Towards Data-Driven Methods for Decarbonizing Reverse Osmosis Desalination

Om Sanan
Scarsdale High School
Scarsdale, USA
om.sanan007@gmail.com

Joshua Sperling
NREL
Golden, USA
joshua.sperling@nrel.gov

David Greene
NREL
Golden, USA
david.greene@nrel.gov

Ross Greer
UC San Diego
La Jolla, USA
regreer@ucsd.edu

***Abstract*—Desalination, when combined with energy-efficient operations and clean energy, has significant potential to address water security, resilience, and costs. Energy demands of desalination must be met, yet current inefficiencies increase costs, and the use of non-renewable sources exacerbates climate change. This research seeks to fill these gaps by advancing integrated water-energy system decarbonization, using data from multiple U.S. desalination plants while defining optimization functions and constraints to reduce energy costs and carbon emissions. A framework is designed for the optimal sizing of grid-connected hybrid renewable energy and storage systems using Artificial Intelligence algorithms to utilize at least 50% renewables.**

***Index Terms*—energy optimization, multi-source data-driven modeling, desalination, hybrid renewable energy systems, renewable energy sources, artificial intelligence, water-energy nexus**

I. Introduction and Related Research

Water scarcity and carbon emissions are critical issues globally that require integrated and proactive responses to droughts, flooding, already limited freshwater resources and as part of future cost-effective and sustainable infrastructure development [1]. Today's decisions regarding the management of water and energy are linked inextricably. A large amount of energy is required to collect, treat, and supply water, which means that thoughtful water management strategies can result in a significant reduction in energy use and greenhouse gas emissions. In response to an escalating global freshwater scarcity crisis, alternative resources such as salt-water desalination are being explored. Reverse Osmosis has proven to be the most effective desalination process, but its costs (especially from energy) still need to be reduced further to make it economically viable. Specific Energy Consumption (SEC - measured in kWh/m3) for seawater desalination has reduced from 8 kWh/m3 in the 1970s to 2.5-3.0 kWh/m3 approaching the thermodynamic limit of 0.76 kWh/m3 for seawater reverse osmosis [2] (though still higher than surface water treatment plant energy use of approximately 0.50 kWh/m3). This reduction in energy consumption is attributable to high-efficiency pressure pumps, improvement in reverse osmosis membrane structure, and use of energy recovery devices. In order to reduce the levelized cost of water and energy usage, strategies are emerging including for water reuse, optimization of water demand, energy efficiency, and leveraging renewable energy (RE), whose cost and availability make it a more compelling and sustainable option than fossil energy. For example, large seawater reverse osmosis (SWRO) desalination plants in Israel, such as Sorek (165 MGD) and Ashkelon (87 MGD) have been able to achieve a levelized cost of water of less than $0.6/m3 due to a combination of high-pressure pumps, energy recovery devices, efficient membrane design, and significant water reuse procedures [3]. Decarbonization of water desalination will require a diversified approach, incorporating multiple RE sources, such as photovoltaic (PV), concentrated solar power (CSP), wind, geothermal, hydro, storage options (battery and hydrogen), along with traditional energy sources, such as natural gas.

In reverse osmosis, water molecules are forced through a semipermeable membrane from a higher salt concentration to a lower salt concentration under the application of hydraulic pressure. Currently, seawater (TDS $\geq$ 30,000 mg/L) and brackish water (TDS of 1,000-10,000 mg/L) reverse osmosis are energy-intensive processes primarily dependent on non-renewable fossil fuels, contributing to carbon emissions of 1.6-6.9x generated by treated groundwater [4]. To alleviate both the environmental and economic implications of this dependency, the integration of increasingly cost-effective RE sources into desalination systems has emerged as a promising solution. Despite the evident potential of such renewable energy-powered desalination, it only represents a marginal 1% of the global capacity [5]. However, RE is getting recognition; built in 2018, Al Khafji SWRO plant (16 MGD) in Saudi Arabia is the world's first large-scale grid-connected PV solar-powered seawater desalination plant that is considered a zero emissions plant, producing excess energy during the day and drawing less energy at night [6].

