

RESEARCH ARTICLE

Application of the hybrid ANFIS models for long term wind power density prediction with extrapolation capability

Monowar Hossain^{1*}, Saad Mekhilef^{1*}, Firdaus Afifi², Laith M. Halabi¹, Lanre Olatomiwa³, Mehdi Seyedmahmoudian⁴, Ben Horan⁵, Alex Stojcevski⁴

1 Power Electronics and Renewable Energy Research Laboratory (PEARL), Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia, **2** Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia, **3** Department of Electrical and Electronic Engineering, Federal University of Technology, PMB 65, Minna, Nigeria, **4** School of Software and Electrical Engineering, Swinburne University of Technology, Victoria, Australia, **5** School of Engineering, Deakin University, Victoria, Australia

* saad@um.edu.my (SM); apumonowar@gmail.com (MH)



OPEN ACCESS

Citation: Hossain M, Mekhilef S, Afifi F, Halabi LM, Olatomiwa L, Seyedmahmoudian M, et al. (2018) Application of the hybrid ANFIS models for long term wind power density prediction with extrapolation capability. PLoS ONE 13(4): e0193772. <https://doi.org/10.1371/journal.pone.0193772>

Editor: Xiaosong Hu, Chongqing University, CHINA

Received: November 23, 2017

Accepted: February 16, 2018

Published: April 27, 2018

Copyright: © 2018 Hossain et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors would like to acknowledge the financial support received from the University of Malaya, Malaysia, through Frontier Research Grant No. FG007-17AFR and Innovative Technology Grant No. RP043B-17AET.

Competing interests: The authors have declared that no competing interests exist.

Abstract

In this paper, the suitability and performance of ANFIS (adaptive neuro-fuzzy inference system), ANFIS-PSO (particle swarm optimization), ANFIS-GA (genetic algorithm) and ANFIS-DE (differential evolution) has been investigated for the prediction of monthly and weekly wind power density (WPD) of four different locations named Mersing, Kuala Terengganu, Pulau Langkawi and Bayan Lepas all in Malaysia. For this aim, standalone ANFIS, ANFIS-PSO, ANFIS-GA and ANFIS-DE prediction algorithm are developed in MATLAB platform. The performance of the proposed hybrid ANFIS models is determined by computing different statistical parameters such as mean absolute bias error (MABE), mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination (R^2). The results obtained from ANFIS-PSO and ANFIS-GA enjoy higher performance and accuracy than other models, and they can be suggested for practical application to predict monthly and weekly mean wind power density. Besides, the capability of the proposed hybrid ANFIS models is examined to predict the wind data for the locations where measured wind data are not available, and the results are compared with the measured wind data from nearby stations.

Introduction

The primary energy sources (fossil fuels) will soon be exhausted since they are used at a much higher rate than they are found in the earth's crust. Moreover, the price of fossil fuels is highly unstable, and it causes huge greenhouse gases (GHG) emissions and environmental pollutions [1, 2]. On the other hand, the wind energy is free, environmentally friendly and clean renewable energy. Consequently, in the fight of global climate change, wind energy is a major solution [3–5]. Globally, installed wind power capacity has reached 432.9GW at the end of 2015 where 63GW was added in 2015 alone [6].

Abbreviations: ANFIS, Adaptive neuro-fuzzy inference system; ANN, Artificial neural network; DE, Differential evolution; ELM, Extreme learning machine; GA, Genetic algorithm; GP, Genetic programming; PSO, Particle swarm optimization; SVM, Support vector machine; GMCM, Gaussian mixture copula model; GPR, Gaussian process regression; WT, Wavelet transform; GS, Grid search; GHG, Greenhouse gases; MABE, Mean absolute bias error; MAPE, Mean absolute percentage error; RMSE, Root mean square error; R^2 , Coefficient of determination; MMD, Malaysian Meteorological Department; WPD, Wind power density.

However, wind energy is unstable and subject to intermittent characteristics thus, the accurate prediction of the wind speed and the wind power is a vital part of the successful establishment of the wind energy conversion system [7]. Again, to build a wind farm in any particular location, analysis of wind data, estimation of wind power and energy density are essential [8, 9]. The wind power density (WPD) of a particular location is the measure of the potentiality of wind resources and the chance of extracting wind energy at different wind speed from that location. The knowledge of WPD also helps the designer and investor to understand the performance of wind turbine and to choose the optimal number of a wind turbine with a suitable power rating [10, 11].

The wind power can be computed from several numerical methods [12, 13]. The problem in numerical methods is that they need high computation time. In the recent years, artificial intelligence (AI) techniques have received overwhelming popularity in the field of the wind energy system and other engineering applications as they offer better advantages, including fast computation time, require no knowledge of internal system parameters and compact solutions [14–20]. Generally, wind speed and power prediction are divided into three categories, namely, short-term (30 min to 6 h), medium-term (6h to 24h), and long-term (24h and longer) predictions [7, 21].

Short-term wind prediction

In ref. [22], two different short-term wind power prediction methods namely; individual ANN and hybrid strategy based on the physical and statistical methods were developed, where individual ANN and hybrid strategy resulted in 10.67% and 2.01% root mean square error (RMSE) respectively in the prediction. However, the hybrid strategy was 5 times slower than individual ANN. An ANFIS based hybrid model was developed in [23] to predict short-term wind power in Portugal that resulted in MAPE of 5.41%, outperforming five other approaches. The authors used historical wind power data as inputs. In [24], the authors applied both ANN and ANFIS models for hourly wind power prediction for a wind farm located in Southern Italy and their prediction accuracy resulted worse when the prediction horizon was increased. More literature review regarding the application of AI methodologies for the prediction of short-term wind speed and power can be found in the ref. [16, 21].

Medium-term wind prediction

In [25], an ANN model was employed for the prediction of daily mean wind speed of 11 locations in India where actually measured wind data are not available. The authors used meteorological variables of the target locations from NASA surface meteorology and solar energy database, and the prediction accuracy is compared with measured wind data that was collected from a nearby meteorological station in Hamirpur. A hybrid method consists of wavelet transform (WT), ANFIS, SVM, and GS was proposed in [26] for 6h ahead wind power forecasting. This study showed that the proposed method can predict wind power with MAPE of 12.16% to 13.83%. More literature review regarding the application of soft computing methodologies for the prediction of medium-term wind speed and power can be found in the ref. [21].

Long-term wind prediction

An ELM model was developed in ref. [27] to predict the monthly wind power density of a particular location in Iran. The authors compared the WPD obtained from ELM model with that from ANN, GP, and SVM. The results showed the performance of ELM higher than other models. The suitability of ANFIS to estimate monthly WPD for the location of Aligoodarz, Iran has been shown in the article [28]. For the prediction of long-term wind speed and

power, an ANN and statistical based models were conducted in ref. [7] where the proposed method showed rather promising results in view of the very small mean absolute error (MAE). In study [29], a hybrid model denoted by WT+FA+FF+SVM was reported and the computed MAPE were in the range of 13.46–18.74% for weekly prediction. More literature review regarding the application of soft computing methodologies for the prediction of long-term wind speed and power can be found in the ref. [21, 30].

Long-term prediction of wind speed has become a research hotspot in many different areas such as restructured electricity markets, energy management, and wind farm optimal design. Although ANFIS merges the learning power of the ANNs with the knowledge representation of fuzzy logic, there are still some difficulties in ANFIS in constructing membership functions (MFs). The difficulty of using ANFIS in constructing membership functions lies in tuning the function to build the best model with high accuracy and better performance. Therefore, this study proposed hybrid ANFIS; ANFIS-PSO, ANFIS-GA, and ANFIS-DE to predict long-term (monthly and weekly) wind power density for four different places in Malaysia namely; Mersing, Kuala Terengganu, Pulau Langkawi and Bayan Lepas. The main benefit of combining these three techniques (PSO/GA/DE) with ANFIS is to reduce the error rates by tuning and optimizing the membership functions. Besides, this study examined the wind speed prediction capabilities of the proposed models for the locations where measured wind data are not available, and the result of the wind speed extrapolation is compared with the measured wind data collected from the nearby meteorological station.

Data collection and analysis

Site description

The aim of this study is to predict long-term (monthly and weekly) wind power of four different locations situated under four distinctive state of Malaysia shown in Table 1. For this purpose, wind speed data were collected from Malaysian Meteorological Department (MMD) situated in the respective locations during the period (2004–2014). As presented in Table 1, the wind data were measured at different heights above sea level by a rotating cup-type anemometer. Those are 43.6m, 5.2m, 6.4m and 2.46m for Mersing, Kuala Terengganu, Pulau Langkawi and Bayan Lepas respectively. It is important to mention that the wind data recorded in different heights need to adjust to the same height because of various characteristics of wind speed with altitude. The wind shear, which is the variation of wind velocity with altitudes, is most pronounced near the surface (sea and land). Due to the drag of surface and viscosity of air, the wind blows faster at higher altitudes.

Typically, the variation of wind speed at daytime follows the $1/7^{\text{th}}$ power law whereas, when the temperature become stable or better at night time, the wind speed close to the ground usually subsidies and at turbine altitudes, it does not decrease that much or may even increase. Thus, the daily average wind speed data collected from the meteorological stations were

Table 1. Location and description of the sites.

Locations	State	Geographical coordinate of meteorological stations		Measured height of wind speed (m)
		Latitude	Longitude	
Mersing	Johor	2° 27' N	103° 50' E	43.6
Kuala Terengganu	Terengganu	5° 23' N	103° 60' E	5.2
Pulau Langkawi	Kedah	6° 20' N	99° 44' E	6.4
Bayan Lepas	Pulau Penang	5° 17' N	100° 16' E	2.46

<https://doi.org/10.1371/journal.pone.0193772.t001>

adjusted at turbine hub height of 50m using the power law. The power law for wind speed adjustment at the different hub height is defined as [12]:

$$\frac{v}{v_o} = \left(\frac{h}{h_o}\right)^\alpha \tag{1}$$

where v is the wind speed at is desired height h and v_o is the wind speed at measured height h_o . While α is the power law coefficient. The exponent (α) is an empirically derived coefficient that varies depending upon the stability of the atmosphere. For neutral stable conditions, it is approximately 1/7, or 0.143, which is commonly assumed to be constant in wind resource assessments. This is because the differences between the two levels are not usually so great as to introduce substantial errors into the estimation (usually < 50m). The value of the coefficient varies from less than 0.10 for very flat land, water or ice to more than 0.25 for heavily forested landscapes and the typical value of 0.14 for low roughness surfaces. The value 0.143 for the coefficient has been chosen for this assessment [12, 31, 32].

Wind power density (WPD)

The wind power density (WPD) is an essential indicator to estimate wind potentiality in a specific location. The computation methods of WPD include; application of measured wind data and the use of Weibull distribution function. The power in the wind at a measured wind speed v passing through a blade sweep area can be expressed as [13, 33]:

$$\bar{P} = \frac{1}{2n} * \rho * \sum_{i=1}^n v^3 = \frac{1}{2} * \rho * \bar{v^3} \left(\frac{W}{m^2}\right) \tag{2}$$

where n is the number of data points over a time period, $\bar{v^3}$ is the mean of cube of wind speed and ρ is the air density (kg/m^3), taken $1.175 kg/m^3$ in this study.

As presented in [12, 31], the 2-parameters Weibull distribution function is the most appropriate, recommended and acceptable model for wind potentiality analysis. In Weibull distribution, the probability distribution function (PDF) determines the probability of the wind at a given velocity V and it can be expressed as [12, 34]:

$$f(V) = \left(\frac{k}{c}\right) \left(\frac{V}{c}\right)^{k-1} \exp\left[-\left(\frac{V}{c}\right)^k\right] \tag{3}$$

On the other hand, cumulative distribution function (CDF) of wind speed V indicates the probability that the wind velocity is equal to or lower than V or within a given wind speed range. CDF can be expressed as [12, 34]:

$$F(V) = 1 - \exp\left[-\left(\frac{V}{c}\right)^k\right] \tag{4}$$

where V is wind speed (m/s), k (dimensionless) is shape factor and c (m/s) is scale factor.

As earlier mentioned, the k and c (m/s) can be computed by several empirical methods. Commonly used is standard deviation method, k and c are defined as follows [12, 13]:

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \tag{5}$$

$$c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)} \tag{6}$$

where \bar{v} is the average wind speed (m/s), σ is standard deviation and $\Gamma(x)$ is the gamma function which is defined as [35, 36]:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \tag{7}$$

Another empirical method is power density or energy pattern factor method. In this method, E_{pf} is needed to be estimated to compute the shape factor k and scale factor c (m/s).

The E_{pf} is known as wind pattern factor which is used for wind turbine aerodynamic design and it is defined as follows [33, 37]:

$$E_{pf} = \frac{\frac{1}{n} \sum_{i=1}^n v_i^3}{\left(\frac{1}{n} \sum_{i=1}^n v_i\right)^3} = \frac{\bar{v}^3}{\bar{v}^3} = \frac{\Gamma\left(1 + \frac{3}{k}\right)}{\Gamma^3\left(1 + \frac{1}{k}\right)} \tag{8}$$

In simple word, E_{pf} is the ratio of mean of cube wind speed to cube of mean wind speed. When E_{pf} is known then shape factor k can be easily estimated by following formula [33, 37]:

$$k = 1 + \frac{3.69}{(E_{pf})^2} \tag{9}$$

The wind power density on the basis of Weibull probability density function is estimated using the following equation [13]:

$$\bar{P} = \frac{1}{2} \rho \int_0^{\infty} v^3 f_w(v) dv = \frac{1}{2} \rho c^3 \Gamma\left(1 + \frac{3}{k}\right) \text{ (W/m}^2\text{)} \tag{10}$$

Methodology

As mentioned in section 2.1, to accomplish the study objective, 11 years (2004–2014) long-term wind speed data measured by the rotating cup-type anemometer were collected from the four different locations in Malaysia; Mersing, Kuala Terengganu, Pulau Langkawi and Bayan Lepas. The raw data was thereafter adjusted to turbine hub height of 50m. The raw data used in this study is presented in [S1 Dataset: Supporting Information.rar](#). Afterward, the monthly and weekly mean wind speed at 50m and the corresponding wind power density from measured data were applied on the developed standalone ANFIS and hybrid ANFIS models. In this study, the data size used for the training and testing the prediction models are defined as P and Q respectively. The purpose of the training process in ANFIS model is to minimize the error between the actual target and the ANFIS output. Based on the literature on ANN, the percentage of training data must be higher than testing data for effective learning of the system before the system can produce a good result. The developed models were trained and validated with several data segments such as P = 80%, Q = 20%; P = 70%, Q = 30% and P = 60%, Q = 40% for training and testing respectively. The percentage of data selected for training and testing has been carefully tested based on the minimal error obtained in the statistical indicator. It is important to mention that no specific rules were considered to choose the data size for training and testing the models. Application of different training and testing data size on the prediction models helps to observe the error metrics and to choose best data size providing the minimum error in the prediction.

