

## Forecasting solar still performance from conventional weather data variation by machine learning method

**Authors:** Gao Wenjie

**Date:** 2022-05-30T00:00:00+00:00

### Abstract

Solar stills constitute an effective approach for addressing potable water scarcity; however, reliable methods for forecasting their productivity remain lacking. Herein, a convenient forecasting model requiring only conventional weather forecast data as input is proposed. The model is established using random forest machine learning methodology and optimized via Bayesian algorithms. Training data were obtained from daily measurements conducted over a 9-month period. To validate model accuracy, coefficients of determination for two types of solar stills were calculated as 0.935 and 0.929, respectively, substantially exceeding those of multiple linear regression (0.767) and traditional models (0.829 and 0.847). Moreover, by applying the model, freshwater production was predicted for four cities in China. The reliability of these predictions was confirmed by a high correlation coefficient (0.868) between predicted production and solar insolation. This forecasting model is expected to significantly promote the global application of solar stills.

### Full Text

#### Preamble

#### Forecasting Solar Still Performance from Conventional Weather Data Variation by Machine Learning Method

Wenjie Gao<sup>1</sup>, Leshan Shen<sup>2,3</sup>, Senshan Sun<sup>1</sup>, Guilong Peng<sup>1</sup>, Zhen Shen<sup>2,3</sup>, Yunpeng Wang<sup>1</sup>, AbdAllah Wagih Kandeal<sup>6</sup>, Zhouyang Luo<sup>2,3</sup>, A.E. Kabeel<sup>7</sup>, Jianqun Zhang<sup>4</sup>,, *Hua Bao*<sup>5</sup>,, Nuo Yang<sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Coal Combustion, and School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>2</sup> Key Laboratory of Solar Energy Utilization & Energy Saving Technology of Zhejiang Province, Hangzhou, 311121, China

<sup>4</sup> Mechanical Engineering Department, Faculty of Engineering, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

<sup>5</sup> Mechanical Power Engineering Department, Faculty of Engineering, Tanta University, Tanta, Egypt

**Corresponding authors:** N.Y. (nuo@hust.edu.cn); H.B. (hua.bao@sjtu.edu.cn); J.Z. (zhangjianq888@163.com)

**Highlights** - This work proposes a method to forecast the freshwater production of solar stills using conventional weather data. - The dataset was obtained from eight months of evaporation experiments and measurements. - The forecasting model is established by machine learning and achieves much higher accuracy than traditional models. - By applying the model, freshwater productions of four cities were predicted with high accuracy from their weather data.

## Abstract

Solar stills are considered an effective method to address the scarcity of drinkable water. However, a reliable way to forecast their production has been missing. Herein, we propose a convenient forecasting model that requires only conventional weather forecasting data as input. The model is established using the random forest machine learning method and optimized by a Bayesian algorithm. The required training data was obtained from daily measurements over nine months. To validate the model's accuracy, the determination coefficients for two types of solar stills were calculated as 0.935 and 0.929, respectively, which are much higher than those of both multiple linear regression (0.767) and traditional models (0.829 and 0.847). Moreover, by applying the model, we predicted the freshwater production of four cities in China. The predicted production was confirmed to be reliable by a high correlation value (0.868) between the predicted production and solar insolation. This forecasting model will greatly promote the global application of solar stills.

**Keywords:** Solar still; Production forecasting; Forecasting model; Weather data; Random forest

## Nomenclatures:

RF - Random forest  
BOA - Bayesian optimization algorithm  
 $R^2$  - The determination coefficient of the model  
WS - Wind speed (m/s)  
WD - Wind direction  
Press. - Atmospheric pressure (Pa)  
T - Air temperature ( $^{\circ}\text{C}$ )  
Tmax - Maximum value of air temperature ( $^{\circ}\text{C}$ )

Tmin - Minimum value of air temperature (°C)

RH - Relative humidity (%)

AQI - The air quality index

BIF-SS - The solar still with an interfacial evaporation structure at the bottom and insulation foams at the sidewall

BSI-SS - The solar still with an interfacial evaporation structure at both the bottom and the sidewall

## 1. Introduction

Seawater covers 70% of the earth's surface, while freshwater is mainly distributed in glaciers, ice caps, and underground sources [1, 2]. With increasing population and industrial activities, the shortage of drinkable water has become a catastrophic issue facing the world [3, 4]. As seawater accounts for 97% of water on earth, desalination is an effective solution for freshwater scarcity [5].