In tandem with the push towards RE sources, the rise of Artificial Intelligence (AI) brings substantial potential for optimizing water demand and RE consumption. AI's ability to forecast RE production based on meteorological data, as well as to optimize the operation of complex energy systems and dynamically produce treated water, makes it a potent tool in modern energy-water management. AI can parse through extensive data, identify patterns and make predictions with unrivaled speed and accuracy. Specifically, advancements in Machine Learning (ML) and Deep Learning (DL) show promise in enhancing the efficiency and performance of renewable energy systems within water desalination plants. The self-learning capabilities of these techniques enable them to adapt to changing conditions and make real-time adjustments, thereby improving system reliability and efficiency.

To better understand the practicality of these innovative approaches, various studies have explored the integration of RE sources and AI into desalination. For instance, Carta, Cabrera et al. used an off-grid energy system to power a seawater reverse osmosis plant, thereby addressing the optimal economic sizing of off-grid SWRO plants [7]. This approach demonstrates the potential of off-grid energy sources for water desalination, informing this research.

Building on this, Maisanam, Sharma et al. utilized a Hybrid Optimization Model for Electric Renewable (HOMER)-based techno-economic assessment to design an optimal solar PV-battery-based system for a remote water supply station [8]. This study highlighted the significant reduction in CO2 emissions achievable with RE systems and the importance of site-specific assessments in system optimization. Following this, Rahimi, Shirvani et al. addressed the intermittency challenges associated with solar energy by integrating energy recovery devices and energy storage systems, offering a method to optimize energy consumption in desalination [9].

This energy optimization was further explored by Rodriguez, Fontan et al., who demonstrated the potential of machine learning to predict solar energy generation and improve utilization in desalination systems [10]. Their study laid the groundwork for Abdelshafy, Jurasz et al. to develop a hybrid RE system, powered by photovoltaic modules and wind turbines, to power a reverse osmosis desalination plant using Particle Swarm Optimization-Grey Wolf Optimizer (PSO-GWO) hybrid model [11]. Through this, they demonstrated the potential of AI for optimizing RE use.

In the study by Lai, Pai et al., various machine learning models such as Artificial Neural Network (ANN), Genetic Algorithms (GA), Support Vector Regression (SVR), Long short-term memory (LSTM), etc. were explored for their effectiveness in RE prediction [12]. This research served to underscore the value of AI in RE applications. This research utilized regression models and general algorithms to select system configurations that optimized cost, freshwater production, and energy resource exploitation.

The existing literature lacks a comprehensive approach to optimizing the diversity of grid-connected renewable energy sources across different regions using real-life RO desalination plants to minimize the energy cost and carbon footprint of water production. This study aims to bridge this gap by optimizing a variety of RE systems and storage for desalination in various regions, using the potential of AI algorithms. The study has the framework to achieve decarbonization on a large scale by proposing the use of at least 50% on-site RE for mid-to-large desalination plants in the U.S.

II. Real-World Data to Simulate Power Production and Consumption

This research is grounded in real-world data, which provides a robust foundation for analysis and modeling efforts. This data, collected from multiple sources, offers a high-level view of the operations of several water treatment facilities in the U.S., as well as the environmental conditions that influence the potential for renewable energy generation at these sites.

A. Water Treatment Facilities Data

Operational data from 4 geographically diverse seawater and brackish water desalination facilities in the U.S. was collected via online sources and from plant operators. With approximately 28 million gallons of actual water production per day: the Tampa Bay Desalination Plant, San Antonio Water System, Alameda County Water District, and Kay Bailey Hutchison (see Table I) are facilities that offer a broad representation of the different environmental and operational conditions that can influence the feasibility and efficiency of RE systems for water treatment.

TABLE I: Overview of Data from 4 Seawater/Brackish Water Desalination Plants: Tampa Bay Desalination (TBD), San Antonio Water Services (SAWS), Alameda County Water (ACW), and Kay Bailey Hutchison (KBH). Tampa Bay desalinates seawater and the others desalinate brackish water.

	TDS	Water Production		Energy Usage	SEC
	ppm	MGD	m^3 / year	kWh / year	kWh / m^3
TBD, FL	35000	8.2	11,288,134	43,023,680	3.81
SAWS, TX	1325	3.9	5,349,394	5,019,000	0.94
ACW, CA	1111	6.7	9,198,486	4,205,916	0.46
KBH, TX	2500	9.0	12,433,762	22,380,772	1.8
Total	-	27.7	38,269,777	74,629,368	1.95

The commonalities in data collected from these facilities include daily or monthly energy use, energy costs, water flows, and water quality from 2017 to 2022. This information is crucial for understanding baseline energy demands of these facilities and potential cost savings achieved through the adoption of RE systems. Additionally, the water quality data can provide insights into the energy intensity of the desalination process, which can vary depending on the salinity and other characteristics of the source water.