ANFIS (adaptive neuro-fuzzy inference system)

The term adaptive neuro-fuzzy inference system (ANFIS) was first introduced by Jang in 1993. ANFIS is a hybrid intelligent scheme that merges the learning power of ANNs with the knowledge representation of fuzzy logic to produce a powerful processing tool [38]. The fuzzy inference system (FIS) is the core of ANFIS. FIS is based on expertise expressed in terms of

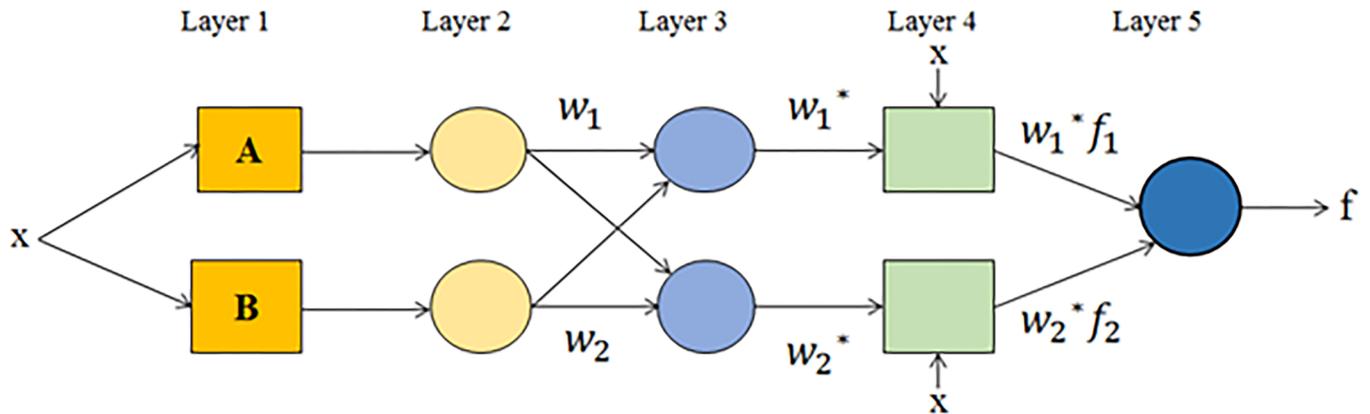


Fig 1. ANFIS structure.

<https://doi.org/10.1371/journal.pone.0193772.g001>

‘IF–THEN’ rules, thus it can be used to predict the behavior of many uncertain systems. One of the advantages of FIS is that it does not require knowledge of the main physical process as a pre-condition for operation. Thus, ANFIS integrates FIS with a back propagation learning algorithm of a neural network. These techniques provide a method for the fuzzy modeling procedure to learn from the available data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data as shown in Fig 1. For one input, two fuzzy ‘IF–THEN’ rule are generated for the maximum equal to 1 and minimum equal to 0. The fuzzy inference system employed in this study uses one input x and one output f . A first-order Sugeno fuzzy model with two fuzzy *if-then* rules is used as follows [39]:

$$\text{Rule 1 : if } x \text{ is } A \text{ then } f_1 = p_1x + t_1 \tag{11}$$

$$\text{Rule 2 : if } x \text{ is } B \text{ then } f_2 = p_2x + t_2 \tag{12}$$

Layer 1 contains membership functions (MFs) of input variables and provides the input values for the next layer. In the 1st layer, each node is adaptive as $O = \mu_{AB}(x)$, where $\mu_{AB}(x)$ are MFs. The bell-shaped MFs is presented in Eq 13, for which the lowest and highest amounts are 0 and 1, respectively.

$$f(x; a, b, c) = \frac{1}{1 + \left(\frac{x-c}{a}\right)^{2b}} \tag{13}$$

The function is subject to the following parameters, namely a , b and c . Each of these parameters defines as follows: a is half width of the curve; b define the gradient together with a , and c is the midpoint of the membership function as shown in Fig 2.

In the 2nd layer (the membership layer), the weight of membership functions (MFs) is considered. Input values for this layer are supplied from the first layer. The nodes in the second layer are a fixed node. The output is the product of all incoming signals and be described as,

$$w_i = \mu(x)_i \cdot \mu(x)_{i+1}, \text{ for } i = 1, 2 \tag{14}$$

The output of each node indicates the weight strength of a rule.

In layer 3 or the rule layer, every node does the pre-condition matching of the fuzzy rules, that calculate each rule’s activation level as well as the normalized firing strength. This is a

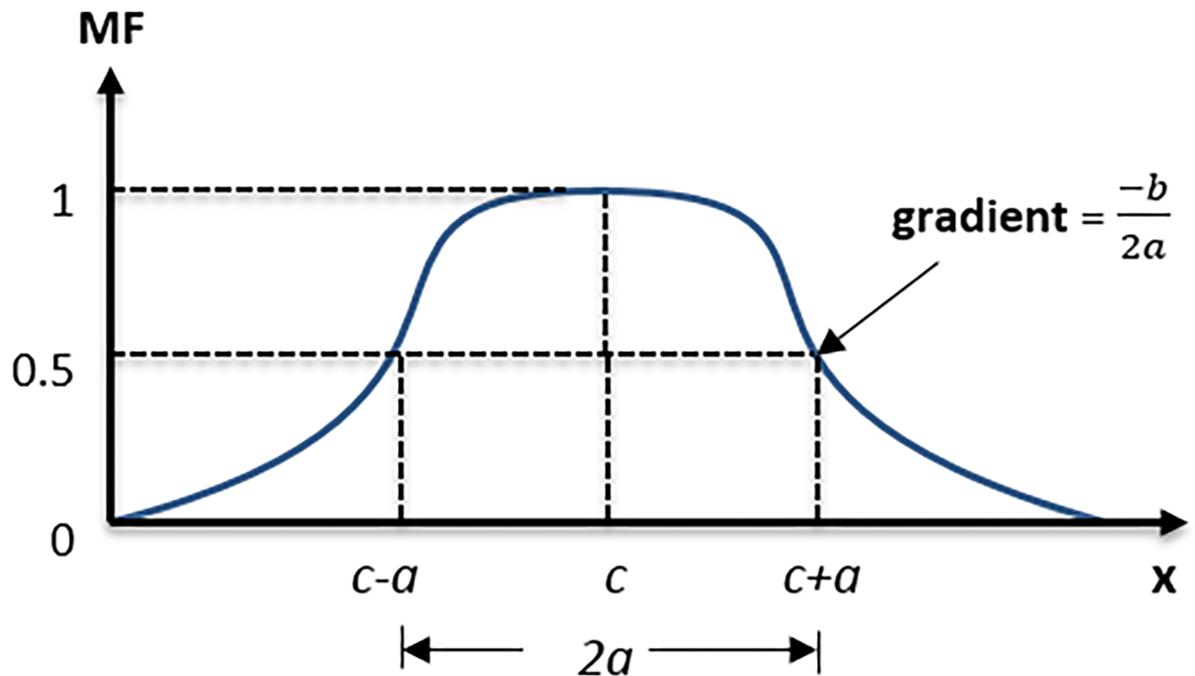


Fig 2. Three parameters bell-shaped membership function; a , b and c .

<https://doi.org/10.1371/journal.pone.0193772.g002>

fixed layer as well, and every node computes the ratio of the i th rule of the firing strength to the sum of i th firing strengths of all rules as:

$$w_i^* = \frac{w_i}{w_1 + w_2}, \text{ for } i = 1, 2 \quad (15)$$

The outputs of this layer are named as normalized weights or firing strengths.

In layer 4 or defuzzification layer, all the adaptive nodes provide the resulting output values from the inference of rules.

$$O_i^4 = w_i^* \cdot f = w_i^* (p_i x + t_i) \quad (16)$$

Here, the parameters set are shown as $\{p_i, t_i\}$.

Layer 5 or the output layer summarizes the inputs and output from layer 4. This layer also converts the results of fuzzy classification into a crisp. Here, the single node is fixed node, and the whole incoming signal is sum up to produce overall output as below,

$$O_i^5 = \sum_i w_i^* \cdot f = \frac{\sum_i w_i \cdot f}{\sum_i w_i} \quad (17)$$

Three optimization techniques namely; PSO, GA, and DE were employed to adjust the ANFIS membership function parameters. The main benefit of combining these three techniques with ANFIS is to reduce the error rates by tuning and optimizing the membership functions.

ANFIS-PSO

Particle swarm optimization (PSO) is an approach for optimizing “continuous” and “discontinuous” decision-making functions as developed by Kennedy and Eberhart in 1995 [40]. PSO

has been used to model animals' sociological and biological behavior such as groups of birds searching for food [41]. PSO has also been employed in population-based search approach, in which a particle of a population is present for each individual potential solution or swarm. In this method, the position of each particle is changed constantly in a search space until getting to the optimum solutions and computational limitations are reached [42]. In PSO, swarm starts with a group of random solutions, each of which is called a particle, and \vec{s}_i represents the particle's position. Likewise, a particle swarm moves in the problem space, where \vec{v}_i expresses the particle's velocity. A function f is evaluated at each time step through input \vec{s}_i . Every particle records its best position related to the best fitness gained to this point, in \vec{p}_i vector. \vec{p}_i^g tracks the most appropriate position identified by any neighborhood member. In universal form of PSO, \vec{p}_i^g represents the best appropriate position in the entire population. A new velocity is achieved for any particle i in each iteration according to the best positions of individual, $\vec{p}_i(t)$ and $\vec{p}_i^g(t)$ neighborhood. The new velocity can be presented by:

$$\vec{v}_i(t + 1) = w\vec{v}_i(t) + c_1 \vec{\theta}_1 \cdot (\vec{p}_i(t) - \vec{x}_i(t)) + c_2 \vec{\theta}_2 \cdot (\vec{p}_i^g(t) - \vec{x}_i(t)) \tag{18}$$

where w represents the inertia weight and positive acceleration coefficients are represented by c_1 and c_2 . $\vec{\theta}_1$ and $\vec{\theta}_2$ represent uniformly-distributed random vectors [0,1], in which a random value is tried for every dimension. \vec{v}_i limit in the $[-\vec{v}_{max}, \vec{v}_{max}]$ range is relent on the problem provided the velocity exceeds the mentioned limit. In some cases, it is rearranged within its suitable limits. The position of every particle alters depending upon the velocities as:

$$\vec{s}_i(t + 1) = \vec{s}_i(t) + \vec{v}_i(t + 1) \tag{19}$$

According to Eqs (18) and (19), the particles incline to gather around the best. Fig 3 depicts the sequential PSO and ANFIS combination [43].

The PSO use for designing a fuzzy system (FS) or parameter optimization is expressed as [44]:

$$R_i : \text{ if } x_1(k) \text{ is } A_{i1} \text{ And } \dots \text{ And } x_n(k) \text{ is } A_{in} , \text{ Then } u(k) \text{ is } a_i \tag{20}$$

where, a_i is a crisp value, k represents the time step, the input variables are $x_1(k), \dots, x_n(k)$, A_{ij} is a fuzzy set and $u(k)$ signifies the system output variable.

For the FS in Eq (20) which comprises r rules and n input variables, it's free parameters are defined through a position vector:

$$\vec{s} = [m_{11}, b_{11}, \dots, m_{1n}, b_{1n}, a_1, \dots, m_{r1}, b_{r1}, \dots, m_{rn}, b_{rn}, a_r] \in \mathfrak{R}^D \tag{21}$$

$$m_{rj} = x_j(k), b_{rj} = b_{fix}, j = 1, \dots, n \tag{22}$$

Following the process of rule creation and initialization, the preliminary antecedent part parameters are outlined. According to Eq (21) and Eq (22), the i th solution vector \vec{s}_i is created as:

$$\begin{aligned} \vec{s}_i &= [s_{i1} \ s_{i2} \ \dots \ s_{iD}] \\ &= [m_{11} + \Delta m_{11}^i, b_{fix} + \Delta b_{11}^i, \dots, m_{1n} + \Delta m_{1n}^i, b_{fix} + \Delta b_{1n}^i, a_1, \dots, m_{r1} + \Delta m_{r1}^i, b_{fix} \\ &\quad + \Delta b_{r1}^i, \dots, m_{rn} + \Delta m_{rn}^i, b_{fix} + \Delta b_{rn}^i, a_r] \end{aligned} \tag{23}$$

In Eq (23), Δm_{ij} and Δb_{ij} represent small random numbers, a_i designates a random number distributed arbitrarily and homogeneously in the FS output range. The f (evaluation function) for \vec{s}_i is calculated based upon the FS performance.

