SUPPORTING INFORMATION

S1. Experimental Data Acquisition

The water vapor flux and salt rejection data used for training the machine learning (ML) models was obtained from the 36 h membrane distillation (MD) operation. The membrane distillation was enabled by three different flat sheet membranes labelled PVDF, CNF/PVDF, and Cu+CNF/PVDF. The membranes were fabricated using the phase separation technique. The thickness of the membranes was maintained at $50\pm5\ \mu$m. The fabricated membranes performance was tested using a custom designed membrane distillation module with an active membrane area of 4 cm^2. The feed containing 3.5 wt% NaCl solution was maintained at 70°C and the permeate was deionized water maintained at 20°C. The feed and the permeate were separated inside the MD module using the fabricated hydrophobic membranes. The weight of the permeate overflow and conductivity was collected real time for the duration of the experiment using a weigh scale and conductivity meter respectively. This real time data was used to calulate the water vapor flux and salt rejection rate.

S2. Non-stationary Behavior of MD Experimental Data

The Cu+CNF/PVDF water vapor flux data obtained from MD experiments is shown in Fig **S1** as an example to demonstrate the non-stationarity of the data being trained for statistical forecasting. Augmented Dickey-Fuller (ADF) was adopted to examine the stationarity using null and alternative hypothesis. Null hypothesis, wherein, water and salt flux time series (TS) data were assumed non-stationary (α=1) while alternate hypothesis assumed data to be stationary in accordance to the equation:

$$Y_t = C + \beta t + aY_{t-1} + \phi\Delta Y_{t-1} + \phi_2\Delta Y_{t-2.....} + \phi_p\Delta Y_{t-p.....} + e_t \qquad (Equation\ 1)$$

where Y_t, is the value of the TS at time t, α is a constant, β is the coefficient on a time trend and p is the lag order of α. Based on the resultant p values that were lower than the leading to elimination of the null hypothesis, the data series were inferred to be stationary. Table S1 reports the results of the Augmented Dickey Fuller Test (ADF Test) and computed p-values that were greater than the significance level (>=0.05) leading to non-stationarity.

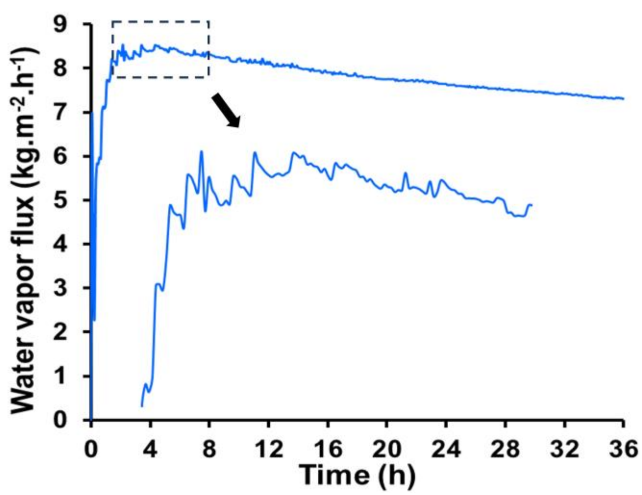

Figure S1: Non-stationary behavior of water vapor flux data obtained from Cu+CNF/PVDF MD experiments and collected for 36 hours.

Table S1: Summary of ADF test parameters for water vapor flux and salt rejection of Cu+CNF/PVDF membrane.

Test Parameters		Value
ADF test statistics	Vapor flux	3.19
	Salt rejection	-0.57
P-value	Vapor flux	1
	Salt rejection	0.87
#Lags used	Vapor flux	17
	Salt rejection	13
Stationary	Vapor flux	No
	Salt rejection	No

S3. Statistical Forecasting: Autoregressive Integrated Moving Average Model

An ARIMA *(p, d, f)* model was then built to convert the data to stationary by defining autoregressive (AR) represented as *p*, moving average (MA) denoted by *q* and the number of times of differencing*, d* performed for water vapor flux and salt rejection data. ARIMA model was implemented using (Seabold and Perktold 2010) and Auto_arima in Python library to obtain an optimal size of ARIMA. Table S2 summarizes ARIMA model parameters (p,d,f) and the resultant error metrics based on probabilistic statistical measures: Akaike information criterion (AIC) (Bozdogan 1987) and the Bayesian information criterion (BIC) for method scoring and model selection. The lower AIC BIC value indicates a better fit. The log-likelihood value is a simpler representation of the maximum likelihood estimation.