Each facility also provided additional data unique to their operations, further enriching this dataset. For instance, some facilities provided data on their existing RE studies, water quality, or water/energy storage, which can serve as a benchmark for future optimization efforts.

B. Weather Data

Extensive weather data for all study locations was collected, cleaned and processed to understand differences, anomalies, and ensure consistency. The National Solar Radiation Database (NSRDB), created by the National Renewable Energy Laboratory (NREL) was the primary database used, providing hourly weather data from 2010 to 2021.

The NSRDB contains many valuable variables for understanding renewable energy potential, including solar irradiance, wind speed, temperature, humidity, and cloud cover. These comprehensive climate variables give critical insights into solar and wind resource availability and variability at each specific site. For example, the solar irradiance data (direct, diffuse, and global horizontal irradiance) enables analysis of solar patterns to estimate photovoltaic energy production and inform solar panel selection and placement. Temperature data also impacts solar panel efficiency, as higher temperatures reduce performance, but this effect can be mitigated by selecting panels with a low temperature coefficient. Similarly, wind speed data helps optimize wind turbine siting and design for maximum energy capture.

In summary, the extensive site-specific weather data from NSRDB allows for tailoring and optimizing RE systems at each location to leverage environmental conditions and maximize efficiency and reliability. The data provides the initial foundations to analyze renewable potential and make informed technology choices for these desalination facilities.

C. Data Analysis and Significance

After obtaining and cleaning the data, analysis aimed to uncover patterns, trends, and correlations was conducted to inform optimization of RE systems for water desalination plants.

One of the key observations from the data was that some plants had periods in the year when they were not operational, demonstrated by the dips in treated water output as shown in the SAWS data in Fig. 2. These operational breaks were due to maintenance, upgrades, or seasonal fluctuations in water demand. These periods of inactivity can impact the energy demand of the plant and, consequently, the sizing of the RE systems. At the same time, an opportunity for flexible operation exists for all plants to reduce energy system sizing and costs via energy efficiency strategies.

The complementary relation of solar and wind energy is evident from Fig. 1, where the Global Horizontal Irradiance peaks in May, while wind speed is at its lowest. This suggests the potential for a balanced energy supply using both sources.

Temperature trends, also depicted in Fig. 1, highlight potential increased energy demands during summer months, are crucial for appropriately sizing of RE systems.

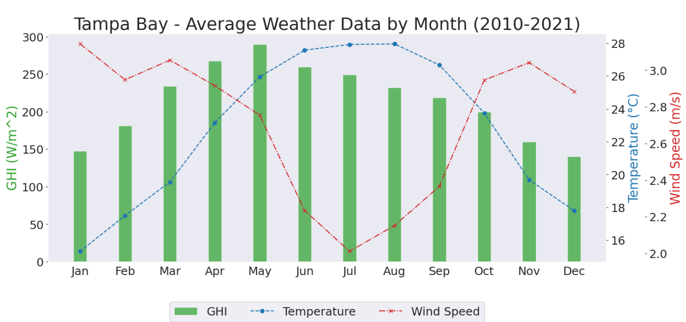

Fig. 1: Average Monthly Weather Data. The bar graph represents the average monthly Global Horizontal Irradiance (GHI) over an 11-year period. The line graphs represent the average monthly temperature and wind speed over the same period.

The relationship between energy usage (measured in kWh) and the volume of treated water is a fundamental aspect of water treatment processes. The energy-intensive nature of these processes means that a larger volume of treated water directly corresponds to increased energy consumption. Conversely, when the volume of treated water decreases, energy usage decreases as well. This strong correlation is clearly demonstrated in Fig. 2, which shows trends of energy use per treated water volume for the San Antonio Water System. The graph provide valuable insights into operations of a desalination facility, highlighting efficiency opportunities.

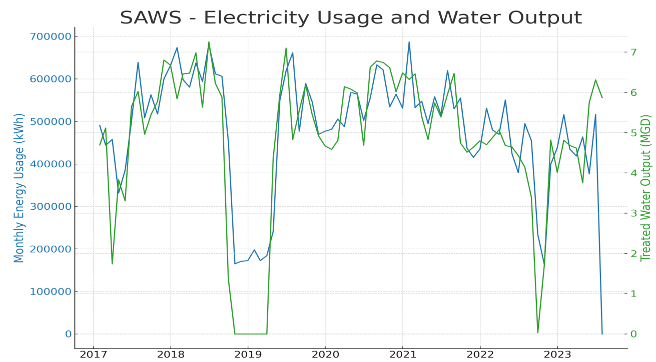

Fig. 2: Energy Usage vs. Treated Water Output Flows. This figure illustrates the relationship between energy usage (in kWh) and treated water output (in Million Gallons per Day).