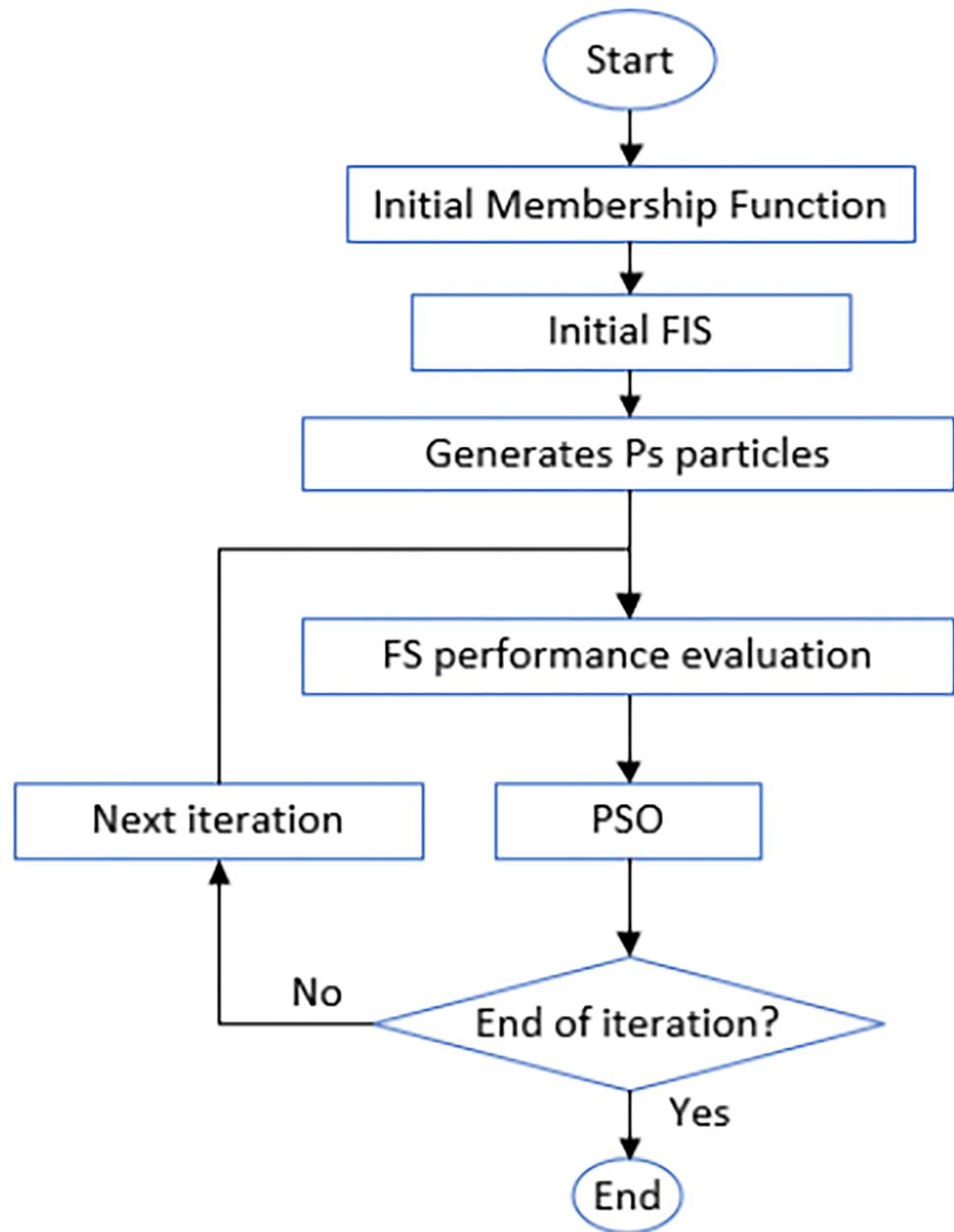


Fig 3. The sequential combination of PSO and ANFIS.

<https://doi.org/10.1371/journal.pone.0193772.g003>

PSO looks for the best originator part parameters. P_s represents the population size. Eq (21) sets the elements in position \vec{s} . When $t = 0$, the $\vec{s}_1(0), \dots, \vec{s}_p(0)$ or initial positions are created arbitrarily according to the best-performing FS found in PSO (\vec{s}_{PSO}). $\vec{s}_1(0)$ is considered similar to \vec{s}_{PSO} . The left $P_s - 1$ particles, $\vec{s}_2(0), \dots, \vec{s}_{P_s}(0)$, are created by addition of uniformly-distributed random numbers to \vec{s}_{PSO} shown as:

$$\vec{s}_i(0) = \vec{s}_{PSO} + \vec{w}_i, \quad i = 2, \dots, P_s \quad (24)$$

where, \vec{w}_i represents a random vector. The primary speed values of all particles, $\vec{v}_i(0)$, $i = 1, \dots, P_s$, are generated randomly. Each particle's performance is evaluated according to

Table 2. Parameter characteristics for ANFIS-PSO.

Population Size	Iterations	Inertia Weight	Damping Ratio	Learning coefficient	
				Personal	Global
100	500	1	0.99	2	2

<https://doi.org/10.1371/journal.pone.0193772.t002>

the FS it signifies. f is described as the $E(t)$ or error index mentioned above. The best position (p_i) of each particle and the best particle p_g in the whole population is obtained according to f . Eqs (18) and (19) overhaul the velocity and position of each particle. The whole learning procedure is accomplished as soon as a pre-defined paradigm is obtained [44].

There are five PSO main parameters used during conducting the experiment as shown in Table 2, this includes; a maximum number of iterations, the population size of the domain, inertia weight damping ratio and inertia weight, global and personal learning coefficient. For this case studies, the optimum values of these parameters are determined by trial and error procedure.

ANFIS-GA

Genetic Algorithm is global search heuristics technique used to find solutions for optimization and solve highly complex search problems. It is a particular class of evolutionary computing method inspired by the idea of natural selection evolutionary process which implemented inheritance, mutation, selection, and recombination [45]. In this hybrid approach, GA is combined with ANFIS to extend its prediction proficiency. GA is implemented to improve ANFIS performance and minimize the error rates by tuning and optimizing the membership functions of a Sugeno type fuzzy inference system. The ANFIS-GA forecast allows reforming of the upcoming behavior of the wind power density and therefore, determines the viability of the wind power plants from any location.

The Hybrid ANFIS-GA model used is shown in Fig 4 [46].

GA model begins with a set of solutions (referred to as chromosomes) represented as population. A new population is drawn from the completion of a previous population. New solutions that formed from selected solution (offspring) are designated according to their fitness.

This process is repeated until some condition is occurred (for example, number of populations or improvement of the optimal solution) is fulfilled. To achieve this, ANFIS algorithm plays an important role as part of the fitness function, $f(x)$. The fitness with intervention of ANFIS fitness function is represented by;

$$f_1(x) = \frac{1}{m} \sqrt{\sum_{i=0}^m (d_i - a_i)^2} \tag{25}$$

where m is a number of feature attributes, a_i is output derived through ANFIS, d_i is desired wind power density.

The next fitness function can be presented as:

$$f_2(x) = \frac{1}{n - m} \sqrt{\sum_{i=m}^n (d_i - a_i)^2} \tag{26}$$

where n is the total number of input features, d_i is set to minimum, a_i is actual value of wind power density and $n - m$ represents remaining undesired features. The final equation is minimized $f(x)$, describe as,

$$f(x) = \frac{f_1(x) + f_2(x)}{2} \tag{27}$$

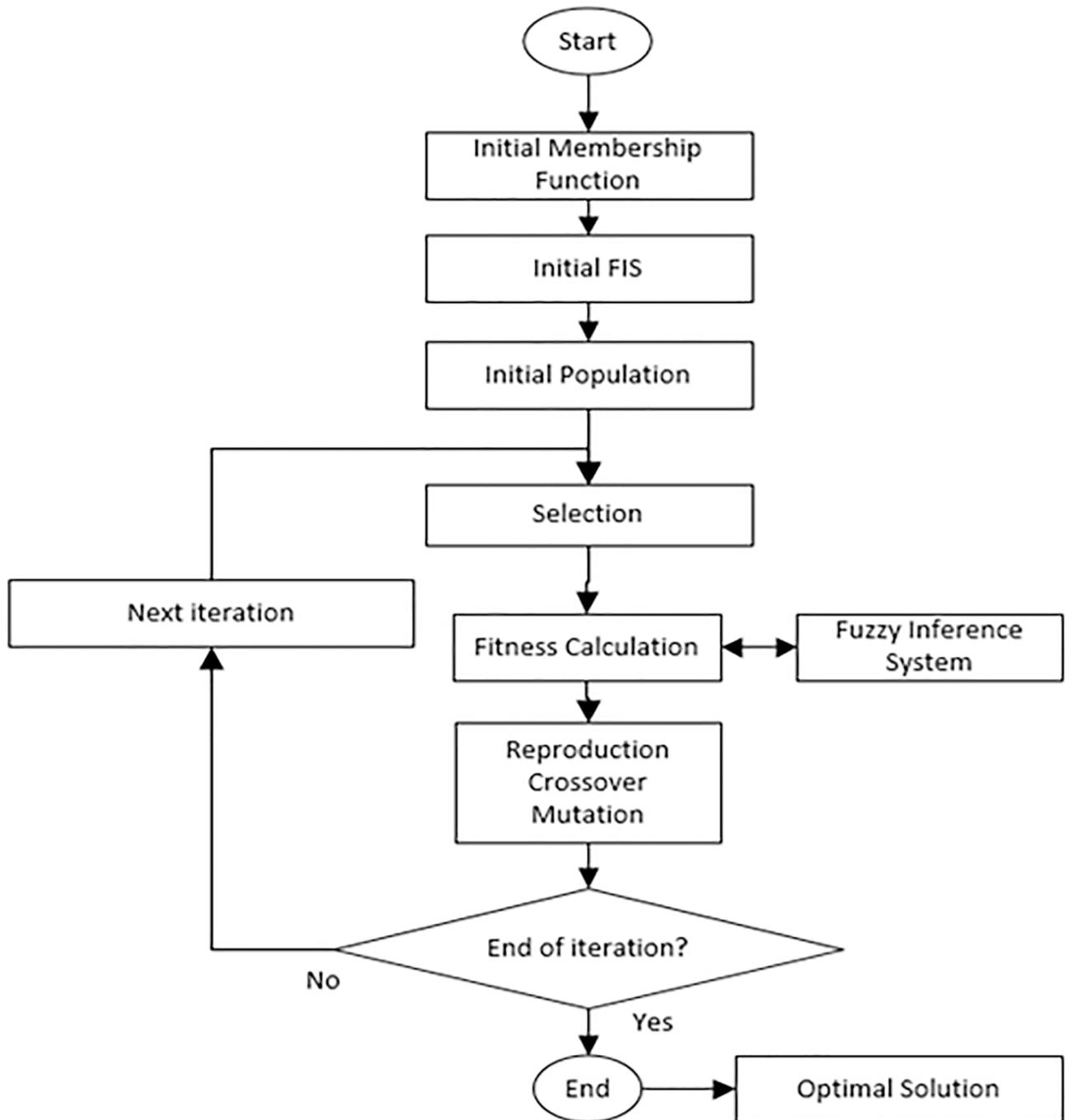


Fig 4. ANFIS-GA model.

<https://doi.org/10.1371/journal.pone.0193772.g004>

For this case study, we determined the specific parameters initialization for the GA. These include; number of iterations, population size, mutation and crossover percentage. Selection of these parameters decides, to a great extent, the ability of the designed controller. The range of the tuning parameters is listed in Table 3.

Table 3. Parameter characteristics of ANFIS-GA.

Population Size	Number of Iterations	Crossover Percentage	Mutation Percentage	Mutation Rate	Selection Pressure	Selection Function
100	500	0.8%	0.3%	0.02	8	Roulette Wheel

<https://doi.org/10.1371/journal.pone.0193772.t003>

After the fitness $f(x)$ of each chromosome x in the population is evaluated, new population is created and following steps are repeated until it is completed. The better fitness gives bigger chance to parent chromosomes from a population to be selected. Then it is crossover the parents to form a new offspring with the crossover probability. Next mutation probability mutates new offspring at each position in chromosome. As the solution goes under Reproduction, Crossover and Mutation with parameters setting from 3, the best solution in current population is returned if end condition is satisfied. The optimal solution calculated will help GA to search for optimized membership function.

ANFIS-DE

Differential evolution (DE) is first introduced as a heuristic method by Storn and Price [47] to solve problems involving global optimization as the solution to minimizing possibly nonlinear and non-differentiable continuous space functions. Since both DE and GA are part of evolutionary computing methodologies, DE functions almost in the same manner as GA. The different in DE is using actual real numbers in a strict mathematical sense, which can be applied to real-valued problems over a continuous space. As a result, the designs of crossover and mutation are significantly different. The idea behind the method of DE is that the difference between two vectors produces a difference vector which is used with a scaling factor to traverse the search space [48].

In ANFIS-DE hybrid approach, DE is combined with ANFIS to improve the performance of ANFIS prediction proficiency. Differential Evolution initialize with population size of (*pop size*) individuals solutions which can be represented as $x_i^t = 1, 2, \dots, \text{popsize}$ for each individual where i represents the population and t_{th} represents the generation to which the population belongs. Then the algorithm depends on the operation of three main operators; mutation, crossover and selection as shown in Fig 5 [46].

Mutation operator is the main operator of DE which differs from other EAs. We implemented DE/rand/1 mutation strategy as described in Eq 28.

$$u_i^t = x_{r_1}^t + F \cdot (x_{r_2}^t - x_{r_3}^t) \tag{28}$$

where u_i^t is the mutant individual for x_i^t and r_1, r_2, r_3 are randomly selected and satisfy individuals. Moreover, they are not equal to running index (i) and mutually different. F is control parameter in the range of $[0, 2]$

Process crossover is carried out after the mutation phase is completed and can be described as,

$$y_{id}^t = \begin{cases} u_{id}^t & \text{if } rand \leq C_R, \\ x_{id}^t & \text{otherwise,} \end{cases} \tag{29}$$

where y_i^t the trial individual, d is representing the d_{th} component of individuals. The *rand* and crossover rate C_R is a parameter in the range of $[0, 1]$.