Among the many desalination technologies, solar desalination [6] is one of the most environmentally friendly approaches. Fortunately, areas where freshwater is scarce often possess abundant solar energy [7]. The solar still is one such solar desalination technology that is easy to install and maintain [8], with broad application prospects in remote coastal areas and islands. Consequently, solar desalination has received widespread attention in recent years [9-18]. However, the daily production value fluctuates greatly and is significantly affected by climatic conditions, making it difficult to forecast.

Traditional models [19-21] show the functional relationship between production and a few important factors. Due to the complexity of heat and mass transfer in reality, these models with simple functions struggle to accurately describe the heat and mass transfer processes inside solar stills and offer limited guidance for solar still design [22]. Recently, machine learning methods have emerged as an effective way to predict solar still performance [23], such as multiple linear regression (MLR) [24], artificial neural network (ANN) [25, 26], and random forest (RF) [27, 28]. Among current algorithms, RF is an ensemble learning algorithm based on decision trees with unexcelled accuracy [29, 30] and shows excellent performance in prediction tasks [28].

However, previous studies only provided functional relationships between performance and professional parameters such as basin plate temperature, glass cover temperature, and feedwater temperature, which are not convenient for customers to measure. More importantly, previous models cannot forecast production in advance, which represents a significant challenge.

Since production is greatly affected by weather and weather forecast data (such as air temperature, humidity, wind, atmospheric pressure, and air quality index) are readily available, establishing a model between production and weather forecasting data would be a convenient and effective way to forecast production.

Production forecasting is significant for promoting the global application of so-

lar stills. Even in remote areas, conventional weather forecasting is not difficult to obtain. Furthermore, forecasting can help ensure a stable water supply or enable controllable desalination capacity. With the help of forecasting, a proper substitute desalination strategy can be planned and chosen, such as using electrically powered desalination as compensation.

This work aims to develop a model to forecast the daily production of solar stills based on convenient weather data. The required training data was obtained through experimental measurements from July 2020 to March 2021. Based on the production and weather data, the forecasting model was constructed using the random forest method. To verify the model's practicality and accuracy, determination coefficients were calculated and compared. By applying the model, the freshwater production of four Chinese cities was forecasted from conventional weather data.

## 2. Experimental Systems

The solar stills consist of a glass cover, basin, foam heat-insulation layer, water feeding tank, freshwater outlet, and required measuring instruments, as shown in Fig. 1 [Figure 1: see original paper] (a). The bottom dimension is 50\$×\$50 cm. Singh and Tiwari [31] reported that the annual solar still yield reaches a maximum value when the condensing glass cover inclination equals the latitude of the location. Thus, the glass cover of the solar stills has an inclination angle of 30°, which is the preferred solar incidence angle at Hangzhou (120.2° E, 30.3° N). The equipment is installed on the roof of a building in Hangzhou, China. The solar still is placed horizontally with a south-facing front.

The schematic of the solar still is shown in Fig. 1 (b). The solar still has an interfacial evaporation structure at the bottom and insulation foams at the sidewall (BIF-SS). The BIF-SS adopts a three-layer composite structure: a floating light absorption layer, a water-conducting layer, and a heat-insulating layer. The light-absorbing layer structure is made of black deerskin velvet fiber cloth with 95% solar absorption. The water-conducting layer is made of cotton fiber cloth with a thickness of about 8 mm and contacts seawater through the water-conducting channel. The sides and bottom are all wrapped with heat-insulating extruded foam XPS board, 2 cm thick. The thermal conductivity of the XPS board is 0.03 W/m-K. Freshwater is obtained from the freshwater collection tank and recorded manually using a cylinder. The solar still with interfacial evaporation structure is designed based on our previous work [32], which has both high energy efficiency and salt rejection capacity. Meanwhile, a control group was set up using a solar still with an interfacial evaporation structure at both the bottom and the sidewall (BSI-SS). The schematic of BSI-SS is shown in Fig. 1 (c).

**Figure 1** (a) The photo of the solar still system for measurement in Hangzhou. The diagrams of two types of solar stills with an interfacial evaporation structure: (b) at the bottom and the insulation foams at the sidewall (BFI), and (c) on

both the bottom and sidewall (BSI).

