

---

**Original Paper**

---

# Flow Factor Prediction of Centrifugal Hydraulic Turbine for Sea Water Reverse Osmosis (SWRO)

**Ying Ma, Eric Kadaj, Kevin Terrasi**

Pump Engineering, an Energy Recovery Inc Company  
17608 Commerce Dr, New Boston, MI, USA, 48164,  
mayingth@hotmail.com, ekadaj@pumpengineering.com, kterrasi@pumpengineering.com

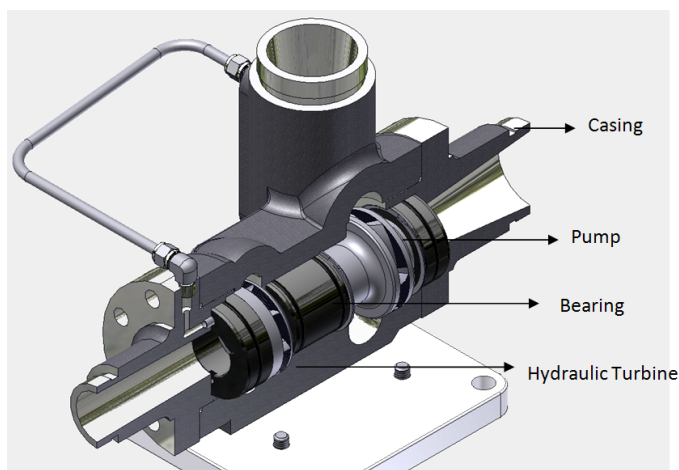
## Abstract

The creation of the hydraulic turbine flow factor map will undoubtedly benefit its design by decreasing both the design cycle time and product cost. In this paper, the geometry and flow variables, which effectively affect the flow factor, are proposed, analyzed and determined. These flow variables are further used to create the operating condition maps by using different model approaches categorized into Response Surface Method (RSM) and Artificial Neural Network (ANN). The accuracies of models created by different approaches are compared and the performances of model approaches are analyzed. The influences of chosen variables and the combination of Principle Component Analysis (PCA) and model approaches are also studied. The comparison results between predicted and actual flow factors suggest that two-hidden-layer Feed-forward Neural Network (FFNN), and one-hidden-layer FFNN with PCA has the best performance on forming this mapping, and are accurate sufficiently for hydraulic turbine design.

**Keywords:** sea water reverse osmosis, hydraulic turbine, flow factor, flow coefficient

## 1. Introduction

A consistent and gradual 10.3% growth rate occurs in the global market for reverse osmosis (RO) membranes and system components [1]. Sea water can be effectively transferred into drinkable water by using the pump to boost the water into high pressure in order to force it through Sea Water Reverse Osmosis (SWRO) membranes. The sea water is separated into two streams by the membranes; the permeate stream and the brine stream. The permeate, or treated water, exits the membrane at low pressure. The brine, or concentrate stream exits the membrane at high pressure. The relative percentages of permeate to brine vary with varying system conditions. Generally, the required feed pressure for SWRO is 800-1000 psi. The higher salt percentage of the water is, the higher the pressure is required by the membranes. The majority of the feed water pressure remains in the brine stream after the membranes. Therefore, a large amount of energy exists in the brine flow. Energy recovery devices, for instance, the hydraulic turbocharger (Fig. 1)



**Fig. 1** 3D Illustration of structure of hydraulic turbocharger

which includes a pump and hydraulic turbine, use the hydraulic turbine to recycle the brine water energy by directly driving the pump, and furthermore greatly decrease the operating cost of SWRO systems.

The two most important performance characteristics of the turbocharger are the efficiency and flow factor of hydraulic turbine. The efficiency plays an important role in the operating cost and is the most important characteristic to consider. However, the flow factor  $Kv$ , also called flow coefficient  $Cv$  in imperial Unit, is also very critical. Compared with the traditional hydraulic turbine, the flow conditions are relatively strict because their operating pressure should match the requirements of SWRO membranes. Incorrect flow factor will result in abnormal operation of SWRO membranes and an undesired recovery ratio. The design of the hydraulic turbine, which guarantees both the efficiency and flow factor, is very challenging and requires the long design cycle and even additional modification after the prototype is tested. Our design files and test records show that the operating condition is more sensitive to the geometry change of the rotors compared with the efficiency. For instance, a 2% change in nozzle size can cause 2% change in the mass flow rate under the same pressure differential while causing less than 1% change in the efficiency. If the hydraulic turbine operating condition can be accurately predicted before the full three dimensional (3D) CFD simulation, it will significantly help to reduce design time and cost, especially for the lower budget or quick delivery projects.

Ghorbanian et al. [2] use four different Multilayer Perceptron (MLP) network and Radial Basis Function Network (RBFN), Regression Neural Network to form axial compressor performance map. Yu et al. [3] use the Feed Forward Neural Network (FFNN) with two hidden layer and back propagation learning algorithm to create compressor's characteristic performance map. Yi et al. [4] use the Response Surface Model and employ a second-order polynomial of the form to create the model between transonic compressor rotor efficiency and rotor geometry variables.

However, no papers have been published on the modeling of operating condition of the hydraulic turbine to the authors' knowledge, especially for SWRO systems. In our study, three groups of hydraulic turbines of different sizes are used, which are referred to as the small, medium and large groups respectively here. First, the impacts of the geometry and flow variables on hydraulic turbine flow factor are studied. Secondly, the geometry variables that effectively influence on flow factor are chosen. Different types of model approaches are used to create the flow factor map. Finally, the accuracies of created models are compared and evaluated with test results, followed by the discussion of the advantages and drawbacks of their corresponding modeling approaches for this specific problem.