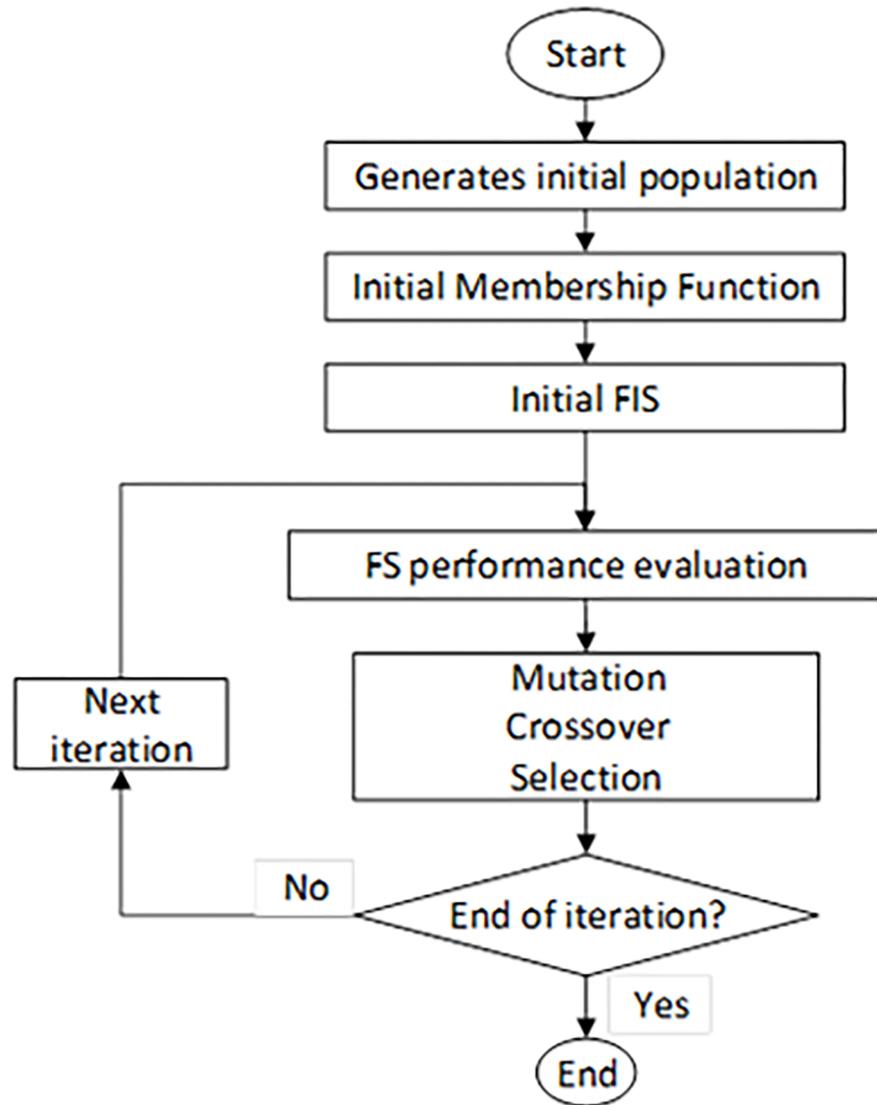


Fig 5. ANFIS-DE model.

<https://doi.org/10.1371/journal.pone.0193772.g005>

Selection operation in DE is also different from other evolutionary algorithms. Once all *pop size* trial individuals are generated, the selection operation is processed as:

$$x_{id}^{t+1} = \begin{cases} y_i^t & \text{if } fitness(y_i^t) \leq fitness(x_i^t), \\ x_i^t & \text{otherwise,} \end{cases} \quad (30)$$

where $fitness(x)$ is the fitness function. Table 4 shows the initial parameters of the ANFIS-DE model.

Table 4. Parameter characteristics for ANFIS-DE.

Population Size	No. of Iterations	Crossover Percentage	Lower bound of scaling factor	Upper bound of scaling factor
100	500	0.2%	0.2	0.8

<https://doi.org/10.1371/journal.pone.0193772.t004>

Statistical indicators model performance evaluation

The performance of the proposed system can be checked by computing several statistical parameters. The most popular statistical error indicators are the mean absolute bias error (MABE), mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination (R^2). The MABE is the average quantity of the summation of all absolute bias error between the predicted and the measured value represented as:

$$MABE = \frac{1}{N} \sum_{i=1}^N |(V_{i,P} - V_{i,M})| \tag{31}$$

The MAPE is the mean absolute percentage difference between the predicted and measured wind power density represented:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \left(\frac{V_{i,P} - V_{i,M}}{V_{i,M}} \right) \right| \times 100 \tag{32}$$

The RMSE presents the accuracy of the model by comparing the deviation between predicted and measured wind power density. The value of RMSE is always positive and it is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_{i,P} - V_{i,M})^2} \tag{33}$$

The coefficient of determination (R^2) indicates the strength of linear relationship between the predicted and measured wind power density. R^2 is obtained by:

$$R^2 = 1 - \frac{\sum_{i=1}^N (V_{i,P} - V_{i,M})^2}{\sum_{i=1}^x (V_{i,M} - V_{M,avg})^2} \tag{34}$$

In the Eqs (31–34), $V_{i,P}$ and $V_{i,M}$ are wind power density estimated from developed prediction models and measured data respectively.

Result and discussion

Monthly wind power density prediction

In the current study, hybrid ANFIS (ANFIS-PSO, ANFIS-GA, ANFIS-DE) and standalone ANFIS models were developed to predict wind power density. The above-mentioned models were trained and tested with different data size. Tables 5 and 6 shows descriptive values of input and output parameters respectively such as; maximum (Max.), minimum (Min.), standard deviation (St Dev.) and range for different locations. On the other hand, Fig 6 shows visual presentation of wind speed at 50m hub height for study locations. It can be found in Fig 6 that maximum wind speed prevails in Mersing and followed by Pulau Langkawi, Bayan Lepas, and Kuala Terengganu. The raw data used in this study is presented in S1 Dataset: Supporting Information.rar.

Table 7 presents the error metrics, including MAPE, MABE, RMSE obtained while training the prediction models with the input-output data set for Mersing, Langkawi, Bayan Lepas and Kuala Terengganu. The data size for training and testing the models were $P = 105$ and $Q = 27$ respectively. In this study, the performance of the prediction models is categorized based on lowest RMSE. It can be observed from Table 7 that ANFIS-PSO model has lowest RMSE, MAPE and MABE in the training stage for the data set of Mersing, Bayan Lepas, and Kuala Terengganu whereas, ANFIS-GA ranks in the first for the data set of Pulau Langkawi. Table 8

Table 5. Descriptive statistical parameters of input (wind speed, m/s) for different locations.

Parameters	Locations			
	Mersing	Kuala Terengganu	Pulau Langkawi	Bayan Lepas
Max.	6.05	4.50	5.35	4.80
Min.	2.15	1.84	1.72	2.12
St dev.	0.87	0.57	0.72	0.52
Range	3.91	2.66	3.63	2.68

<https://doi.org/10.1371/journal.pone.0193772.t005>

summarizes the RMSE, MAPE and MABE performance metrics when testing data set of different underlying locations is applied to the predictions models. As can be seen from Table 8, the ANFIS-GA model ranks in the first for the underlying data sets of Mersing and Kuala Terengganu. On the other hand, ANFIS-PSO model provides the smallest error metrics for the data set of Pulau Langkawi and Bayan Lepas.

It is important to mention that sometimes ANFIS model provides better performance than hybrid ANFIS models but does not provide the best model performance for any of the underlying locations. For instance, ANFIS shows second and third best performance for the data set of Bayan Lepas and Mersing respectively when testing the prediction models.

The R^2 is the correlation between measured and predicted WPD, which has the highest value of one. A pronounced observation of R^2 from Tables 7 and 8 revealed a very good correlation between measured and predicted WPD obtained when training and testing the prediction models using the data set from Mersing, Pulau Langkawi, and Bayan Lepas. On the other hand, the measured and predicted WPD of Kuala Terengganu suffers comparatively lower correlation when training and testing the prediction models.

Furthermore, the performance of the developed prediction models is compared to different training and testing data size to illustrate the effect of data size on the prediction accuracy. Tables (9–12) summarize the RMSE, MABE, MAPE, and R^2 when input-output data sets of Mersing, Pulau Langkawi, Bayan Lepas and Kuala Terengganu respectively are applied to the prediction models. In this case, training and testing dataset consists of $P = 92$ (70%) and $Q = 40$ (30%) observations during the period January 2004 to August 2011 and September 2011 to December 2014 respectively. Furthermore, in order to choose the best data size that will provide optimal error in WPD prediction, all the developed models were trained and tested with another input-output data size, which consists of $P = 80$ (60%) and $Q = 52$ (40%) observations during the period January 2004 to August 2010 and September 2010 to December 2014 respectively. The statistical error metrics (RMSE, MABE, MAPE, and R^2) for this category of training and testing data set are also shown in Tables (9–12).

A profound observation on Tables (7–12) reveals that:

- i. For WPD prediction of Mersing, ANFIS-PSO is the best model when training the models and the value of RMSE obtained are 4.96, 5.23 and 5.41 for the data size $P = 105$, $Q = 27$;

Table 6. Descriptive statistical parameters of measured WPD (W/m^2) for different locations.

Parameters	Locations			
	Mersing	Kuala Terengganu	Pulau Langkawi	Bayan Lepas
Max.	152.85	90.35	104.50	78.31
Min.	6.44	4.49	4.07	6.66
St dev.	33.80	17.75	17.28	13.73
Range	146.41	100.43	71.66	85.86

<https://doi.org/10.1371/journal.pone.0193772.t006>

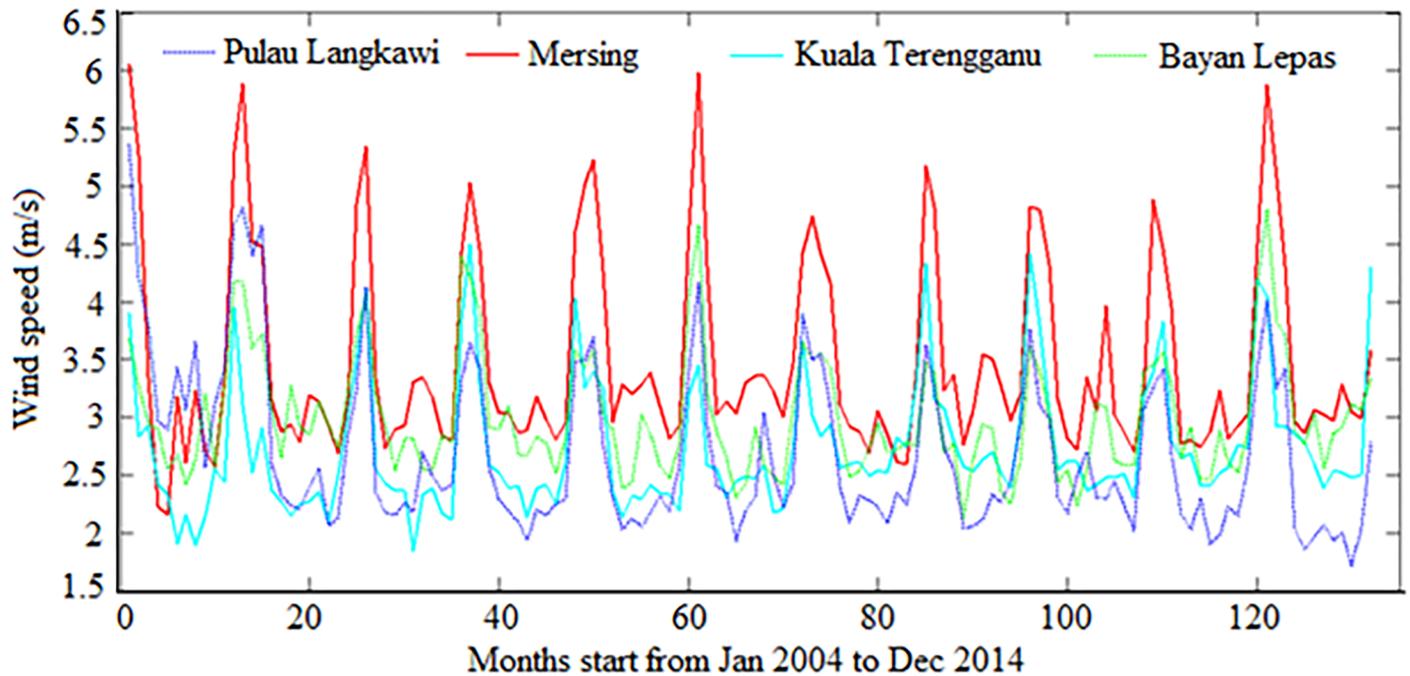


Fig 6. Visual presentation of wind speed at 50m hub height of underlying locations.

<https://doi.org/10.1371/journal.pone.0193772.g006>

P = 92, Q = 40, and P = 80, Q = 52 respectively. On the other hand, when testing the prediction models ANFIS-GA is the best for the data size P = 105, Q = 27 and having RMSE of 5.04 whereas, ANFIS-PSO is the best model when data size P = 92, Q = 40, and P = 80, Q = 52 resulting RMSE of 5.37 and 5.22 respectively.

Table 7. A Statistical model comparison in training the models when P = 105 and Q = 27.