Table S2: Optimal ARIMA parameters (p, d, f.) for training water vapor flux and salt rejection data.

Target	Cu+CNF/PVDF	
	Vapor Flux	**Salt Rejection**
ARIMA	(3,2,1)	(3,2,1)
Log likelihood	33.68	2775.39
AIC	-53.36	-5538.70
BIC	-24.86	-5514.33

S3.1 Akaike Information Criterion (AIC) and size of ARIMA

The "size" of an ARIMA model referred to by the order of its components: AR (autoregressive), I (integrated), and MA (moving average) or ARIMA(p, d, q) are enlisted in Table S3. Where, p specifies the number of lags; d, the degree of differencing; and q, the order of the moving average part. This is the number of lagged forecast errors that are used to predict the future values. These were calculated fitting AIC model to compare the fit of ARIMA model as summarized in Table S3. As observed, AIC can be positive or negative. The negative AIC value obtained for all salt rejection rates simply means that the likelihood of the model is greater than 1 which happen when the number of data points is much larger than the number of parameters in the model. The absolute value of the AIC are insignificant, but the lower AIC values are desirable for the better fitting model, and therefore, ARIMA is not suited for this study.

Table S3: Different size of ARIMA and the AIC error associated with ARIMA model.

Quantity	ARIMA	AIC
Vapor flux	(1,0,1)	535.01
	(2,0,1)	272.86
	(3,0,2)	109.42
	(2,0,2)	58.30
	(1,0,2)	394.81
	(1,0,3)	106.56
Salt rejection	(1,1,1)	-4999.80
	(0,1,0)	-5300.03
	(1,1,0)	-5317.36
	(3,1,2)	-4692.67
	(2,1,2)	-5400.91
	(3,1,1)	-5533.40

S4. Statistical Forecasting: Exponential Smoothing (ES)

Additionally, a triple Exponential Smoothing comprising: value, trend, and seasonality components was also tested based on Overall, Trend and Seasonal smoothing represented by following correlations:

$$s_t = \ a\ \frac{y_t}{I_{t-L}} + (1-a)(s_{t-1} + b_{t-1}) \qquad (equation\ S2)$$

$$b_t = \ \gamma(s_t - s_{t-1}) + (1-\gamma)_{b_{t-1}} \qquad (equation\ S3)$$

$$I_t = \beta\ \frac{y_t}{S_t} \ + (1-B)_{I_{t-L}} \qquad (equation\ S4)$$

Where, *y* is the observation, L is the period, *s* is the smoothed observation, *b* corresponding to trend factor. *I* and *t* represent the seasonal and time indices respectively Additionally, *α, β, γ* are the parameters that need to be determined through estimation. Fig 2b shows the comparison between the water vapor flux test data of Cu+CNF membrane and water vapor flux obtained by tuning *α, β,* and *γ.* The closest prediction to test data was with *α* and *β* equal to 0.3 and *γ* equal to 0.5.

Grid search technique was used to compute Mean Absolute Error (MAE) to measure the average magnitude of the errors in a set of forecasts without considering their direction. Three parameters were used to quantify the exponential smoothing. α determined the level of smoothing coefficient, whereas β represents the smoothing coefficient trend, and γ represents the seasonal smoothing coefficient by adjusting repeating patterns, of the time series data. The lowest values obtained for all three parameters and MAE confirms the suitability of exponential smoothing for time series forecasting studies of water vapor flux and salt rejection data.