## 2. Parameterization

The hydraulic turbine flow condition includes the flow rate and pressure differential. However, flow rate and pressure differential are coupled with the hydraulic turbine operating conditions. It is better to propose a parameter to represent the basic characteristic instead of these two coupled variables. Based on the affinity laws for constant diameter [5], relationships between the flow rate  $Q$  or head  $H$  and rotating speed  $N$  for hydraulic turbine are expressed as:

$$\frac{Q_1}{Q_2} = \frac{N_1}{N_2} \quad (1)$$

$$\frac{H_1}{H_2} = \frac{N_1^2}{N_2^2} \quad (2)$$

Because the head  $H$  is linear with the pressure differential  $\Delta P$  for incompressible fluid. Therefore, eq. (1) and (2) can be written into the relationship between  $\Delta P$  and  $Q$ :

$$\frac{\sqrt{\Delta P_1}}{\sqrt{\Delta P_2}} = \frac{N_1}{N_2} \quad (3)$$

The above equation can also be rewritten into:

$$\frac{Q_1}{\sqrt{\Delta P_1}} = \frac{Q_2}{\sqrt{\Delta P_2}} = \dots = \frac{Q_N}{\sqrt{\Delta P_N}} \quad (4)$$

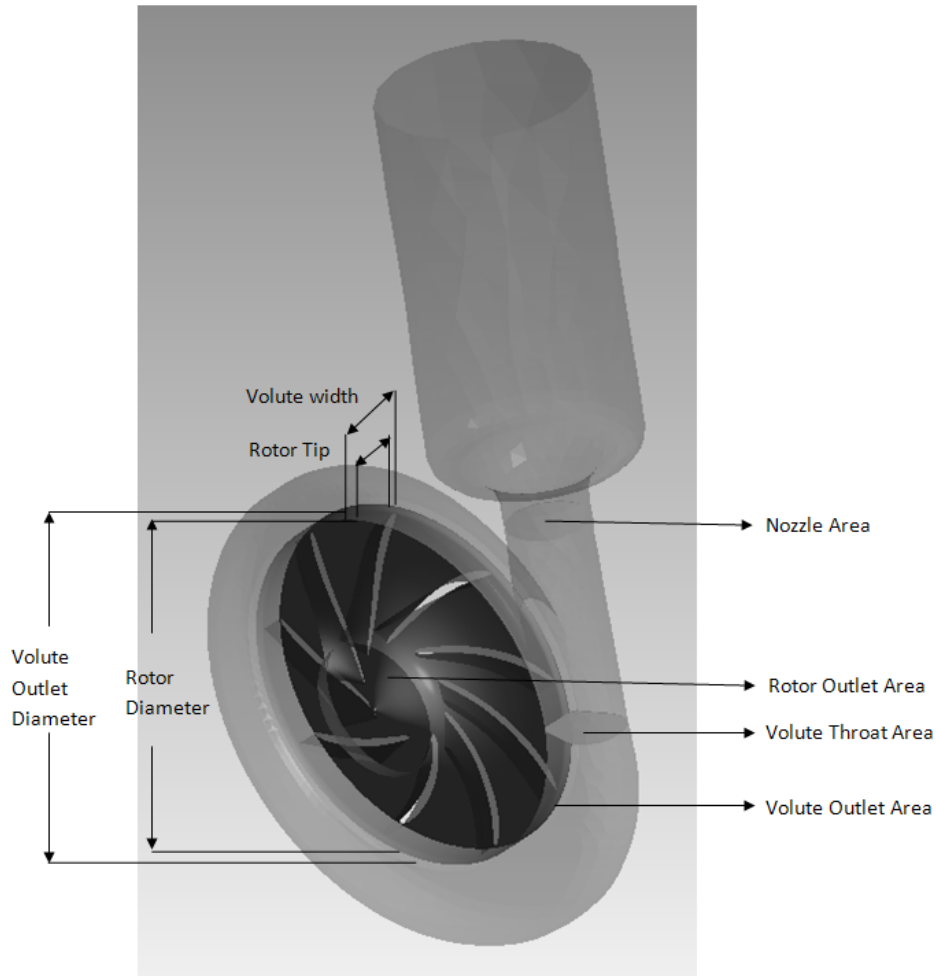
Equation (4) illustrates that the flow rate square is linear with pressure differential. However, the flow factor  $Kv$  remains the same for a turbine with a constant geometry and can be considered as one basic characteristic of the turbine. Therefore the flow factor, which is expressed in eq. (5), is used as a performance parameter to create the modeling map. The physical meaning of the flow factor  $Kv$  indicates the passable flow rate with square root of pressure differential. The higher its value is, the lower flow resistance of the turbine is and the more flow is capable to pass under unit pressure differential. The unit of specific gravity  $\gamma$  is  $1 \text{ kg/m}^3$ . Therefore, the flow factor  $Kv$  is dimensionless.

$$Kv = Q \sqrt{\frac{\gamma}{\Delta P}} \quad (5)$$

The design and experimental data of three groups of the hydraulic turbines with different sizes are used. The basic structures of three groups of turbines are the same, shown in Fig. 1. However, rotor outside diameters in the group of small size vary from 0.0861 to 0.0864 m, those of medium size from 0.0973 to 0.1057 m, and those of large size from 0.1148 to 0.1260 m.

A hydraulic turbine includes three main components: the nozzle, volute and rotor. There are dozens of geometry variables for each component. It is impossible to consider all of these parameters in the modeling. Therefore, the key variables that affect the flow factor should be chosen from them first. Ten potential key variables are chosen to evaluate the effects on the flow factor

based on the expertise, which includes the nozzle area, rotor outlet area, volute throat area, rotor inlet meridional absolute Velocity  $C_m$ , volute outlet area, rotor tip, volute width, volute outlet diameter, rotor outside diameter (OD) and rotor wrap angle. All of these ten potential key variables except the rotor inlet meridional absolute velocity  $C_m$  and rotor wrap angle are illustrated in Fig. 2. The rotor inlet meridional absolute velocity  $C_m$  is the component of absolute velocity in meridional direction at the rotor inlet, and equal to the flow rate divided by the rotor inlet area. The rotor wrap angle, also called the theta angle, represents how many angles of the rotor twists from the leading edge to the trailing edge in the absolute coordinate system  $(R, Z, \theta)$ .