Locations	Rank	Prediction Models	Training			
			RMSE	MAPE (%)	MABE	R ²
Mersing	1	ANFIS-PSO	4.96	5.14	2.61	0.9781
	2	ANFIS-GA	5.41	5.77	2.63	0.9756
	3	ANFIS-DE	5.50	9.70	3.03	0.9749
	4	ANFIS	5.82	10.88	3.07	0.9728
Pulau Langkawi	1	ANFIS-GA	3.13	10.17	1.89	0.9709
	2	ANFIS-PSO	3.16	10.96	1.90	0.9703
	3	ANFIS	3.26	11.70	1.96	0.9684
	4	ANFIS-DE	3.30	10.10	1.90	0.9676
Bayan Lepas	1	ANFIS-PSO	2.14	7.88	1.65	0.9762
	2	ANFIS	2.28	9.06	1.74	0.9709
	3	ANFIS-GA	2.34	8.52	1.72	0.9692
	4	ANFIS-DE	2.54	9.04	1.83	0.9638
Kuala Terengganu	1	ANFIS-PSO	3.73	8.90	1.91	0.9523
	2	ANFIS-GA	4.0	10.30	2.19	0.9449
	3	ANFIS	4.03	14.57	2.29	0.9448
	4	ANFIS-DE	4.37	18.30	2.88	0.9392

<https://doi.org/10.1371/journal.pone.0193772.t007>

Table 8. A Statistical model comparison in testing the models when P = 105 and Q = 27.

Locations	Rank	Prediction models	Testing			
			RMSE	MAPE (%)	MABE	R ²
Mersing	1	ANFIS-GA	5.04	5.25	2.52	0.9701
	2	ANFIS-PSO	5.10	5.88	2.67	0.9691
	3	ANFIS	5.14	6.65	2.74	0.9690
	4	ANFIS-DE	5.56	9.22	3.26	0.9636
Pulau Langkawi	1	ANFIS-PSO	1.53	11.11	1.21	0.9774
	2	ANFIS-DE	1.59	13.07	1.19	0.9775
	3	ANFIS-GA	1.72	10.27	1.23	0.9737
	4	ANFIS	2.17	23.07	1.79	0.9578
Bayan Lepas	1	ANFIS-PSO	2.37	8.17	1.82	0.9749
	2	ANFIS	2.49	9.02	1.92	0.9720
	3	ANFIS-GA	2.75	7.27	1.83	0.9659
	4	ANFIS-DE	2.81	8.84	1.96	0.9643
Kuala Terengganu	1	ANFIS-GA	4.79	13.65	3.06	0.9412
	2	ANFIS-PSO	4.92	10.77	3.54	0.9456
	3	ANFIS-DE	5.01	21.67	3.70	0.9355
	4	ANFIS	5.37	12.97	3.15	0.9261

<https://doi.org/10.1371/journal.pone.0193772.t008>

- ii. For WPD prediction of Bayan Lepas, ANFIS-PSO shows the best performance for all above data size when training the prediction models. In this case, the computed values of RMSE are 2.14, 2.17 and 2.18 when the training data sizes are of P = 105, Q = 27; P = 92, Q = 40, and P = 80, Q = 52 respectively. Then again, ANFIS-PSO shows best performance for a testing data size of P = 105, Q = 27 and P = 92, Q = 40 resulting RMSE of 2.37 and 2.54 respectively. It is not surprising that ANFIS model has the best accuracy when testing data size of P = 80, Q = 52 having RMSE of 2.65.
- iii. For WPD prediction of Pulau Langkawi, ANFIS-GA and ANFIS-PSO show the best performance in training and testing the models respectively with data size of P = 105, Q = 27 which results in RMSE of 3.13 and 1.53 in training and testing respectively. In case of the data size P = 92 and Q = 40, ANFIS-PSO ranks in first both training and validation with RMSE of 3.29 and 1.69 respectively. For the training and testing data size P = 80, Q = 52; ANFIS and ANFIS-GA model show optimal RMSE, which are 3.45 and 2.04 in training and testing respectively.

Table 9. Effect of data size for the prediction wind power density of Mersing.

Data Size	Training						Testing					
	Rank	Prediction models	RMSE	MABE	MAPE (%)	R ²	Rank	Prediction models	RMSE (W)	MABE (W)	MAPE (%)	R ²
P = 92, Q = 40	1	ANFIS-PSO	5.23	2.64	6.88	0.9786	1	ANFIS-PSO	5.37	2.75	5.95	0.9663
	2	ANFIS-GA	5.36	2.92	9.45	0.9712	2	ANFIS-GA	5.46	2.89	6.53	0.9651
	3	ANFIS	5.39	2.92	10.28	0.9767	3	ANFIS	5.55	2.94	7.06	0.9640
	4	ANFIS-DE	5.52	2.81	7.84	0.9756	4	ANFIS-DE	5.72	2.84	5.68	0.9617
P = 80, Q = 52	1	ANFIS-PSO	5.41	2.82	6.40	0.9780	1	ANFIS-PSO	5.22	2.79	6.40	0.9662
	2	ANFIS-GA	5.59	3.33	8.95	0.9765	2	ANFIS-DE	5.33	3.58	11.56	0.9652
	3	ANFIS	5.82	3.13	12.46	0.9745	3	ANFIS-GA	5.46	3.34	8.95	0.9634
	4	ANFIS-DE	6.21	4.29	11.56	0.9709	4	ANFIS	5.48	3.12	7.55	0.9632

<https://doi.org/10.1371/journal.pone.0193772.t009>

Table 10. Effect of data size for the prediction wind power density of Bayan Lepas.

Data Size	Training						Testing					
	Rank	Prediction models	RMSE	MABE	MAPE (%)	R ²	Rank	Prediction models	RMSE (W)	MABE (W)	MAPE (%)	R ²
P = 92, Q = 40	1	ANFIS-PSO	2.17	1.64	7.85	0.9748	1	ANFIS-PSO	2.54	1.74	7.83	0.9684
	2	ANFIS-GA	2.22	1.69	7.99	0.9740	2	ANFIS-GA	2.62	1.78	7.67	0.9584
	3	ANFIS	2.24	1.72	8.95	0.9698	3	ANFIS	2.73	1.93	9.95	0.9541
	4	ANFIS-DE	2.38	1.73	7.98	0.9738	4	ANFIS-DE	2.80	1.87	8.25	0.9577
P = 80, Q = 52	1	ANFIS-PSO	2.18	1.56	7.29	0.9770	1	ANFIS	2.65	1.96	11.67	0.9441
	2	ANFIS	2.25	1.73	8.35	0.9756	2	ANFIS-PSO	2.74	1.98	11.14	0.9409
	3	ANFIS-GA	2.46	1.77	7.71	0.9707	3	ANFIS-DE	2.77	2.18	11.76	0.9398
	4	ANFIS-DE	3.34	2.34	10.23	0.9461	4	ANFIS-GA	3.38	2.0	9.39	0.9252

<https://doi.org/10.1371/journal.pone.0193772.t010>

iv. For WPD prediction of Kuala Terengganu, ANFIS-PSO and ANFIS-GA show the best performance in training and testing the models respectively with data size of P = 105, Q = 27 which results in RMSE of 3.73 and 4.79 respectively. For the data size P = 92, Q = 40; ANFIS model shows optimal RMSE, which is 3.45 in training stage and ANFIS-PSO presents optimal RMSE of 2.23 in the testing stage. In case of data size P = 80 and Q = 52, ANFIS-PSO ranks in first both training and validation of the prediction models with RMSE of 3.99 and 4.58 respectively.

Based on the above discussion, it is clear that P = 105, Q = 27 data size applied to the prediction models provides minimal error for WPD prediction of all underlying locations. For this data size, overall ANFIS-PSO and ANFIS-GA showed a higher correlation between measured and predicted WPD all locations both in training and testing stages.

On the other hand, when P = 92, Q = 40 data set are applied to the prediction models, ANFIS-PSO shows the best performance in both training and testing stages for predicting WPD of Mersing, Pulau Langkawi, and Bayan Lepas. However, when predicting WPD of Kuala Terengganu, ANFIS-PSO ranks second in training and first in the testing stage. It is important to mention that P = 80, Q = 52 data size shows the largest error for predicting WPD of all underlying locations, and therefore, we do not want to use P = 80, Q = 52 data set for model performance justification.

Figs 7–10 depicts the visual presentation of measured and predicted WPD when testing using ANFIS-PSO and ANFIS-GA for Mersing, Bayan Lepas, Pulau Langkawi and Kuala Terengganu respectively with best data size i.e. P = 105, Q = 27.

However, both ANFIS-PSO and ANFIS-GA seem to be able to provide an overall good performance both in training and testing stages. Therefore, hybrid ANFIS, especially ANFIS-PSO

Table 11. Effect of data size for the prediction wind power density of Pulau Langkawi.

Data Size	Training						Testing					
	Rank	Prediction models	RMSE	MABE	MAPE (%)	R ²	Rank	Prediction models	RMSE (W)	MABE (W)	MAPE (%)	R ²
P = 92, Q = 40	1	ANFIS-PSO	3.29	1.92	10.28	0.9707	1	ANFIS-PSO	1.69	1.38	12.28	0.9717
	2	ANFIS-GA	3.34	1.92	10.09	0.9698	2	ANFIS-GA	1.71	1.31	11.89	0.9707
	3	ANFIS	3.38	2.03	11.60	0.9690	3	ANFIS-DE	1.88	1.44	11.55	0.9650
	4	ANFIS-DE	3.43	2.21	13.08	0.9682	4	ANFIS	2.23	1.84	20.39	0.9508
P = 80, Q = 52	1	ANFIS	3.45	2.14	11.64	0.9705	1	ANFIS-GA	2.04	1.64	18.18	0.9572
	2	ANFIS-GA	3.58	2.19	12.61	0.9668	2	ANFIS	2.20	1.66	17.19	0.9501
	3	ANFIS-DE	4.06	3.04	19.24	0.9591	3	ANFIS-PSO	2.37	1.68	13.91	0.9421
	4	ANFIS-PSO	4.15	2.55	11.60	0.9572	4	ANFIS-DE	3.23	2.68	26.17	0.8921

<https://doi.org/10.1371/journal.pone.0193772.t011>

Table 12. Effect of data size for the prediction wind power density of Kaula Terengganu.

Data Size	Training						Testing					
	Rank	Prediction models	RMSE	MABE	MAPE (%)	R ²	Rank	Prediction models	RMSE (W)	MABE (W)	MAPE (%)	R ²
P = 92, Q = 40	1	ANFIS	4.07	2.28	12.61	0.9407	1	ANFIS-PSO	4.15	2.72	12.74	0.9436
	2	ANFIS-PSO	4.15	2.32	11.61	0.9385	2	ANFIS-DE	4.24	2.50	13.07	0.9531
	3	ANFIS-GA	4.32	2.33	12.40	0.9421	3	ANFIS	4.61	2.68	11.83	0.9448
	4	ANFIS-DE	4.39	2.37	12.06	0.9307	4	ANFIS-GA	5.16	2.78	11.95	0.9307
P = 80, Q = 52	1	ANFIS-PSO	3.99	2.05	10.06	0.9438	1	ANFIS-PSO	4.58	2.85	12.43	0.9405
	2	ANFIS-GA	4.06	2.36	13.13	0.9415	2	ANFIS	4.73	3.27	16.18	0.9307
	3	ANFIS	4.11	2.23	12.85	0.9401	3	ANFIS-DE	4.92	3.74	24.15	0.9363
	4	ANFIS-DE	4.70	3.04	21.10	0.9217	4	ANFIS-GA	4.94	3.64	19.35	0.9312

<https://doi.org/10.1371/journal.pone.0193772.t012>

and ANFIS-GA can be suggested for practical utilization in WPD prediction for the locations having similar wind resource conditions. The Fig 11 shows the error distribution of WPD prediction using ANFIS-PSO and ANFIS-GA for underlying locations. According to the definition of relative percentage error (RPE) presented in [27, 28], the RPE falls in an interval of -10% to 10% can be considered acceptable. The computed value of RPE presented in the Fig 11 is obtained from the proposed ANFIS-PSO and ANFIS-GA when 27 months testing data set of underlying locations applied to the models. It can be observed that most of the wind power

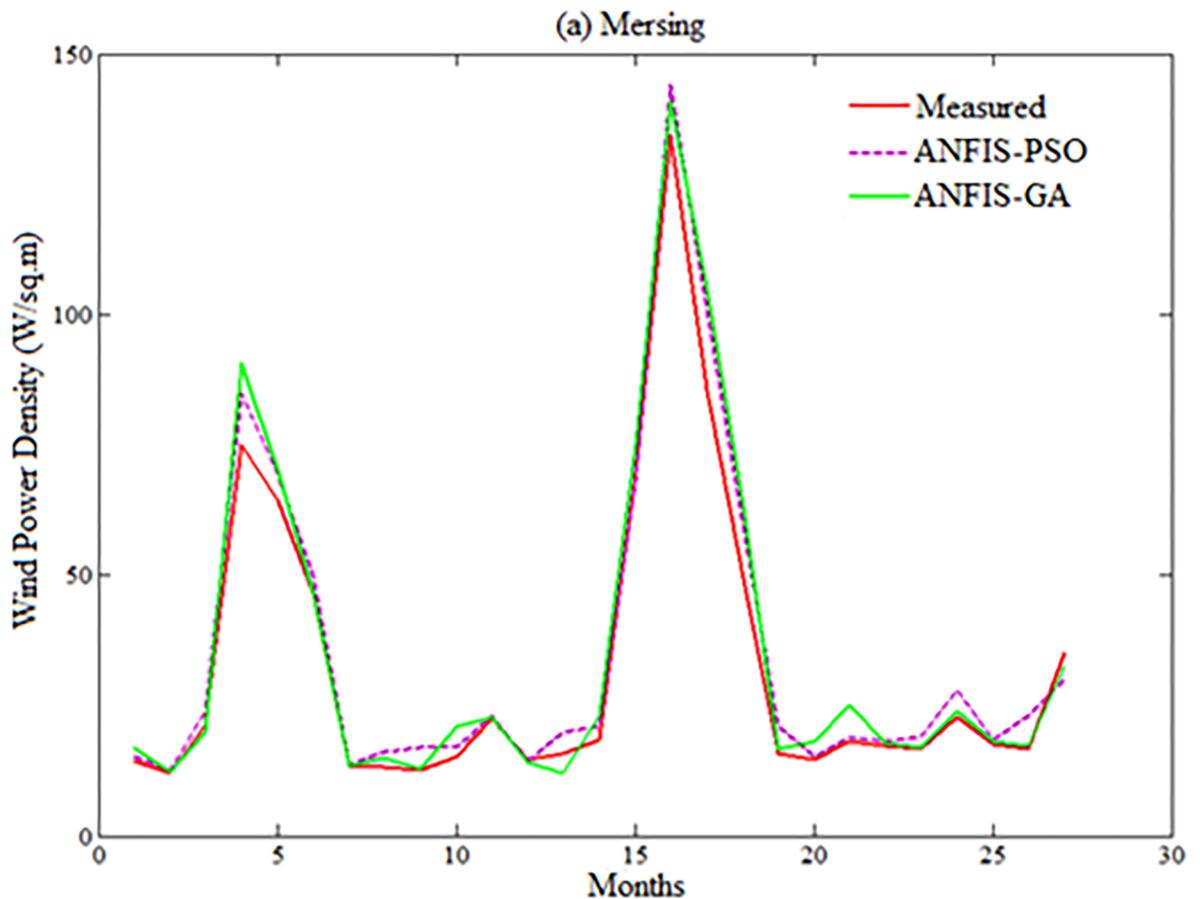


Fig 7. Comparison of WPD prediction from proposed methods when testing with measured WPD value for Mersing.