Our analysis of desalination processes reveals significant correlations between operational parameters. Energy use and energy costs show a strong positive correlation (Pearson correlation coefficient) of 0.93, indicating a direct relationship between energy use and operational expenses. Similarly, the treated water flows and energy use have a positive correlation (0.77), highlighting the energy-intensive nature of water distribution. Evidently, the treated water flows and the input water quality measured by turbidity show a negative correlation (-0.30), suggesting that the lower quality of water, the more time needed to treat, thus a lower output rate. These correlations underscore the need to further explore energy optimization and process performance in desalination.

Based on actual empirical data, this analysis aims to offer a significant benefit compared to pilot plants and hypothetical scenarios, which often oversimplify the problem by making broad assumptions. These oversimplifications may not capture the challenging dynamics of actual desalination plant operations. This research paper employs a method that accounts for these variables, allowing for more accurate and reliable results. For example, in a study by Abdelshafy et al. [11], the researchers assumed a solar panel with a 1 kW power rating in a proposed RO plant, while solar panel data shows the average is less than half of that power output [13]. In addition to providing a framework for decarbonization on a large, commercial scale, this study generates a dataset across multiple facilities, allowing for a comparative focus and understanding across multiple types of renewable energy and desalination systems.

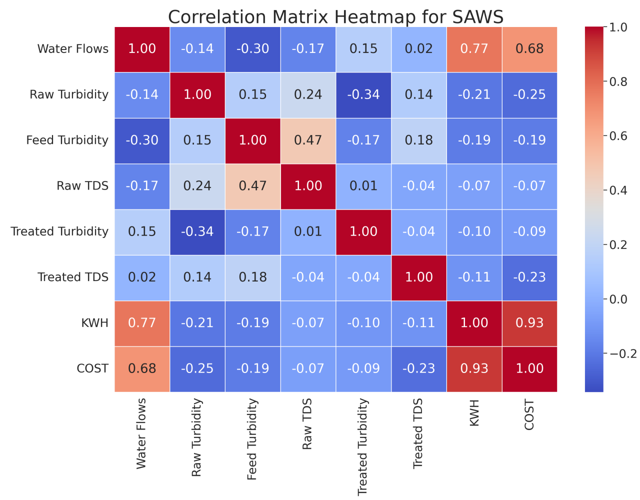

Fig. 3: Correlation Matrices for San Antonio Water System. This heatmap shows the correlation between the plant operation variables given by this desalination plant.

III. Optimization Problem and Constraints

To obtain the optimal system component sizing, we formalize the optimization expressions, decision variables, and constraints for the multi-task problem of minimizing energy costs and carbon footprint while meeting the power requirements of desalination demand to the extent possible by a given array of energy sources. Currently none of the plants have on-site RE generation, though the utilities have a portion of RE generation, which is expected to increase. This study can be used to model and compare the following grid-connected scenarios: (i) 100% renewable energy including storage, (ii) minimum 75% / 50% / 25% RE with the balance being sourced from the grid, and (iii) 100% grid-sourced energy.

When configuring the hybrid renewable energy system, we aim to incorporate state-of-the-art PV panels, wind turbines, and other advanced technologies that can significantly enhance energy generation, thereby expediting the decarbonization process. Additionally, we foresee improvement in power capacity and efficiency; and a decline in the costs of diverse renewable energy systems in the coming years.

A. Objective Function and Decision Variables

The objective is to minimize the total annualized cost and CO2 emissions of the hybrid renewable desalination system. The decision variables are:

- Number of PV arrays (N_{pv})
- Number of wind turbines (N_w)
- Number of concentrated solar panels (N_{csp})
- Number of geothermal units (N_g)
- Number of hydro turbines (N_h)
- Number of battery storage units (N_{batt})
- Number of hydrogen storage units (N_{h2})
- Number of inverters (N_{inv})
- Number of fuel cells (N_{fc})
- Number of buses (N_{bus})
- Number of electrolyzers (N_{elec})

The Levelized Cost of Energy (LCOE) of each RE system is given by:

$$LCOE = \frac{CRF \times CAPEX + FOM}{CF} + VOM \tag{1}$$

where CRF is the capital recovery factor, CAPEX is the capital expenditures, FOM is the fixed operations and maintenance costs, CF is the capacity factor, and VOM is variable operations and maintenance costs.