**Fig. 2** 3D Illustration of potential key variables

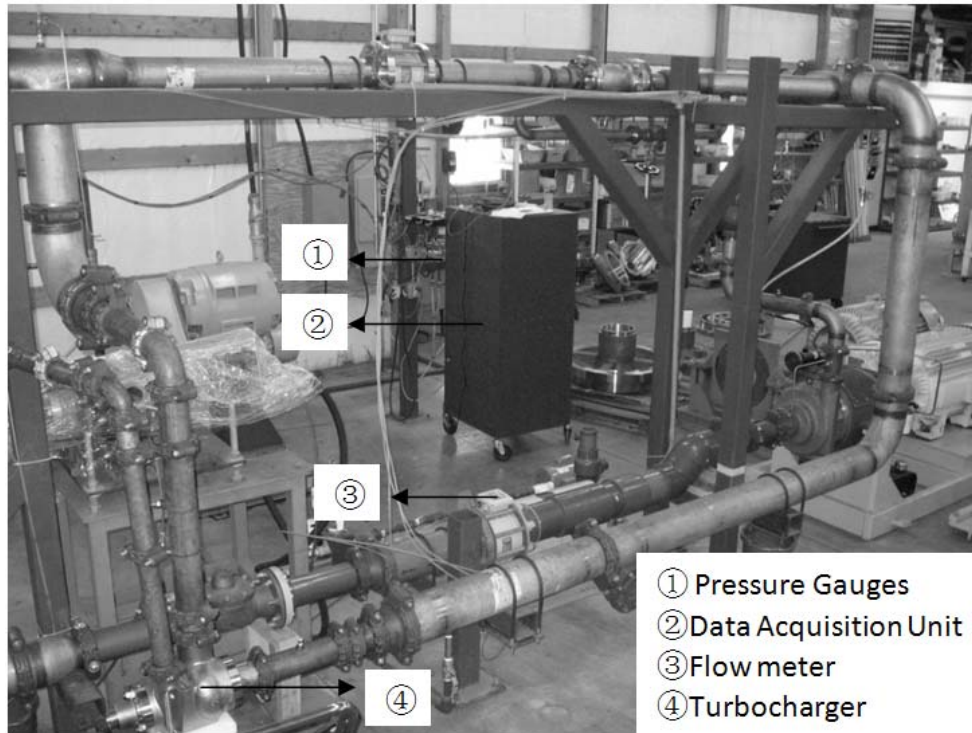
### 3. Selection of Key Variables

The correlation coefficients between each variable and the flow factor are calculated to evaluate the key variables. The formula used to calculate the correlation coefficient is expressed as below, where  $X, Y$  are two vectors:

$$C(X, Y) = \frac{E\{[Y - E(Y)]^T [X - E(X)]\}}{\sqrt{E\{[X - E(X)]^T [X - E(X)]\}} \sqrt{E\{[Y - E(Y)]^T [Y - E(Y)]\}}} \quad (6)$$

$Y$  represents the flow factor, and has only one column.  $X$  represents ten variables and has ten columns.  $E$  denotes mean operator. As mentioned above, three groups of turbines of different sizes are used and named as the group of small, medium and large size groups. There are 47, 46 and 88 units in the small, medium and large size groups. All the parameters including flow factor and key variables have been normalized into  $[-1, 1]$  to calculate the correlation coefficients.

The test loop is shown in Fig. 3. The flow, differential pressure and rotating speed of turbine are self-determined compared to pump. Pump flow, boost pressure and rotating speed can be adjusted independently by changing the frequency of Variable Frequency Drive (VFD) and throttle valve. One parameter can be changed while other two remains the same. As for the turbine, once the loading of turbine is determined, relationship between flow, differential pressure and rotating speed will be obeyed the affinity. And one parameter of three changed will definitely changed other two. The chosen loading is mainly determined to allow rotor to run under the best efficiency point (BEP) speed in the experiments. The loading is changed by the flow entering into the pump, which connects with the turbine. The pressure differential is calculated based on the turbine inlet and outlet pressure measured by two pressure gauges. The flow factor is calculated based on the flow rate measured from the flow meter and calculated pressure differential using eq. (4). Most of the potential key variables are measured from the machined unit except volute width and rotor wrap angle are obtained from the design geometry files because of the limits of the measurement tools. Therefore, there is likely 0.1-0.5% machining errors for these two variables. The correlation coefficients and their corresponding  $p$  values for three different sizes are calculated and listed in Table 1 and 2 respectively.



**Fig. 3** Illustration of test loop

**Table 1** Correlation coefficients between potential key variables and flow factor

Correlation coefficient	Small	Medium	Large
Nozzle area	0.8596	0.9725	0.9129
Rotor outlet area	0.5733	0.9299	0.6854
Volute throat area	0.4234	0.9003	0.8546
Rotor inlet Cm	0.5754	0.4064	0.4982
Volute outlet area	0.6472	0.7835	-0.0638
Rotor tip	0.6553	0.7359	-0.1180
Volute width	0.3795	0.9599	0.8486
Volute outlet diameter	0.4765	0.2134	0.2162
Rotor outside diameter	0.0261	0.6502	0.3033
Rotor wrap angle	0.1313	0.2998	0.1599