The measurements require a series of sensors shown in Table 1. Weather parameters were recorded every minute, including wind speed (WS), wind direction (WD), atmospheric pressure (Press.), air temperature (T), and relative humidity (RH). The air quality index (AQI) data was obtained from the website [www.tianqi.com](http://www.tianqi.com). The recorded weather data for Hangzhou is shown in Fig. 2 [Figure 2: see original paper], expressed as daily average values. Affected by El Niño, the average temperature in August is the highest, significantly higher than in July. Meanwhile, August is the driest month with the lowest average air humidity. Fig. 2 also shows that AQI and atmospheric pressure are higher in winter.

**Table 1** The test platform of meteorological data

Device	Model	Range	Accuracy	Resolution
Wind speed sensor	011E-MetOne	0-60 m/s	$\pm 0.1$ m/s	0.04 m/s
Wind direction sensor	020C-MetOne	0-360°	$< 0.1^\circ$	-
Environmental humidity sensor	HC2S3-Campbell	0-100% RH	$\pm 0.8\%$ RH	0.1% RH
Atmospheric pressure sensor	CS106-Campbell	500-1100 kPa	$\pm 0.3$ kPa	$\pm 0.1$ kPa
Ambient temperature sensor	110PV-Campbell	-40-135°C	$\pm 0.2^\circ\text{C}$	-
Data collector	CR100-Campbell	0-4200 g	0.01 g	-

**Figure 2** The recorded weather data of Hangzhou used as input in the model predicting the production of solar still. (a) Air temperature. (b) Relative humidity. (c) The air quality index. (d) Atmospheric pressure.

Fig. 3 [Figure 3: see original paper] (a) shows the hourly production of the BIF-SS on March 9th. The freshwater productivity gradually increases from 8:00 and reaches its peak at 12:00 at about  $0.8 \text{ kg/m}^2 \cdot \text{h}$ . By 20:00, the productivity approaches zero. Fig. 3 (b) shows the recorded water production of the solar still from July 2020 to March 2021. Affected by weather conditions, daily production varies significantly. The freshwater production in August was the highest and substantially greater than in other months, with the highest daily production reaching  $6.0 \text{ kg/m}^2 \cdot \text{day}$ . The weather and production data are listed in Supporting Materials (SM) I.

**Figure 3** (a) Both the accumulated production (red dots) and the hourly production (black squares) of the BIF solar still on March 9th, 2021. (b) The daily production of BIF-SS measured from July 2020 to March 2021, which forms part of the dataset for building the forecasting model.

### 3. Machine Learning Methods

The forecasting model is established based on the dataset. The solar still dataset is given as  $F = \{X, y\}_1$ , where  $X$  represents the input parameters, including Week, WS, WD, T, Press, RH, and AQI, and  $y$  represents the daily production (the target value corresponding to  $X$ ).

The basic steps include data preprocessing, model construction, and algorithm optimization. Data preprocessing involves scaling the data attributes to a specific range because attributes with larger magnitudes will dominate and affect model accuracy. The standardized method (Z-Scale) is used to scale the input data parameters. The Z-Scale method is based on the mean and standard deviation of the original data, maintaining the sample spacing. After data standardization, the RF method is used to establish the forecasting model.

First, samples are randomly selected and divided into training and test sets. Then, a decision tree is built for each data piece to obtain the predicting result. Finally, all results are voted on to obtain the final result. The Bayesian optimization algorithm (BOA) [28] is used to search for the most appropriate hyper-parameters of the RF model. The diagram of the forecasting model establishment is shown in Fig. 4 [Figure 4: see original paper] (details in SM II).

**Figure 4** The flowchart of making the forecasting model, which includes data preprocessing, model construction, and algorithm optimization.

### 4. Results and Discussions

#### Forecasting Results of RF Model

Fig. 5 [Figure 5: see original paper] shows the performance of the forecasting model for three different cases of testing dataset. The determination coefficient ( $R^2$ ) and mean square error (MSE) are used to evaluate the performance of the forecasting model (details in SM II). With increasing/decreasing training/testing dataset size,  $R^2$  of the random forest models remains at a high level and improves gradually, indicating good model convergence. The  $R^2$  and MSE values are 0.935 and 0.209, respectively, when the test size is 10%.

**Figure 5** For BIF-SS, the predicted values versus the measured values corresponding to different sizes of testing dataset: (a) 10%, (b) 20%, and (c) 30% of the dataset, respectively. The  $R^2$  value is much higher than that of multiple linear regression (0.767).