The total carbon emissions (E_{total}) from each RE source and the utility are also calculated. This is achieved by multiplying the power production of each source by its respective Life Cycle Emissions (LCE) constant:

$$E_{total} = \sum_{i \in \{pv, wind, csp, geo, hydro, utility\}} P_i \times C_i \tag{2}$$

where P_i is the power production and C_i is the LCE constant for each source.

In these formulas, P represents the power production of each source, and C represents the LCE constant for each source. The specific LCE constants are: $C_{pv} = 43$, $C_{wind} = 13$, $C_{csp} = 28$, $C_{geo} = 37$, $C_{hydro} = 21$, and $C_{utility}$ varies by location (average of 422.5).

B. Utility Functions of Decision Variables

We compute the energy output to meet power supply constraints and to reduce energy costs using the utility functions described below, which depend on weather and design parameters.

a) Wind Turbine Power: The power output of a wind turbine, P_{wind} (W), is given by:

$$P_{\text{wind}} = \frac{\rho A v^3 C_p}{2} \tag{3}$$

where:

- ρ = Air density ($\frac{kg}{m^3}$)
- A = Rotor swept area (m^2)
- v = Wind speed ($\frac{m}{s}$)
- C_p = Power coefficient

b) Photovoltaic (Solar) Power: The power output of a photovoltaic (PV) panel, P_{pv} (W), is given by:

$$P_{\text{pv}} = \eta_{\text{inv}} \cdot \eta_{\text{B}} \cdot \eta_{\text{r}} \cdot T_c \cdot \frac{A_{\text{PV}}}{\text{module}} \cdot I \tag{4}$$

where:

- η_{inv} = Inverter efficiency
- η_{B} = Battery efficiency
- η_{r} = Rated solar cell efficiency
- T_c = Temperature correction factor
- A_{PV} = Area of each module (m^2)
- I = Average daily insolation

c) Concentrated Solar Power: The power output P_{CSP} of Concentrated Solar Power (CSP) systems is given by:

$$P_{\text{CSP}} = AG\eta CF \tag{5}$$

where:

- A = Area of the solar collector
- G = Solar radiation incident on the collector (solar irradiance)
- η = Efficiency of the solar collector
- CF = Capacity factor (considers system's operational hours and downtime)

d) Geothermal Power: The power output of a geothermal power plant P_{GT} is given by:

$$P_{\text{GT}} = \eta m c \Delta T \tag{6}$$

where:

- η = Geothermal system efficiency
- m = Mass flow rate of the geothermal fluid
- c = Specific heat capacity of the fluid
- ΔT = Change in temperature of the fluid

e) Hydro Power: The power output of a hydroelectric power plant P_H is given by:

$$P_H = \eta \rho g h Q \tag{7}$$

where:

- η = Efficiency of the turbine and generator
- ρ = Density of water (approx. 1000 $\frac{kg}{m^3}$)
- g = Acceleration due to gravity (approx. 9.81 $\frac{m}{s^2}$)
- h = Height of the water column (hydraulic head)
- Q = Flow rate of the water

C. Constraints

The optimization is subject to the following constraints[1]:

1) All decision variables must be greater than or equal to zero:

$$N_{pv}, N_w, N_{csp}, N_g, N_h, N_{batt}, N_{h2}, N_{inv}, N_{fc}, N_{bus}, N_{elec} \geq 0 \tag{8}$$

2) Renewable fraction (RF) must be greater than a minimum percentage (50%):

$$RF = \frac{P_{pv} + P_w + P_{csp} + P_{geo} + P_{hyd}}{P_{total}} \geq 0.5 \tag{9}$$

3) Maximum discharge depth for battery storage (e.g. 80%):

$$DOD \leq 0.8 \tag{10}$$

4) RO plant capacity limits:

$$P_{FC_{min}} \leq P_{FC} \leq P_{FC_{max}} \tag{11}$$

IV. Discussion and Concluding Remarks

This comprehensive study on four U.S. desalination plants reveals substantial potential for implementing hybrid renewable energy systems. By utilizing renewable energy sources, the current annual output of 32 billion grams of CO2 can be reduced by 50%, equivalent to 16 billion gallons or 16,000 metric tons of CO2 reduced annual emissions, akin to the impact of planting 600,000 trees [14]. This optimized framework offers a viable solution applicable to other water treatment and desalination facilities, promoting eco-friendly practices and sustainable water production.