**Table 2** p-value of the correlation coefficients between potential key variables and flow factor

Correlation coefficient	Small	Medium	Large
Nozzle area	0.0000	0.0000	0.0000
Rotor outlet area	0.0000	0.0000	0.0000
Volute throat area	0.0034	0.0000	0.0000
Rotor inlet Cm	0.0000	0.0000	0.0000
Volute outlet area	0.0000	0.0000	0.5545
Rotor tip	0.0000	0.0000	0.2734
Volute width	0.0093	0.0000	0.0000
Volute outlet diameter	0.0008	0.1544	0.0430
Rotor outside diameter	0.8633	0.0000	0.0041
Rotor wrap angle	0.3845	0.0429	0.1368

The correlation coefficient is a value in the range from -1 to 1. The absolute value of the correlation coefficient indicates the correlation level. 1 or -1 indicates perfect correlation while the positive suggest two variables vary in the same direction. A correlation value close to 0 indicates no association or independency between two variables. If the absolute value of the correlation coefficient is larger than 0.75, it assumes that a strong correlation exists between two variables. *P*-value is a number

between 0 and 1 to specify a confidence level. 0.05 of p-value, which indicates 95% confidence intervals, is considered as the standard value. When the p-values are above 0.05, the corresponding correlation coefficients are not convincing.

All results of three different groups of hydraulic turbines show that nozzle area, volute throat area, rotor outlet area has significant impact on the flow factor. Moreover, rotor inlet Cm also has a strong correlation with flow factor. Moreover, their maximum corresponding p-values is only 0.0034 and much smaller than 0.05. Therefore, these four variables: nozzle area, volute throat area, rotor outlet area and rotor inlet Cm are chosen as key variables to create the hydraulic turbine operating condition map.

The three groups of hydraulic turbines show the contradictory results on volute outlet area, rotor tip and volute width. The results of small and medium size groups indicate that volute outlet area and rotor tip has the great association with flow factor while large group shows that those two parameters are almost independent with the flow factor. Their corresponding p-values are zero in the small or medium size group while 0.2734 or 0.5545 in the group of large size. Therefore, geometry variable volute outlet area or rotor tip is also considered as the key variable to create the map, which further compare with the map created without consideration of volute outlet area or rotor tip.

Because volute width can be directly calculated from volute throat area and volute outlet area, which are already considered as the key parameter, therefore the volute outlet width is negligible here.

All results show that volute outlet diameter, rotor wrap angle and rotor OD have no association with flow factor, and therefore are not considered as the key variables.

## 4. Modeling Approaches

In this study, two general modeling approaches: Response Surface Methodology (RSM) and Artificial Neural Network (ANN) are applied to create the map between key variables and the flow factor.

### 4.1 Response Surface Methodology

In regard to the Response Surface Methodology (RSM), a second-order polynomial in eq. (7) is used to represent the relationship between key variables  $X_i$  and flow factor  $K_v$ . There are four terms in eq. (7), which are constant, linear, square and interaction ones. The coefficients  $\beta_0$ ,  $\beta_i$ ,  $\beta_{ii}$  and  $\beta_{ik}$  are trained via the least squares regression which minimize the sum of the squares of the deviations between the predict and actual values.

$$R = \beta_0 + \sum_{i=0}^n \beta_i X_i + \sum_{i=0}^n \beta_{ii} X_i^2 + \sum_{i<k}^n \beta_{ik} X_i X_k \quad (7)$$

Models are different if different terms are included. In the current study, five different models with different equations can be generally categorized into multiple linear and nonlinear regressions. The included terms of each RSM model are explained as followed:

Multiple Linear Regression:

$$1^{\text{st}} \text{ RMS: Included the linear terms } R = \sum_{i=0}^n \beta_i X_i$$

$$2^{\text{nd}} \text{ RMS: Included the constant and linear term } R = \beta_0 + \sum_{i=0}^n \beta_i X_i$$

Multiple Nonlinear Regression:

$$3^{\text{rd}} \text{ RMS: Included the constant, linear and cross products terms } R = \beta_0 + \sum_{i=0}^n \beta_i X_i + \sum_{i<k}^n \beta_{ik} X_i X_k$$

$$4^{\text{th}} \text{ RMS: Included the constant, linear and squared terms } R = \beta_0 + \sum_{i=0}^n \beta_i X_i + \sum_{i=0}^n \beta_{ii} X_i^2$$

$$5^{\text{th}} \text{ RMS: Included all terms: } R = \beta_0 + \sum_{i=0}^n \beta_i X_i + \sum_{i=0}^n \beta_{ii} X_i^2 + \sum_{i<k}^n \beta_{ik} X_i X_k$$

### 4.2 Artificial Neural Network

Models are different if different terms are included. In the current study, five different models with different equations can be generally categorized into multiple linear and nonlinear regressions. The included terms of each RSM model are explained as followed:

Artificial Neural Network (ANN), inspired by biological neural system is one type of nonlinear mathematical model. The widely used Artificial Neural Networks (ANNs) for creating maps are Feed-forward Neural Network (FFNN) and Radial Basis Function Network (RBFN). FFNN, also called multi-layer perceptrons, is generally composed of 3-5 interconnected neuron layers: one input layer, one output layer and 1 to 3 hidden layers. Information flow moves only in one direction from the input layer to the output layer without any loops from the subsequent layer to the previous one. The values of neurons in the input and output layers are equal to the values of input and output variables respectively. The value of each neuron in the hidden layers are calculated based on every neuron in the previous layer and also contributes to the value of neurons in the subsequent layer using Equation  $X_i = \sum_j W_{ij} f(X_j) + b_{ij}$ . Where  $X_i$  and  $X_j$  are neurons in the current and previous layers respectively;  $f$  is the activation function and has three categories: linear, threshold and sigmoid. The type of activation function used is determined by structure designers;  $W_{ij}$  and  $b_{ij}$  are the weigh factor and bias between neurons  $X_i$  and  $X_j$  in two different layers. The training of ANN is an adaptive process in which parameters including weights and bias continue to be changed as internal or external information that flow through ANN until the some certain criteria has been achieved. RBFN, another type of ANN, has a less flexible structure compared to FFNN. RBFN typically has only three layers: an input layer, one hidden layer with a nonlinear radial basis function (RBF) used as the activation function, and one linear output layer. The Gaussian RBF  $X_i = \sum_j W_{ij} \exp(-\beta_{ij} \|X_j - C_j\|^2) + b_{ij}$  is widely used