The  $R^2$  value is much higher than that of multiple linear regression (0.767) and traditional models. For example, Kumar [20] developed a thermal model to predict the exact performance of solar stills for different ranges of Grashof Number, with an  $R^2$  value of only 0.829. In Panchal's work [21], the main parameters of the theoretical model were water temperature and inner glass cover temperature, and the model's  $R^2$  was 0.847. The results in Fig. 5 indicate that the RF method possesses much higher predicting accuracy than traditional models (calculation details in SM III).

### Correlation Between Productions and Weather Parameters

We evaluated the degree of correlation between solar still production and conventional weather forecasting parameters. The random forest method was preferred due to its superior forecasting performance, and the results are shown in Fig. 6 [Figure 6: see original paper]. The three most important parameters are the daily highest temperature (Tmax), relative humidity (RH), and the daily lowest temperature (Tmin), with importance values of 41%, 20%, and 18%, respectively. Moreover, Press., WS, and WD have similar importance values in the range of 2.3% to 3.6%, which is close to that of random orders (2.1%). The random orders were generated randomly, serving as a factor with no correlation to production and used as a baseline for comparison.

This indicates that Tmax, RH, and Tmin are the three most highly correlated factors with production, with Tmax having the highest correlation value. When temperature rises due to increasing solar radiation, the evaporation rate increases. Relative humidity shows a higher degree of correlation because it directly reflects weather conditions and solar radiation. When air humidity is high, conditions are usually cloudy or rainy with low radiation intensity. Besides these three highest-correlated factors, the air quality index has an importance value of 6%. AQI can also affect solar radiation energy; when AQI is high, air quality is poor and particulate matter scatters sunlight, reducing the solar energy entering solar stills.

**Figure 6** The degree of correlation between the production of solar stills and the conventional weather forecasting parameters. The three parameters with highest values are the daily highest temperature (Tmax), relative humidity (RH), and the daily lowest temperature (Tmin), with values of 41%, 20%, and 18%, respectively.

### Forecasting Results Between Different Types of Solar Still

A control group was set up to verify the accuracy and applicability of the predicting RF method. Solar evaporation experiments were conducted on the solar still with an interfacial evaporation structure at both the bottom and the sidewall (BSI-SS). Fig. 7 [Figure 7: see original paper] shows the predicting performance results based on the production data of BSI-SS. The predicting results are comparable to those of BIF-SS. As shown in Fig. 8 [Figure 8: see original paper], when 20% of production data is used for testing accuracy, the  $R^2$  values of BIF

and BSI are 0.927 and 0.939, respectively. These results verify the high accuracy and applicability of the forecasting model.

**Figure 7** The results of the predicting performance between different test sizes (BSI-SS). (a) 10%, (b) 20%, and (c) 30%.

**Figure 8** Comparison of the production values between measured and predicted using 20% of production data as the test set. (a) BIF-SS; (b) BSI-SS.

## 5. Applying Forecasting Model

By applying the forecasting model, freshwater production of four Chinese cities (Wuhan, Hefei, Chongqing, and Linzhi) was calculated and predicted from weather data obtained from the China Meteorological Data Center (<http://data.cma.cn>). The weather data from July 2020 to February 2021 includes air temperature, atmospheric pressure, wind speed and direction, relative humidity, and air quality index. These four cities were selected because they have similar latitudes to Hangzhou ( $\sim 30^\circ\text{N}$ ). Daily productions from July 2020 to February 2021 were calculated and predicted based on daily weather data.

The average daily productions of the four cities are shown in Fig. 9 [Figure 9: see original paper]. The average daily productions in Hefei and Wuhan are similar to that of Hangzhou at  $2.18 \text{ kg/m}^2$  per day, as these three cities have similar latitudes and are located close to the Yangtze River, resulting in similar climates. Chongqing's production is the lowest among these cities at  $2.1 \text{ kg/m}^2$  per day because it is foggy year-round with lower solar radiation intensity than other cities. Linzhi's production is the highest at  $2.48 \text{ kg/m}^2$  per day because it is located on the Qinghai-Tibet Plateau with high altitude (3.1 km) and high insolation. The predicted daily production of the three cities is shown in SM IV.

**Figure 9** The predicted average daily production of five cities in China using the RF model. Linzhi's production is the highest due to its high elevation and insolation, while Chongqing is the lowest due to its dense mist and lower radiation.