Table S4: Summary of α, β, and γ parameters used for Exponential Smoothing and resulatant MAE

Target	Cu+CNF/PVDF	
	α, β, and γ	MAE
Vapor flux	(0.1, 0.1, 0.1)	0.01
	(0.1, 0.1, 0.7)	0.02
	(0.1, 0.7, 0.1)	0.03
	(0.3, 0.7, 0.1)	0.08
	(0.3, 0.9, 0.1)	0.1
Salt rejection	(0.1, 0.7, 0.1)	0.01
	(0.3, 0.9, 0.9)	0.03
	(0.7, 0.7, 0.5)	0.1

S5. Outliers in MD experimental data

Time-series data mostly contains outliers due to the influence of unusual and non-repetitive events. The impact of these outliers can be significant which can cause the reduction of the predictions accuracy, or bias in our models(Chen and Liu 1993). In order to study the significance of outlier's data points in this study, we used a method called Z testing. The Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values. It is measured in terms of standard deviations from the mean. Z score can be formulated as follow:

$$Z = \frac{X - \mu}{\sigma} \qquad \text{(equation } S5\text{)}$$

where X indicates our experimental value, μ is the mean and σ stands for Standard Deviation. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean. The Z-score is being used to identify outliers in the dataset. If a data point has a Z-score that is too high (for example, greater than 3 or -3 in absolute value), it could be considered an outlier because it is too far from the mean compared to the other data points (Aggarwal, Gupta et al. 2019). The Z-scores for water vapor flux and salt rejection experimental data is shown in figure S2 demonstrates the deviation of experimental observations from the most likely outcome, the mean. Based on the Z-scores the impact on accuracy is low due to the low number of outliers observed in the experimental data.

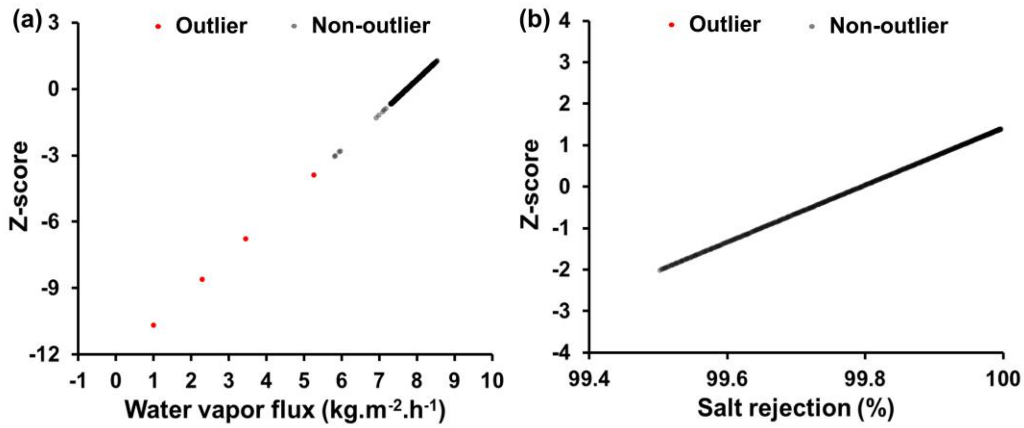

Figure S2: Outlier (Red) identification based on Z-scores for water vapor flux and salt rejection dataset of Cu+CNF/PVDF membrane used for training the time series models.

References

Aggarwal, V., V. Gupta, P. Singh, K. Sharma and N. Sharma (2019). "Detection of spatial outlier by using improved Z-score test." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI): 788-790.https://doi.org/10.1109/ICOEI.2019.8862582

Bozdogan, H. (1987). "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions." Psychometrika 52(3): 345-370.https://doi.org/10.1007/BF02294361

Chen, C. and L. M. Liu (1993). "Forecasting time series with outliers." Journal of forecasting 12(1): 13-35.https://doi.org/10.1002/for.3980120103

Seabold, S. and J. Perktold (2010). "Statsmodels: econometric and statistical modeling with python." SciPy **7**: 1