in the hidden layer while the output is a weighted summation of neurons in the hidden layer. Where  $W_{ij}$  is the weight factor;  $\beta_{ij}$  the coefficient of Gaussian Radial Basis Function;  $C_j$  the center vector for neuron in the hidden layer.  $W_{ij}$ ,  $\beta_{ij}$  and  $C_j$  are adaptively changed during the training of RBFN. In current study, FFNN with one hidden layer, FFNN with two hidden layers and RBFN are used to create the model.

### 4.3 Application of Principle Component Analysis

Principal Component Analysis (PCA) is used in this study to improve the modeling accuracy. Physically, PCA is a mathematical technique to transform a high-dimensional space into reduced dimension system, in which the coordinates has the maximum variances. Mathematically, PCA is used to search the eigenvectors related to the largest eigenvalues. In our study, PCA is used as an orthogonal transformation, which preserves the original dimension of data and only transforms data into a new coordinate system on which maximum variances of data are projected.

## 5. Results and Discussions

### 5.1 General Comparison of Different Model Approaches

The eight different model approaches are used. 50% of units in the database in each group are selected in the training database arbitrarily to train the model. And other 50% left are selected into the test database to evaluate the created model. Therefore, each group of small, medium and large sizes has two databases: training and test. The error  $\varepsilon$  between the actual and predicted values of flow factor is calculated for each unit in both the training and test database. The Root Mean Square Errors (RMSEs) of training and test databases are further calculated using eq. (8) and used to evaluate the average performances of models:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i}{n}} \quad (8)$$

The RMSEs of the test databases are used to evaluate the model approaches. Theoretically, the RMSEs of test databases are more important than RMSEs of training databases because they are able to indicate the accuracies of the trained models. The smaller the RMSE of the test database is, the better performance of the model approach has. RMSEs can be used to evaluate the average performances of model approaches. However, the worst situation and the maximum error are also concerned when considering a specific design. Therefore, the model approach with the higher value of RMSE but acceptable maximum value is better than the one with relatively lower value on RMSE but a large value on the maximum error.

Key variables and flow factor are normalized in the range [0, 1] before creating maps. Therefore, the values of errors can physically represent the percentage of the errors and be simply written in percentage format.

The RMSE and maximum errors with four key variables used to create maps are shown in Table 3. In Table 3, the average values of RMSEs in the three groups and maximum errors are calculated based on three groups, which depict the average performance of the model approaches.

**Table 3** Average RMSE and maximum errors with four key variables used to create flow factor maps

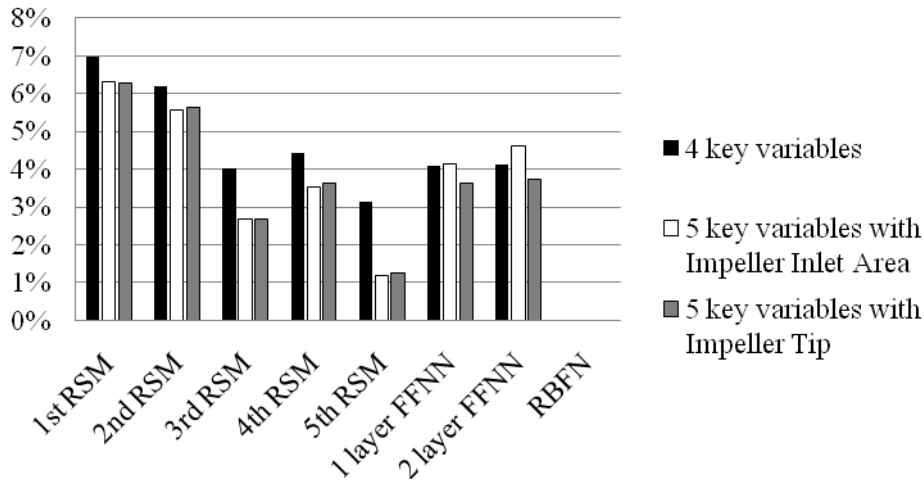
Model approaches	RMSE training	RMSE test	Max error
1 <sup>st</sup> RSM	6.98%	7.62%	22.05%
2 <sup>nd</sup> RSM	6.19%	7.01%	20.12%
3 <sup>rd</sup> RSM	4.03%	18.32%	45.77%
4 <sup>th</sup> RSM	4.44%	7.01%	16.74%
5 <sup>th</sup> RSM	3.14%	8.52%	20.48%
One hidden layer FFNN	4.10%	8.18%	21.43%
Two hidden layer FFNN	4.12%	6.46%	14.94%
RBFN	0.00%	45.73%	141.38%

The results in Tables 3 suggest that two hidden layer FFNNs has the best performance comprehensively while 4<sup>th</sup> RSM, 5<sup>th</sup> RSM and one hidden layer hidden layer FFNNs have the similar performances. 1<sup>st</sup> - 3<sup>rd</sup> RSM has lower performances mainly because there are no square terms. Models created using RBFN has the very low value of RMSE in the training database but very high value in the testing database. Therefore, RBFN over-trains the database and doesn't succeed to create the accurate model.