Furthermore, daily solar insolation data was obtained from the China Meteorological Data Center to analyze prediction accuracy. A gauge is needed to check the predicted values because there are no measured production values. As noted above, solar insolation was not used in building the model, meaning the solar insolation values are independent of the predicted production. Generally, solar insolation is directly proportional to production, which can serve as a gauge to check predicted values. Fig. 10 [Figure 10: see original paper] shows the comparison of predicted daily production and solar insolation from July 2020 to February 2021 in Wuhan. Due to higher/lower radiation intensity and temperature, production should be higher in summer/winter. The changing trend of predicted production is similar to daily solar insolation, and the correlation

coefficient between the two datasets is 0.868, indicating that the forecasting model possesses high accuracy.

**Figure 10** A comparison of the predicted daily production and the solar insolation from July 2020 to February 2021 in Wuhan. The correlation coefficient between predicted daily production and solar insolation is 0.868, indicating that the forecasting model possesses high accuracy.

## 6. Conclusions

In conclusion, a forecasting model has been built that can forecast freshwater production using convenient weather data. To collect the dataset, a series of solar evaporation experiments were conducted from July 2020 to March 2021 based on two types of solar stills, with production and weather data recorded. The model to forecast solar still production was then established using the random forest method and optimized by the Bayesian optimization algorithm.

The forecasting model demonstrates high accuracy. The determination coefficient ( $R^2$ ) on the training dataset and test dataset can reach 0.946 and 0.935, respectively, when the test size is 10% of production data. A control group was set up to verify the accuracy and applicability of the predicting RF method, and the determination coefficients of the two types of solar stills were calculated as 0.935 and 0.929.

To identify closely related parameters, we also calculated the degree of correlation between production and weather parameters. The three most highly correlated parameters are maximum air temperature, relative humidity, and minimum air temperature, with correlation degrees of 41%, 20%, and 18%, respectively.

By applying the model, productions of four cities were predicted with high accuracy from their weather data. To verify the reliability of the predicted results, the predictions were compared with daily solar insolation data. The correlation coefficient between predicted production and solar insolation is 0.864, indicating that the predictions have high accuracy.

With the help of this forecasting model, the global application of solar stills will be greatly promoted.

**Data Availability Statement.** Replication data and code can be found on the website for this project: [http://nanoheat.energy.hust.edu.cn/Code\\_{22}1.rar](http://nanoheat.energy.hust.edu.cn/Code_{22}1.rar).

## Acknowledgment

This work was sponsored by the National Key Research and Development Program of China (2018YFE0127800), Science, Technology & Innovation Funding Authority (STIFA), Egypt grant (40517), China Postdoctoral Science Foundation (2020M682411), and Fundamental Research Funds for the Central Universities (2019kfyRCPY045).

## References

- [1] El-Samadony Y, Kabeel A E. Theoretical estimation of the optimum glass cover water film cooling parameters combinations of a stepped solar still. *Energy*, 2014, 68:744-750.
- [2] Abujazar M S S, Fatihah S, Lotfy E R, et al. Performance evaluation of inclined copper-stepped solar still in a wet tropical climate[J]. *Desalination*, 2018, 425: 94-103.
- [3] Kabeel A E, Arunkumar T, Denkenberger D C, et al. Performance enhancement of solar still through efficient heat exchange mechanism- A review. *Applied Thermal Engineering*, 2017, 114:815-836.
- [4] Kumar P V, Kumar A, Prakash O, et al. Solar stills system design: A review[J]. *Renewable and sustainable energy reviews*, 2015, 51: 153-181.
- [5] Elimelech M, Phillip W A. The Future of Seawater Desalination: Energy, Technology, and the Environmen. *Science*, 2011, 333(6043):712-717.
- [6] Guilong Peng, Swellam W. Sharshir, Yunpeng Wang, et al. Potential and challenges of improving solar still by micro/nano-particles and porous materials- A review. *Journal of Cleaner Production*, 2021, 311(9):127432.
- [7] Mekonnen M M, Hoekstra A Y. Four billion people facing severe water scarcity. *Science Advances*, 2016, 2(2):e1500323-e1500323.
- [8] Swellam W. Sharshir, Guilong Peng, Ammar H. Elsheikh, et al. Influence of basin metals and novel wick-metal chips pad on the thermal performance of solar desalination process. *Journal of Cleaner Production*, 2020, 248, 119-224.
- [9] Shalaby S M, Sharshir S W, Kabeel A E, et al. Reverse osmosis desalination systems powered by solar energy: Preheating techniques and brine disposal challenges-A detailed review. *Energy Conversion and Management*, 2022, 251: 114971.
- [10] Kandeal A W, El-Shafai N M, Abdo M R, et al. Improved thermo-economic performance of solar desalination via copper chips, nanofluid, and nano-based phase change material. *Solar Energy*, 2021, 224: 1313-1325.
- [11] Qian Chen, Alrowais R, Burhan M, et al. A self-sustainable solar desalination system using direct spray technology. *Energy*, 2020,205:118037.
- [12] Gao M, Zhu L, Peh C K, et al. Solar absorber material and system designs for photothermal water vaporization towards clean water and energy production. *Energy & Environmental Science*, 2019, 12(3): 841-864.
- [13] Guilong Peng, Shichen Deng, Swellam W. Sharshir, et al. High efficient solar evaporation by airing multifunctional textile. *International Journal of Heat and Mass Transfer*, 2019, 147:118866.