AI models like Multi-Objective Particle Swarm Optimization (MOPSO), Multi-Objective Genetic Algorithm (MOGA), and Deep Reinforcement Learning (DRL) can enhance the efficiency of hybrid RE systems, optimizing component sizing, and reducing energy costs and CO2 emissions.

Implementing AI models in real-world desalination plants can reduce carbon footprint and boost operational efficiency. Further research should focus on context-specific AI optimization and advanced models for integrating multiple renewable energy sources and energy storage management.

In conclusion, this study demonstrates the feasibility and benefits of adopting hybrid RE systems in functioning desalination plants, providing a roadmap for other facilities. AI optimization ensures dynamic system adjustments based on environmental changes, contributing to a greener future. This research catalyzes the move from fossil fuels to renewable energy in water desalination, offering a scalable and sustainable solution to address global water production needs.

[1] Additional constraints may be added for individual sites and energy types reflecting the total space available and operational simplicity to implement renewable energy systems. For example, the Tampa Bay facility was built in 2007 after Tampa Bay Water was required to reduce the use of groundwater dramatically. The plant has since had occasional shutdowns due to issues with membrane maintenance and cost, but continues to run on a consistent basis and support the town with clean water production.

References

[1] M. H. Saray and A. T. Haghighi, "Energy analysis in water-energy-food-carbon nexus," *Energy Nexus*, p. 100223, 2023.

[2] J. M. Gordon and T. C. Hui, "Thermodynamic perspective for the specific energy consumption of seawater desalination," *Desalination*, vol. 386, pp. 13–18, 2016.

[3] A. Tal, "Addressing desalination's carbon footprint: the israeli experience," *Water*, vol. 10, no. 2, p. 197, 2018.

[4] Jun 2021. [Online]. Available: https://www.danfoss.com/en/about-danfoss/articles/dhs/the-carbon-footprint-of-potable-water/

[5] Mar 2012. [Online]. Available: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2012/IRENA-ETSAP-Tech-Brief-I12-Water-Desalination.pdf

[6] E. T. Sayed, A. Olabi, K. Elsaid, M. Al Radi, R. Alqadi, and M. A. Abdelkareem, "Recent progress in renewable energy based-desalination in the middle east and north africa mena region," *Journal of Advanced Research*, vol. 48, pp. 125–156, 2023.

[7] J. A. Carta and P. Cabrera, "Optimal sizing of stand-alone wind-powered seawater reverse osmosis plants without use of massive energy storage," *Applied Energy*, vol. 304, p. 117888, 2021.

[8] A. K. S. Maisanam, B. Podder, A. Biswas, and K. K. Sharma, "Site-specific tailoring of an optimal design of renewable energy system for remote water supply station in silchar, india," *Sustainable Energy Technologies and Assessments*, vol. 36, p. 100558, 2019.

[9] B. Rahimi, H. Shirvani, A. A. Alamolhoda, F. Farhadi, and M. Karimi, "A feasibility study of solar-powered reverse osmosis processes," *Desalination*, vol. 500, p. 114885, 2021.

[10] F. Rodríguez, A. Fleetwood, A. Galarza, and L. Fontán, "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control," *Renewable energy*, vol. 126, pp. 855–864, 2018.

[11] A. M. Abdelshafy, H. Hassan, and J. Jurasz, "Optimal design of a grid-connected desalination plant powered by renewable energy resources using a hybrid pso–gwo approach," *Energy conversion and management*, vol. 173, pp. 331–347, 2018.

[12] J.-P. Lai, Y.-M. Chang, C.-H. Chen, and P.-F. Pai, "A survey of machine learning models in renewable energy predictions," *Applied Sciences*, vol. 10, no. 17, p. 5975, 2020.

[13] V. Aggarwal, "What is the output of a solar panel?: Energysage," 2022, accessed: 2023-08-03.

[14] A. Moseman, "How many new trees would we need to offset our carbon emissions?" https://climate.mit.edu/ask-mit/how-many-new-trees-would-we-need-offset-our-carbon-emissions, 2022, accessed: 2023-08-03.