### 5.2 Comparison between 4 and 5 Key Variables

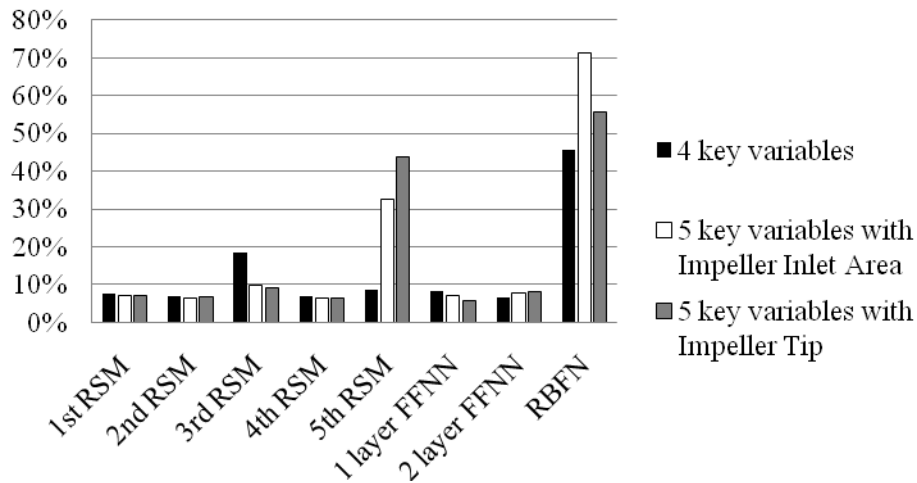
It has been mentioned that rotor inlet area or rotor tip may have an association with the flow factor. To better compare the models between 4 and 5 variables, the average values of RMSEs and maximum errors are calculated and compared each other shown in Fig. 4.

### RMSE of training database



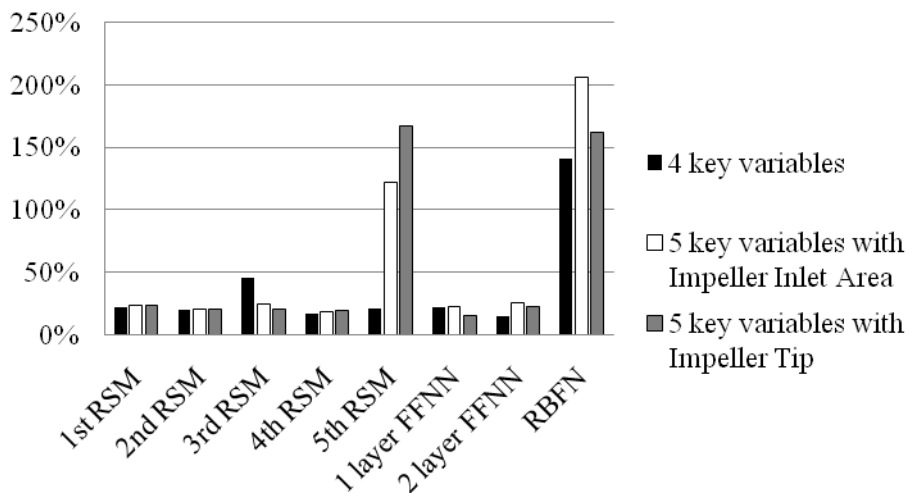
(a) Comparison of RMSEs in training databases

### RMSE of test database



(b) Comparison of RMSEs in test databases

### Maximum error



(c) Comparison of maximum errors

**Fig. 4** Comparison among models using 4 and 5 key variables

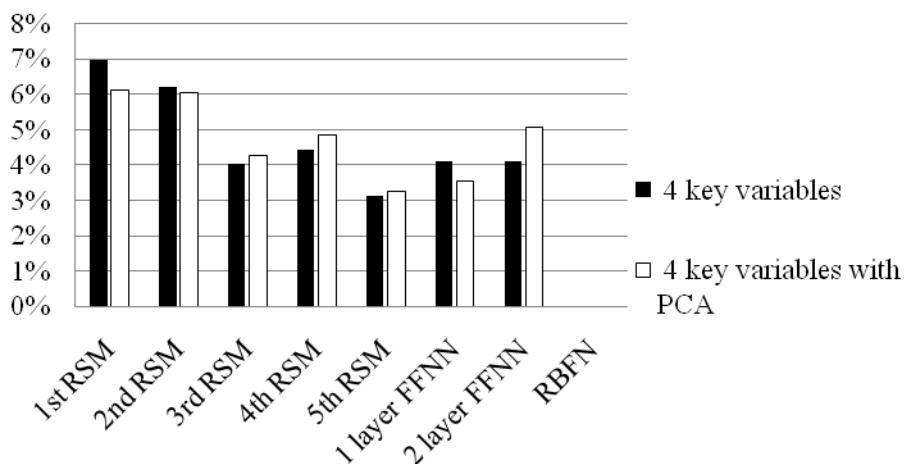
The results in Fig. 4a show that RMSEs of training database using the 3rd, 4th RSM decrease and that in using the 5th RSM significantly decreases by adding another key variable. However, the RMSE in the test databases in Fig. 4b and maximum error Fig. 4c using the 5th RSM increase dramatically while those of 4th RSM almost remain the same. The RMSE in the test databases

and maximum error using the 3rd RSM slightly decrease. However, the accuracy of models created using the 3rd RSM is still lower than those of 4th, 5th RSM and one-hidden-layer or two-hidden-layer FFNNs. Therefore, the new accuracy of 3rd RSM with 5 components key variable has no improvement on the optimal one but bring the higher computational cost of modeling. Therefore, it can be concluded that the introduction of additional key variables besides 4 key variables are sufficient to model the hydraulic turbine operating condition map.

