923 (2022).

29. Energy Recovery, Inc. Energy saving retrofit for enhanced plant capacity at Red Sea resort (Interwater / Safa Water Technology, Hurghada, Egypt). Vendor case study (2024). https://energyrecovery.com/resources/energy-saving-retrofit-for-enhanced-plant-capacity/

30. Danfoss. Trouble-free and energy efficient water production with Danfoss technology (Interwater, Hurghada, Egypt). Vendor case story, 12 July 2021. https://www.danfoss.com/en/service-and-support/case-stories/hpp/trouble-free-and-energy-efficient-water-production-with-danfoss-technology

31. Zitzler, E., Deb, K. & Thiele, L. Comparison of multiobjective evolutionary algorithms: empirical results. Evol. Comput. 8, 173–195 (2000).

32. Rasmussen, C. E. & Williams, C. K. I. Gaussian Processes for Machine Learning (MIT Press, 2006).

33. Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. J. Glob. Optim. 13, 455–492 (1998).

34. Zitzler, E. & Thiele, L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans. Evol. Comput. 3, 257–271 (1999).

35. Pedregosa, F. et al. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).

36. The Dow Chemical Company. FILMTEC Membranes. System Design: System Performance Projection, Design Equations and Parameters. Form No. 609-02057. https://www.lenntech.com/data-sheets/filmtec-design-equations-l.pdf (accessed 13 October 2026).

37. The Dow Chemical Company. FILMTEC BW30-400/34i element: product data sheet. https://www.lenntech.com/data-sheets/Dow-Filmtec-BW30-400%2034i-L.pdf (accessed 13 October 2026).

38. Nicot, J.-P., Gross, B., Walden, S. & Baier, R. Self-Sealing Evaporation Ponds for Desalination Facilities in Texas. Final report to the Texas Water Development Board (Bureau of Economic Geology, Univ. Texas at Austin, 2007).

39. Strong, M., Oakley, J. E. & Brennan, A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med. Decis. Making 34, 311–326 (2014).

40. Saltelli, A. et al. Global Sensitivity Analysis: The Primer (Wiley, 2008).

41. Wilcoxon, F. Individual comparisons by ranking methods. Biometrics Bull. 1, 80–83 (1945).

42. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272 (2020).

43. Guerreiro, A. P., Fonseca, C. M. & Paquete, L. The hypervolume indicator: computational problems and algorithms. ACM Comput. Surv. 54, 119 (2021). https://doi.org/10.1145/3453474

44. Knowles, J. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Trans. Evol. Comput. 10, 50–66 (2006). https://doi.org/10.1109/TEVC.2005.851274

45. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & de Freitas, N. Taking the human out of the loop: a review of Bayesian optimization. Proc. IEEE 104, 148–175 (2016). https://doi.org/10.1109/JPROC.2015.2494218

46. Friedman, J. H. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232 (2001).

47. Breiman, L. Random forests. Mach. Learn. 45, 5–32 (2001).

48. Sutton, R. S. & Barto, A. G. Reinforcement Learning: An Introduction 2nd edn (MIT Press, 2018).

49. Federgruen, A. & Groenevelt, H. The greedy procedure for resource allocation problems: necessary and sufficient conditions for optimality. Oper. Res. 34, 909–918 (1986). https://doi.org/10.1287/opre.34.6.909

50. Zhu, A., Christofides, P. D. & Cohen, Y. Effect of thermodynamic restriction on energy cost optimization of RO membrane water desalination. Ind. Eng. Chem. Res. 48, 6010–6021 (2009). https://doi.org/10.1021/ie800735q

51. Pearson, J. L., Michael, P. R., Ghaffour, N. & Missimer, T. M. Economics and energy consumption of brackish water reverse osmosis desalination: innovations and impacts of feedwater quality. Membranes 11, 616 (2021). https://doi.org/10.3390/membranes11080616

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryadditionalsectionsS9S12final.docx