<https://doi.org/10.1371/journal.pone.0193772.g007>

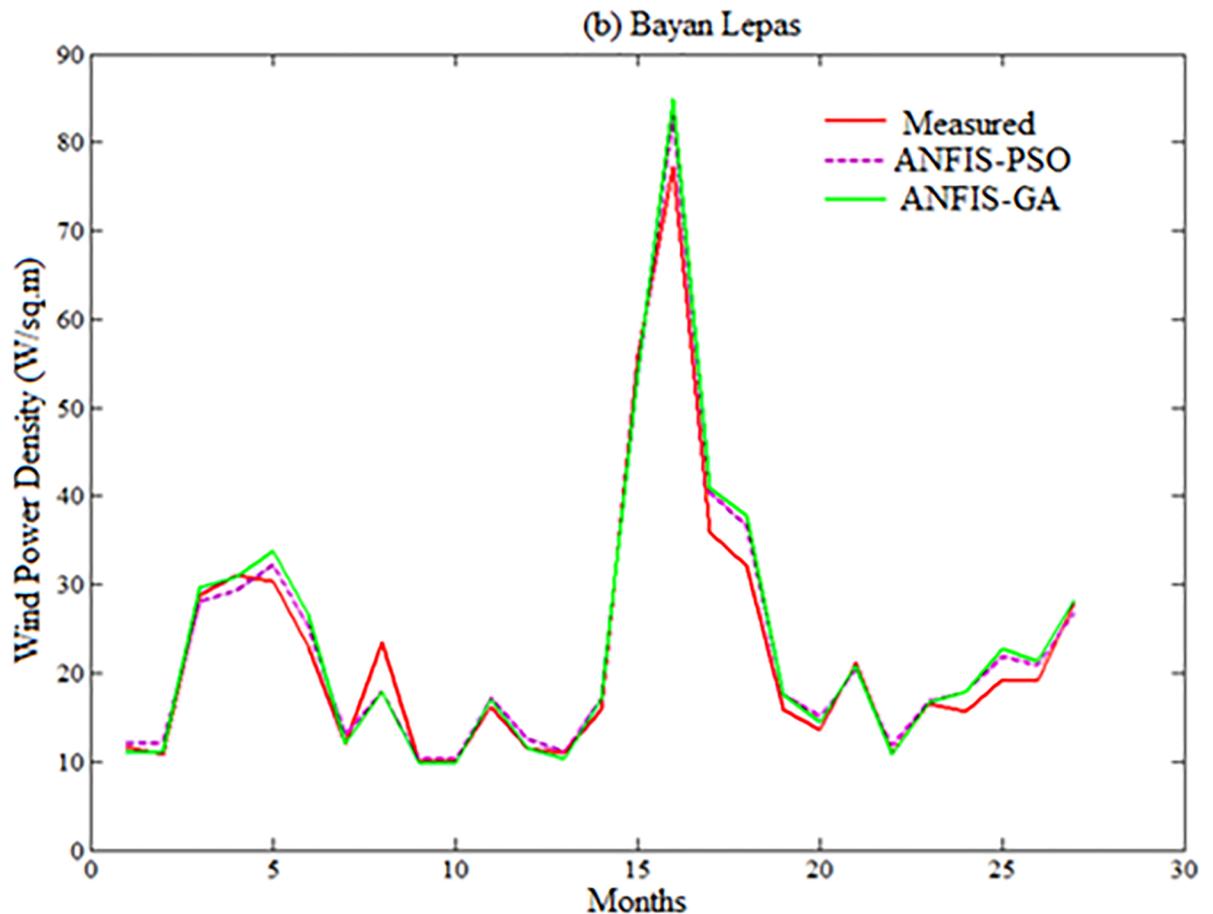


Fig 8. Comparison of WPD prediction from proposed methods when testing with measured WPD value for Bayan Lepas.

<https://doi.org/10.1371/journal.pone.0193772.g008>

values obtained via the proposed ANFIS-PSO and ANFIS-GA model fall within the range of -5% up to 5%. In case of Mersing and Kuala Terengganu, only one prediction RPE falls outside (-10% to 10%) range for both ANFIS-PSO and ANFIS-GA.

Based on the above discussion, the performance of ANFIS-PSO can be considered the best among other the models. The Figs 12 and 13 show the visual presentation of regression analysis of measured and predicted WPD computed from ANFIS-PSO for all underlying locations when $P = 105$, $Q = 27$. The Fig 12 shows the correlation between actual and predicted WPD obtained from ANFIS-PSO method. The correlation coefficient (R^2) is the agreement between measured and predicted values, which has the highest value of one. The R^2 obtained when training proposed the ANFIS-PSO model are 0.9781, 0.9703, 0.9762 and 0.9523 for Mersing, Pulau Langkawi, Bayan Lepas and Kuala Terengganu respectively, which indicates a very high and adequate prediction performance of ANFIS-PSO model.

Again, the linear regression analysis presented in Fig 13 shows the correlation between actual and predicted WPD obtained from ANFIS-PSO when testing the prediction models with data from different underlying locations. The R^2 obtained while testing the proposed ANFIS-PSO model for Mersing, Pulau Langkawi, Bayan Lepas and Kuala Terengganu are 0.9691, 0.9774, 0.9749 and 0.9456 respectively, which supports that ANFIS-PSO has a high precision for the prediction of WPD.

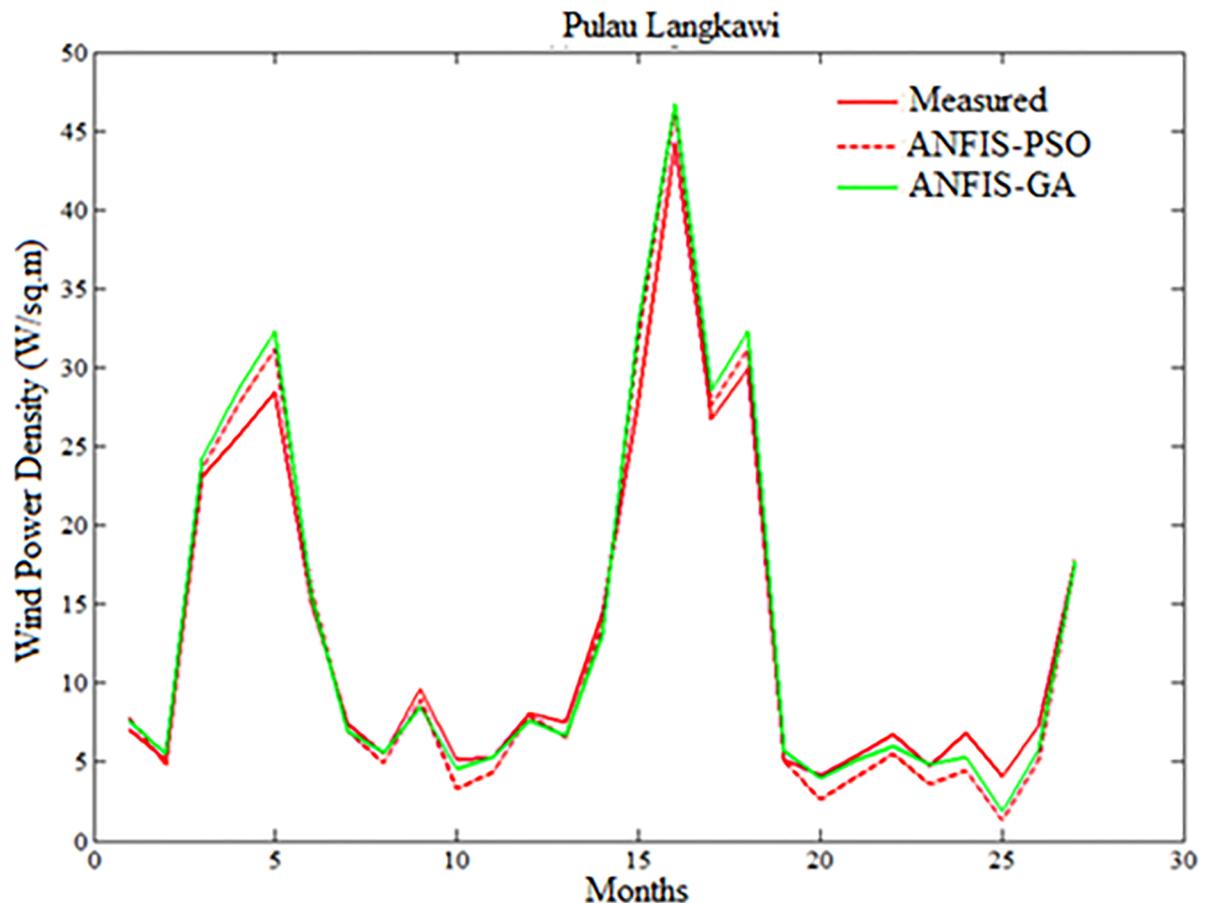


Fig 9. Comparison of WPD prediction of the proposed method with measured value for Pulau Langkawi.

<https://doi.org/10.1371/journal.pone.0193772.g009>

Weekly wind power density prediction

The detailed explanation of data collection and site description can be found in section 2. For weekly wind power density prediction, 11 year's (2004–2014) long-term daily average wind speed data were converted into the weekly average for all underlying locations in Malaysia namely Mersing, Kuala Terengganu, Pulau Langkawi and Bayan Lepas. Afterward, the weekly mean wind speed at 50m and the corresponding wind power density from measured data were applied on the developed standalone ANFIS and hybrid ANFIS models. The first 80% data and rest 20% data for training and testing were used respectively. The results presented in Figs (14–19) mainly obtained when training and testing the ANFIS-PSO and ANFIS-GA models as they were best prediction models for monthly wind power density prediction presented in section 5.1.

Figs 14 and 15 show the error distribution, RMSE, and MAPE whereas, Fig 16 presents regression analysis when training and testing data sets of Mersing were applied on ANFIS-PSO and ANFIS-GA prediction models. It can be observed from Figs 15 and 16 that most of the wind power values obtained via the proposed ANFIS-PSO and ANFIS-GA model fall within the range of -5% up to +5%. The RMSE were 5.72 and 4.82 when testing the ANFIS-PSO and ANFIS-GA respectively. It is mentioned in section 5.1 that the R^2 is the correlation between measured and predicted WPD, which has the highest value of one. In Fig 17, the R^2 were 0.9899 and 0.9703 when testing the ANFIS-PSO and ANFIS-GA respectively.

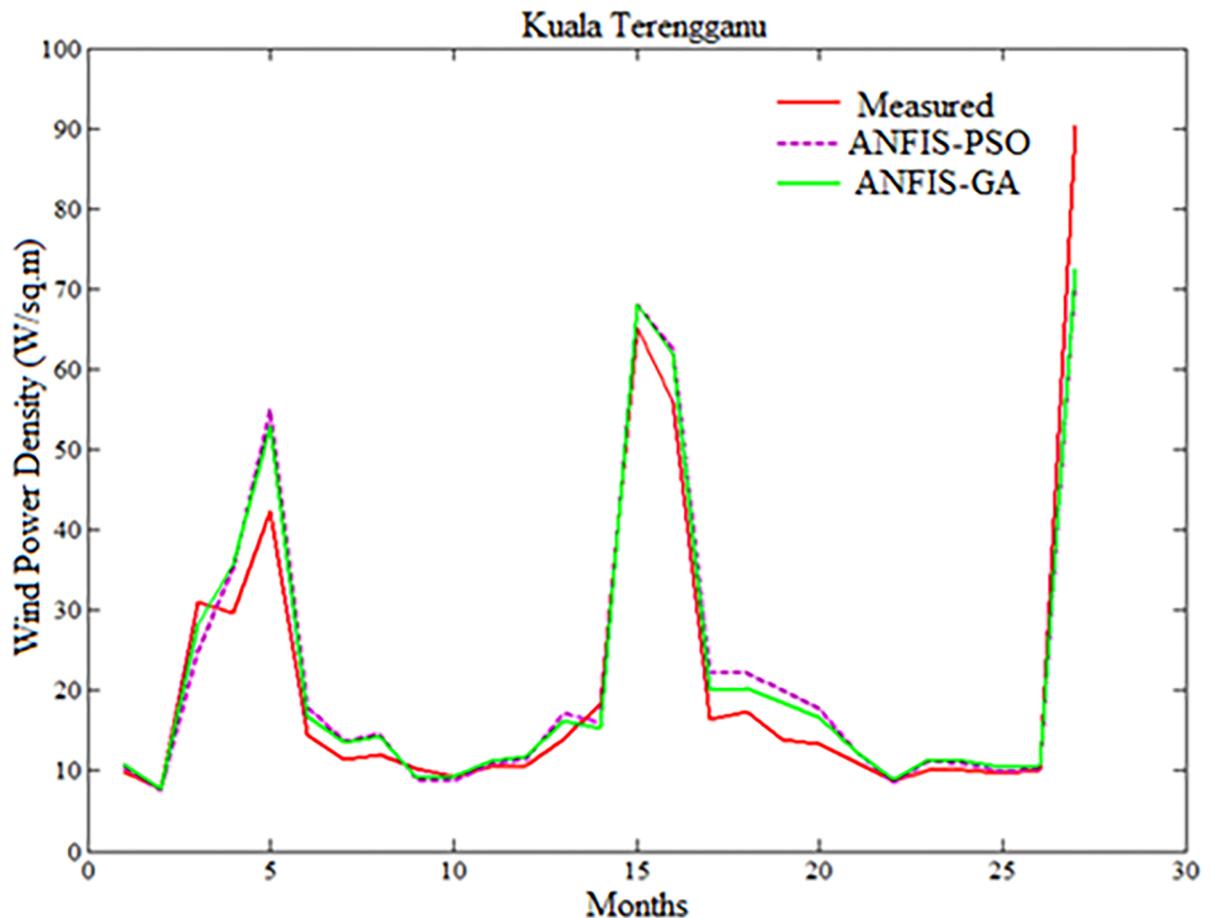


Fig 10. Comparison of WPD prediction of proposed methods with measured value for Kuala Terengganu.