### 5.3 Influences of Principle Component Analysis (PCA)

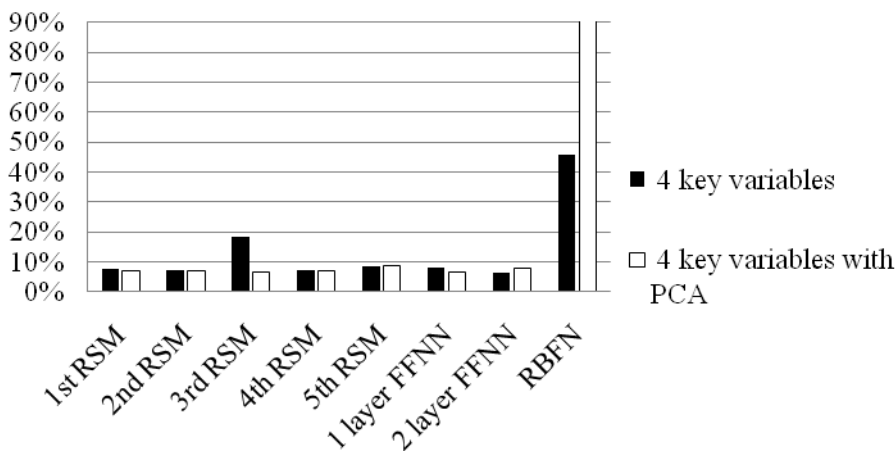
The results between using and not using PCA are compared and shown in Fig. 5. All results in Fig. 5a-c indicate that there is almost no change between using and not using PCA for 1<sup>st</sup>, 2<sup>nd</sup> RSM. Errors of 4<sup>th</sup> RSM, 5<sup>th</sup> RSM and two hidden layer slightly increase while those of RBFN significantly increase by adding PCA. However, one hidden layer FFNN still has slightly better performance compared to 3<sup>rd</sup> RSM. In summary, on average the best model approach for this specific problem is two-hidden-layer FFNN, and the second best one is the one-hidden-layer FFNN with using PCA. There are 4 and 1 neurons for the input and output layer. There are 8 neurons for the one-hidden-layer FFNN while 10 and 8 neurons for the two-hidden-layer FFNN. Linear transfer functions are used for the input and output layer while sigmoid transfer functions are used for hidden layer. The training epoch is 10,000.