- [14] Chen S, Zhao P, Xie G, et al. A floating solar still inspired by continuous root water intake. *Desalination*, 2021, 512: 115133.
- [15] Rahmani A, Kemmar F, Saadi Z. Experimental investigation on the negative effect of the external condenser on the conventional solar still performance. *Desalination*, 2021, 501:114914.
- [16] Guo Y, Dundas C M, Zhou X, et al. Molecular engineering of hydrogels for rapid water disinfection and sustainable solar vapor generation. *Advanced Materials*, 2021, 33(35): 2102994.
- [17] Swellam W. Sharshir, Guilong Peng, Wu L, et al. Enhancing the solar still performance using nanofluids and glass cover cooling: Experimental study. *Applied Thermal Engineering*, 2017, 113:
- [18] Cheng D, Gong W, Li N. Response surface modeling and optimization of direct contact membrane distillation for water desalination. *Desalination*, 2016, 394: 108-122.
- [19] R. Dunkle. Solar water distillation the roof type still and a multiple effect diffusion still. *International Development in Heat Transfer*, 5 (1961) 895-902.
- [20] S. Kumar, G.N. Tiwari, Estimation of convective mass transfer in solar distillation systems. *Sol. Energy*, 57 (1996) 459-464.
- [21] Panchal H. Performance investigation on variations of glass cover thickness on solar still: experimental and theoretical analysis[J]. *Technology and Economics of Smart Grids and Sustainable Energy*, 2016, 1(1): 1-11.
- [22] C. Elango, N. Gunasekaran, K. Sampathkumar, Thermal models of solar still—A comprehensive review. *Renewable Sustainable Energy Rev.*, 47 (2015) 856-911.
- [23] Ham A, Mbb C, D A, et al. Performance estimation of a mini-passive solar still via machine learning. *Renewable Energy*, 2020, 162:489-503.
- [24] Mashaly AF, Alazba AA. MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment. *Computers and Electronics in Agriculture*. 2016;122:146-55.
- [25] Ren Y-S, Lei L, Deng X, Zheng Y, Li Y, Li J, et al. Novel application of neural network modelling for multicomponent herbal medicine optimization. *Scientific Reports*. 2019;9:15442
- [26] Pei J, Deng L, Song S, Zhao M, Zhang Y, Wu S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature*. 2019;572:106.
- [27] Belmokre A, Mihoubi M K, D Santillán. Analysis of Dam Behavior by Statistical Models: Application of the Random Forest Approach. *KSCE Journal of Civil Engineering*, 2019(2).
- [28] Yunpeng Wang, A.W.Kandeal, Swellam W. Sharshir, et al. Prediction of tubular solar still performance by machine learning integrated with Bayesian

optimization algorithm. Applied Thermal Engineering.2020;186: 116233.

[29] Svetnik V. Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information & Computer Sciences, 2003;43(6):1947-58

[30] Chan C W, Paelinckx D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 2008, 112(6):2999-3011.

[31] Singh H N, Tiwari G N. Monthly performance of passive and active solar stills for different Indian climatic conditions. Desalination, 2004, 168:145-150.

[32] Shi J, Luo X, Liu Z, et al. Efficient and antifouling interfacial solar desalination guided by a transient salt capacitance model. Cell Reports Physical Science, 2021, 2(2):100330.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv –Machine translation. Verify with original.*