<https://doi.org/10.1371/journal.pone.0193772.g010>

Similarly, Figs 17 and 18 show the error distribution, RMSE, and MAPE whereas, Fig 19 presents regression analysis when training and testing data sets of Langkawi were applied on ANFIS-PSO and ANFIS-GA prediction models. It can be observed from Figs 18 and 19 that most of the wind power values obtained via the proposed ANFIS-PSO and ANFIS-GA model fall within the range of -5% up to +5%. The RMSE were 1.76 and 2.11 when testing the ANFIS-PSO and ANFIS-GA respectively. It is mentioned in section 5.1 that the R^2 is the correlation between measured and predicted WPD, which has the highest value of one. In Fig 19, the R^2 were 0.9844 and 0.9701 when testing the ANFIS-PSO and ANFIS-GA respectively.

Extrapolation capabilities of the proposed models

Measured wind speed data are not available for many locations in Malaysia, including remote islands (less than 200 km²) and decentralized places to identify possible wind energy applications. In this dissertation, extrapolation capabilities of the wind speed of the proposed hybrid ANFIS models have been examined for a location in Tioman Island having latitude 2° 48' 30" N and longitude 104° 8' 29" E, where measured wind data are not available. And then, the result is compared with the measured wind data at Mersing station (latitude 2° 27' N and longitude 103° 50' E) which is the nearby station having similar climate conditions. As the study location does not have any actual measured meteorological data, the daily average solar

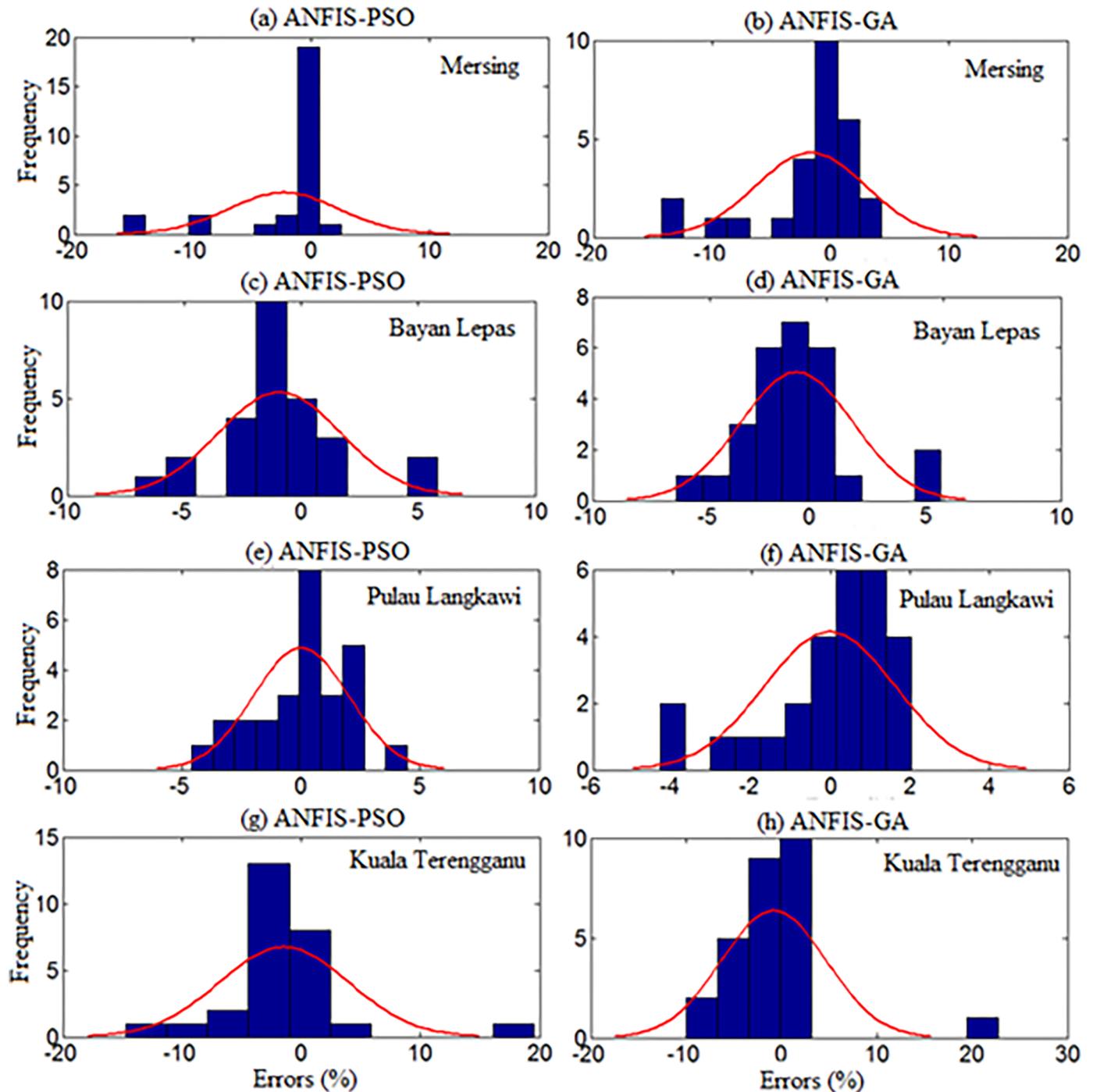


Fig 11. Error distribution when testing ANFIS-PSO and ANFIS-GA for (a, b) Mersing, (c, d) Bayan Lepas (e, f) Pulau Langkawi, (g, h) Kuala Terengganu.

<https://doi.org/10.1371/journal.pone.0193772.g011>

radiation (kWh/m²/day), daily average air temperature, maximum and minimum air temperature, air pressure, relative humidity and altitude data were collected from NASA surface meteorology and solar energy database for the whole year of 2004 for the prediction of daily average wind speed (m/s). The Figs 20 and 21 show the prediction of daily average wind speed for Tioman Island using ANFIS-PSO and ANFIS-GA respectively, which are compared with

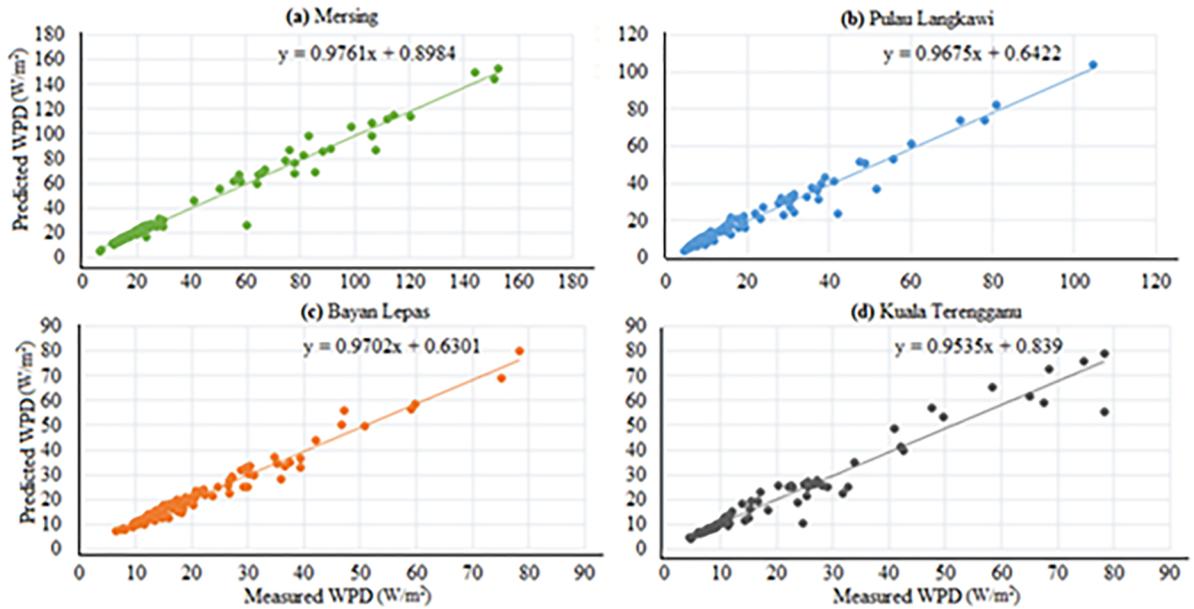


Fig 12. Measured versus predicted WPD when training ANFIS-PSO model (a) Mersing, (b) Pulau Langkawi, (c) Bayan Lepas and (d) Kuala Terengganu.

<https://doi.org/10.1371/journal.pone.0193772.g012>

measured wind speed at Mersing meteorological station that was measured at 50m above the sea level in 2004.

From error analysis, it was found that the MAPE and MABE for ANFIS-PSO were 31.56%, 0.9672% respectively whereas, they were found to be 32.21% and 0.9872% for ANFIS-GA respectively. It can be observed from the figures that the wind characteristics in Tioman Island,

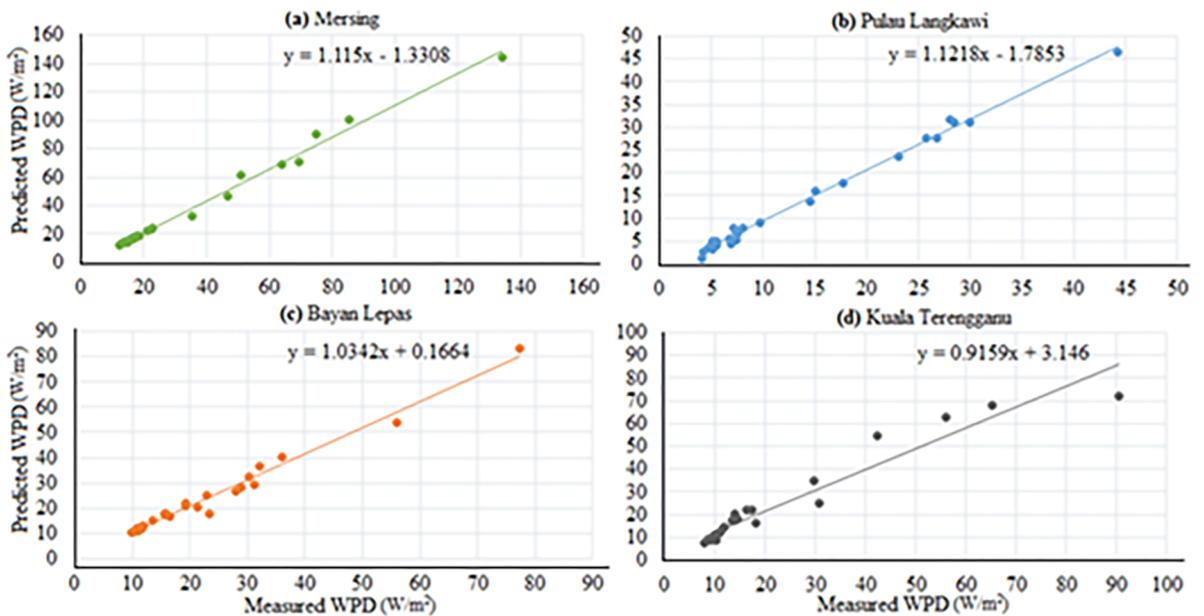


Fig 13. Measured versus predicted WPD when testing ANFIS-PSO model (a) Mersing, (b) Pulau Langkawi, (c) Bayan Lepas and (d) Kuala Terengganu.