**RMSE of training database**



(a) Comparison of RMSEs in training databases

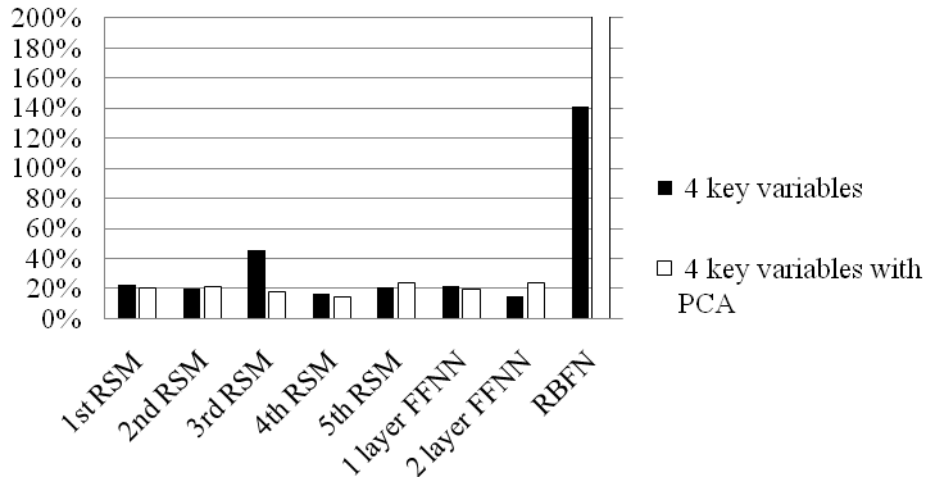
**RMSE of test database**



(b) Comparison of RMSEs in test databases



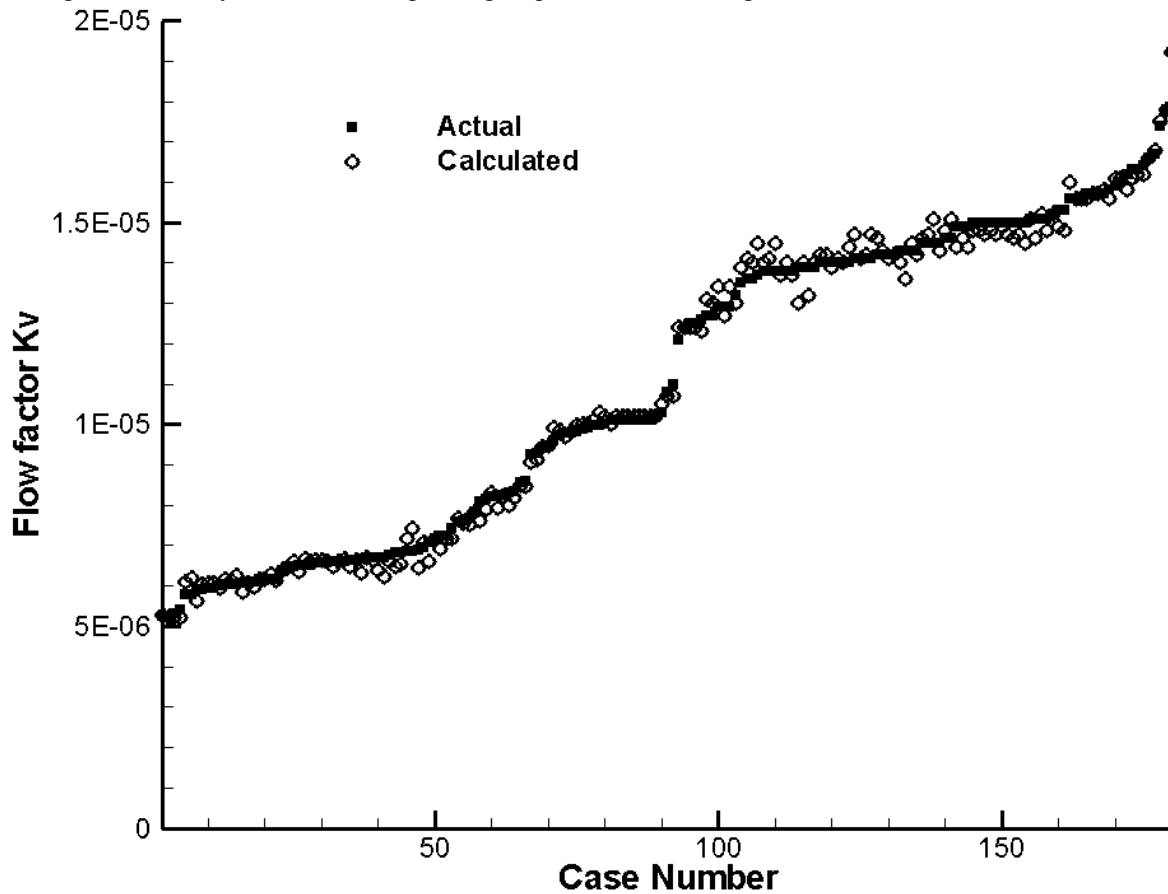
### Maximum error



(c) Comparison of maximum errors

**Fig. 5** Comparison between using and without using PCA

The comparison between the actual and predicted flow factor using two-hidden-layer FFNN, which has been proved as the best modeling method for hydraulic turbine operating map, is as shown in Fig. 6.



**Fig. 6** Comparison between the actual and predicted flow factor

In order to illustrate the comparison clearly in Fig. 6, cases are plotted in the order from the smallest to the largest actual value of flow factor. It can be seen that actual and predicted values match very well for most of the cases. Therefore, it can be concluded that the created flow factor map is able to calculate sufficiently accurate and easily used in one-dimensional design. The product cycle can be greatly decreased by using this flow factor map. However, there are still errors between the actual and predicted ones. The accuracy of the model created using FFNN for the medium part, where the flow factor from  $1.22E-5$  to  $1.48E-5$  in Fig. 6, is lower than other parts. There are many reasons to create the errors. First, the basic principles of FFNN, in which an approximation model is used to represent the real relationship, definitely determine their limits and bring the method error. Secondly, the turbulence flow occurs in the hydraulic turbine. Pressure and flow rate are not steady for short period of time, which introduces

random errors. Finally, human and machine error is also an important contributing factor to the errors. The error in the test record may be caused by the sensor precision. The noises in the key variables may be caused by the machine or measurement errors.

## 6. Conclusion

In this paper, the importance of the accurate prediction of hydraulic turbine flow factor has been explained. The potential key variables correlated to the ratio have been proposed. The correlation coefficients between potential key variables and the flow factor were calculated, based on which four key variables were finally determined to create the operating condition map.

Eight modeling methods categorized into Response Surface Method (RSM) and Artificial Neural Network (ANN) were used to create the flow factor map. The accuracies of maps were compared and comparison results showed that both 4<sup>th</sup>, 5<sup>th</sup> RSM and Feed-forward Neural Network (FFNN) had the good performance while two-layer FFNN had the best accuracies. The rotor inlet area or rotor tip was also considered as the key variable and was used to increase the accuracies of models. However, the comparison results between models with 4 and 5 key variables suggested that the model with four key variables has the higher accuracies and the introduction of extra key variable seemed unnecessary.

Principle Component Analysis (PCA) was employed to combine with model approaches. The advantages of PCA on modeling approaches are shown and discussed. All the results demonstrated that two-hidden-layer FFNN and one-hidden-layer FFNN with PCA has the best performance for this specific mapping. The actual and predicted flow factor, which were calculated based on trained two-hidden-layer FFNN, were shown and evaluated. The comparison showed good matching between actual and predicted ones and the causes of errors were also analyzed.

## Nomenclature

$b$	Bias in ANN	$Q$	Flow rate [m <sup>3</sup> /s]
$C$	Correlation coefficient	$R$	Function of RSM
$C_j$	Center vector in RBF	$X$	Normalized geometry parameter vector
$C_v$	Flow coefficient	$Y$	Normalized flow coefficient vector
$E$	Mean operator	$\Delta P$	Pressure differential
$f$	Activation function in ANN	$\beta$	Weights factor of ANN or coefficient of polynomial equation
$H$	Head	$\gamma$	Specific gravity
$K_v$	Flow factor	$\varepsilon$	Deviation between test and calculating valves
$N$	Rotating speed		
$P$	Pressure		

## References

- [1] Hanft, S., 2005, "Major Reverse Osmosis System Components for Water Treatment: The Global Market," MST049B.
- [2] Ghorbanian, K. and Gholamrezaei, M., 2007, "Axial compressor performance map prediction using artificial neural network," ASME Turbo Expo, Montreal, Canada, GT-2007-27165.
- [3] Yu, Y., Chen, L., Sun, F. and Wu, C., 2007, "Neural-Network based Analysis and Prediction of a Compressor's characteristic Performance Map," Applied Energy, Vol. 84, No.1, pp. 48-55.
- [4] Yi, W., Huang, H. and Han, W., 2006, "Neural Design Optimization of Transonic Compressor Rotor using CFD and Genetic Algorithm," ASME Turbo Expo, Barcelona, Spain, GT-2006-90155.
- [5] Alison-Youel, S.D., 2008, "Observation and Analysis of Affinity Law Deviations through Tested Performance of Liquefied Gas Reaction Turbines," ISROMAC-12, Honolulu, Hawaii, Article ID 737285.