<https://doi.org/10.1371/journal.pone.0193772.g013>

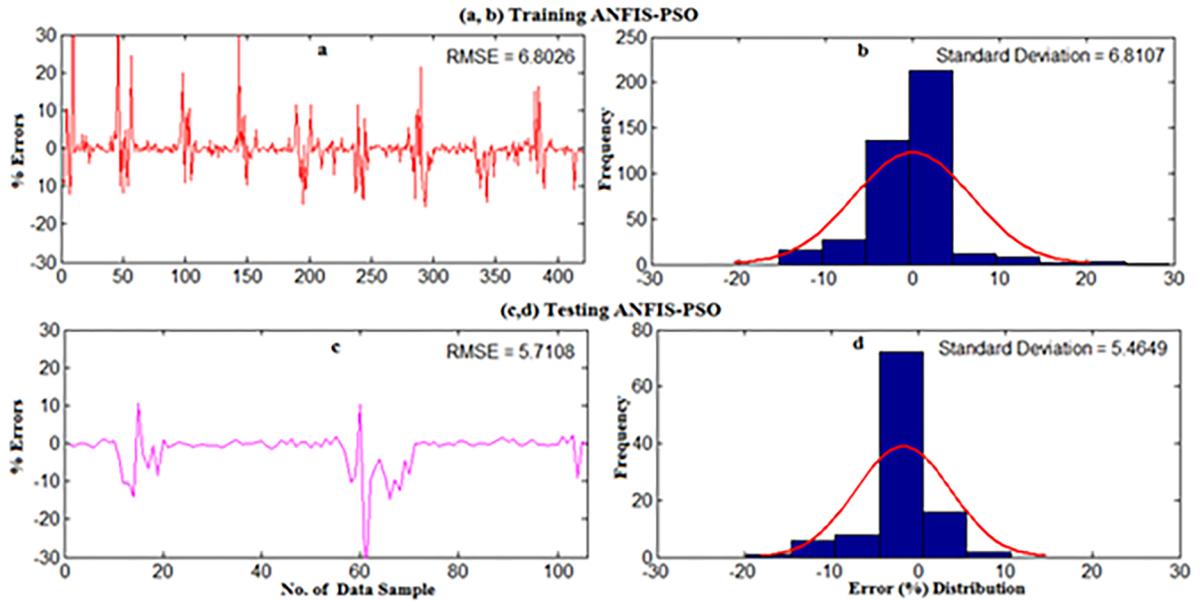


Fig 14. Error distribution when (a,b) Training and (c,d) Testing the ANFIS-PSO model for weekly WPD prediction of Mersing.

<https://doi.org/10.1371/journal.pone.0193772.g014>

obtained with the proposed ANFI-PSO and ANFIS-GA, is not exactly similar with the measured wind at Mersing in many cases. However, it is important to take note that the performance of the prediction models can be justified accurately when measured wind data in Tioman Island will be available for comparison.

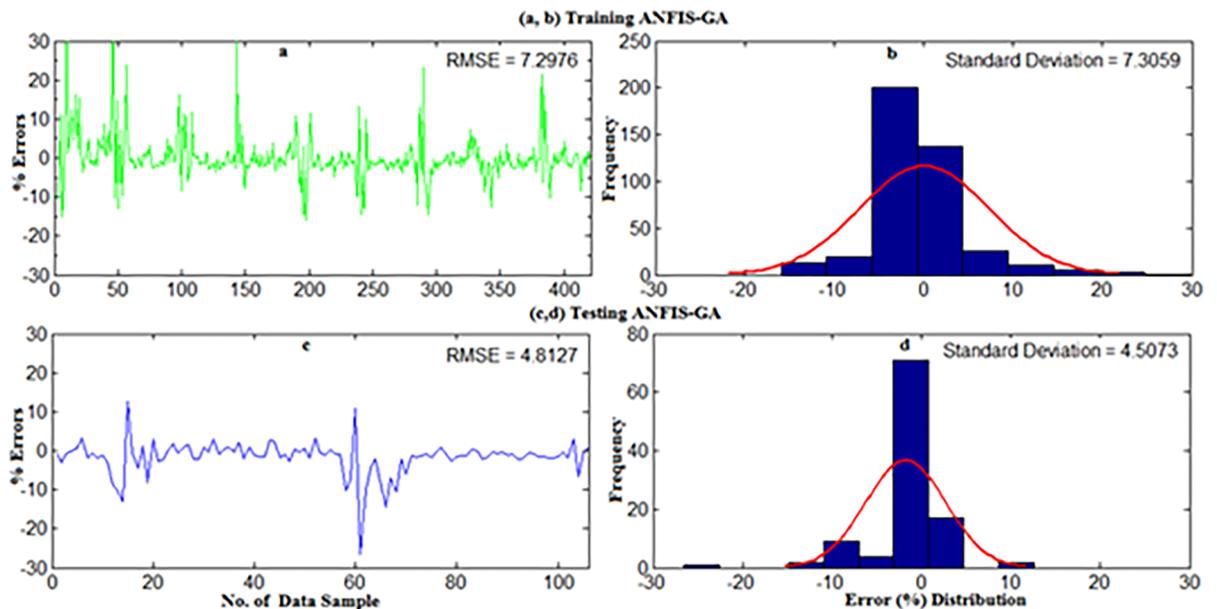


Fig 15. Error distribution when (a,b) Training and (c,d) Testing the ANFIS-GA model for weekly WPD prediction of Mersing.

<https://doi.org/10.1371/journal.pone.0193772.g015>

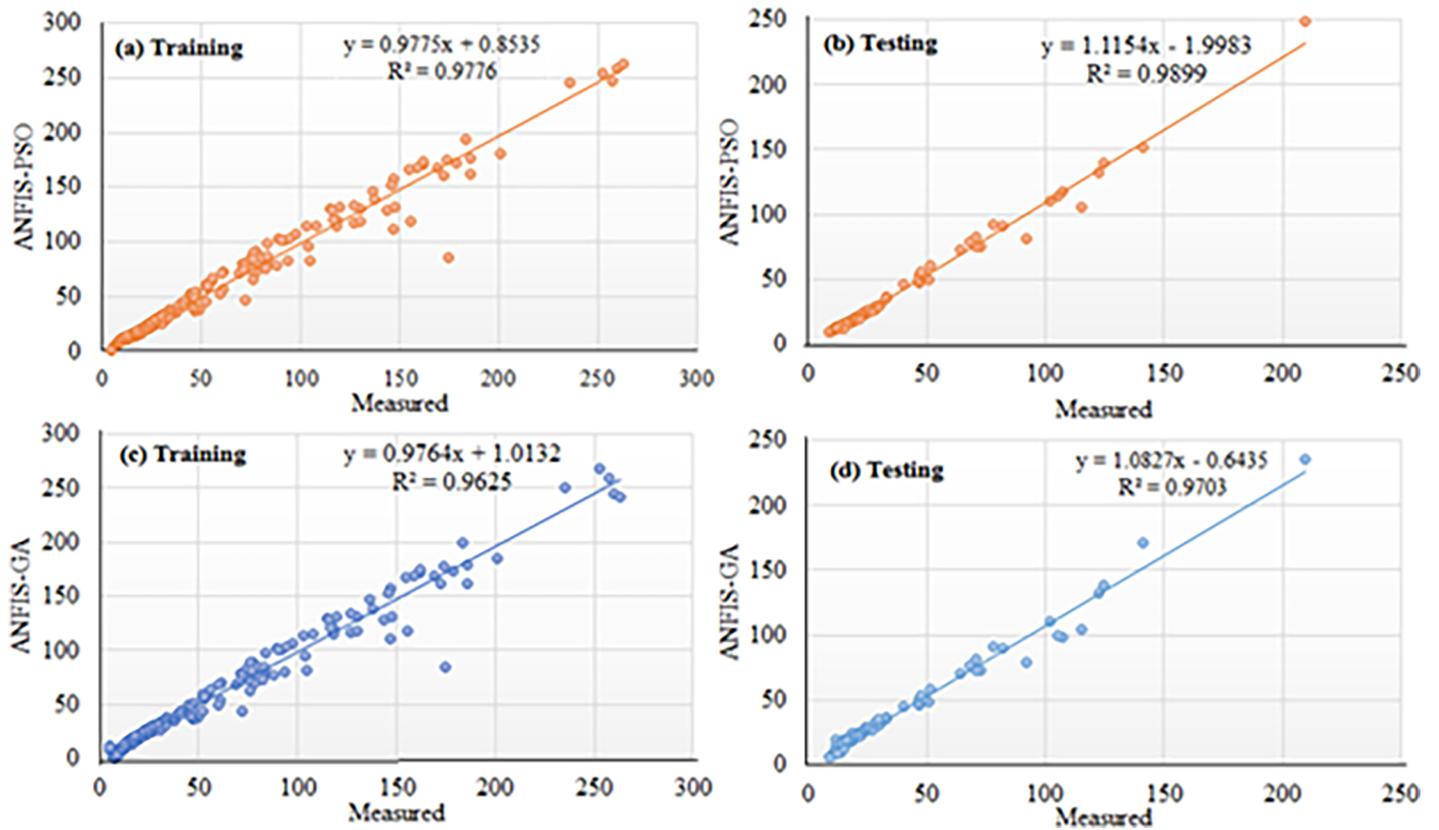


Fig 16. Regression plot of weekly WPD prediction for Mersing when training and testing the ANFIS-PSO and ANFIS-GA prediction model.

<https://doi.org/10.1371/journal.pone.0193772.g016>

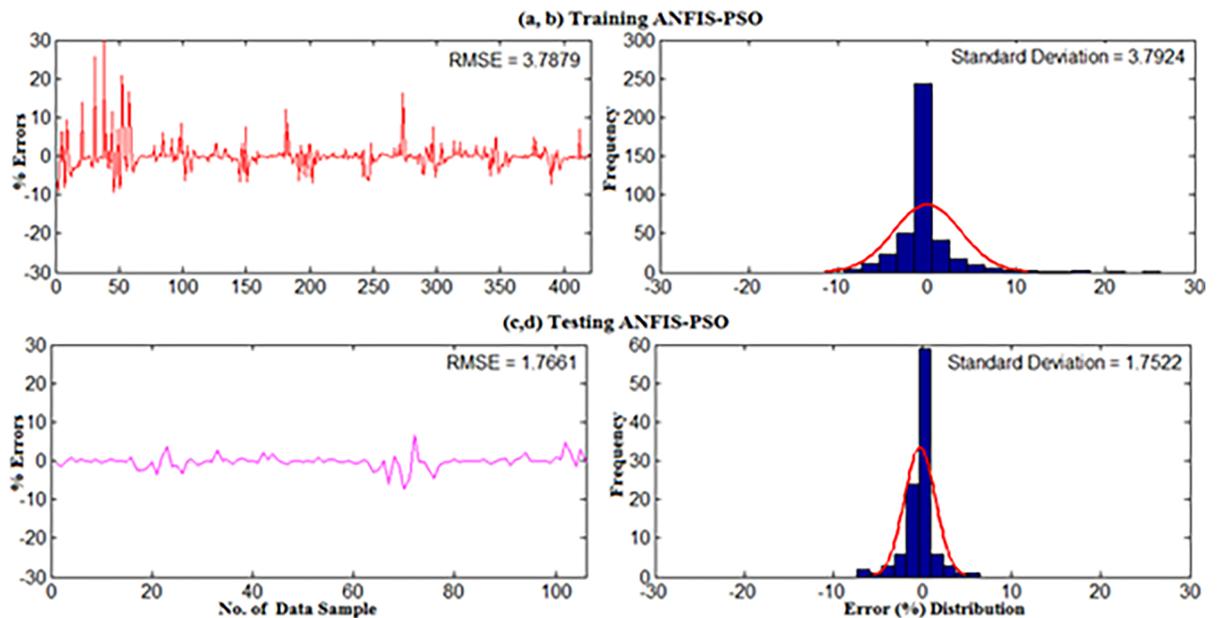


Fig 17. Error distribution when (a,b) Training and (c,d) Testing the ANFIS-PSO model for weekly WPD prediction of Pulau Langkawi.

<https://doi.org/10.1371/journal.pone.0193772.g017>

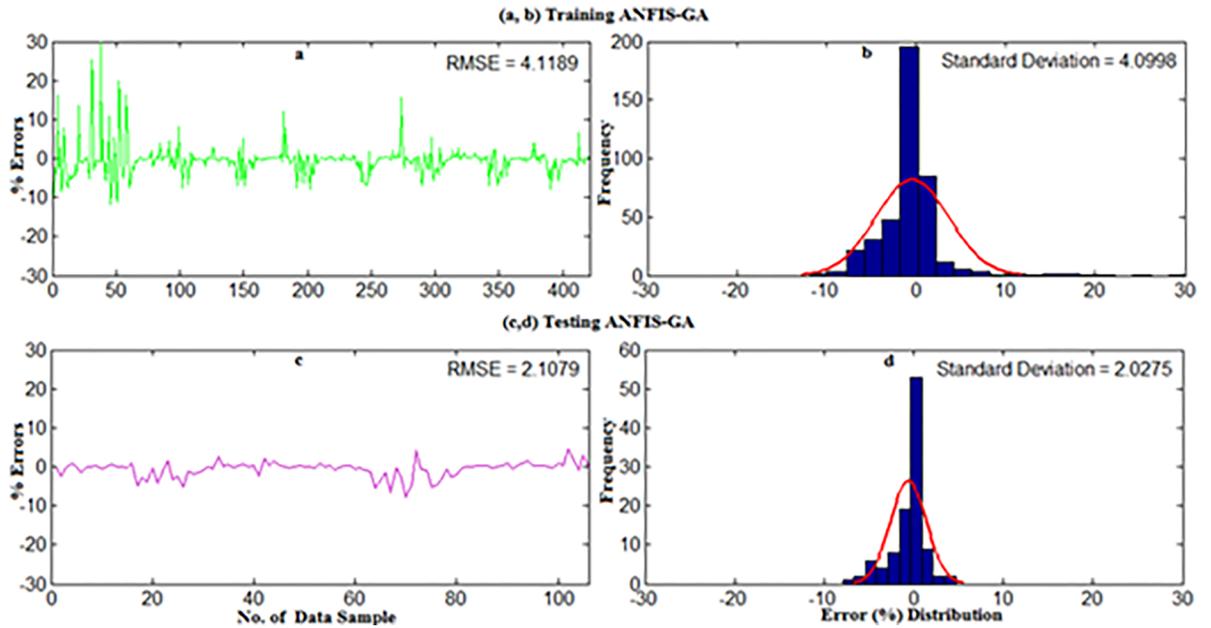


Fig 18. Error distribution when (a, b) Training and (c, d) Testing the ANFIS-GA model for weekly WPD prediction of Pulau Langkawi.

<https://doi.org/10.1371/journal.pone.0193772.g018>

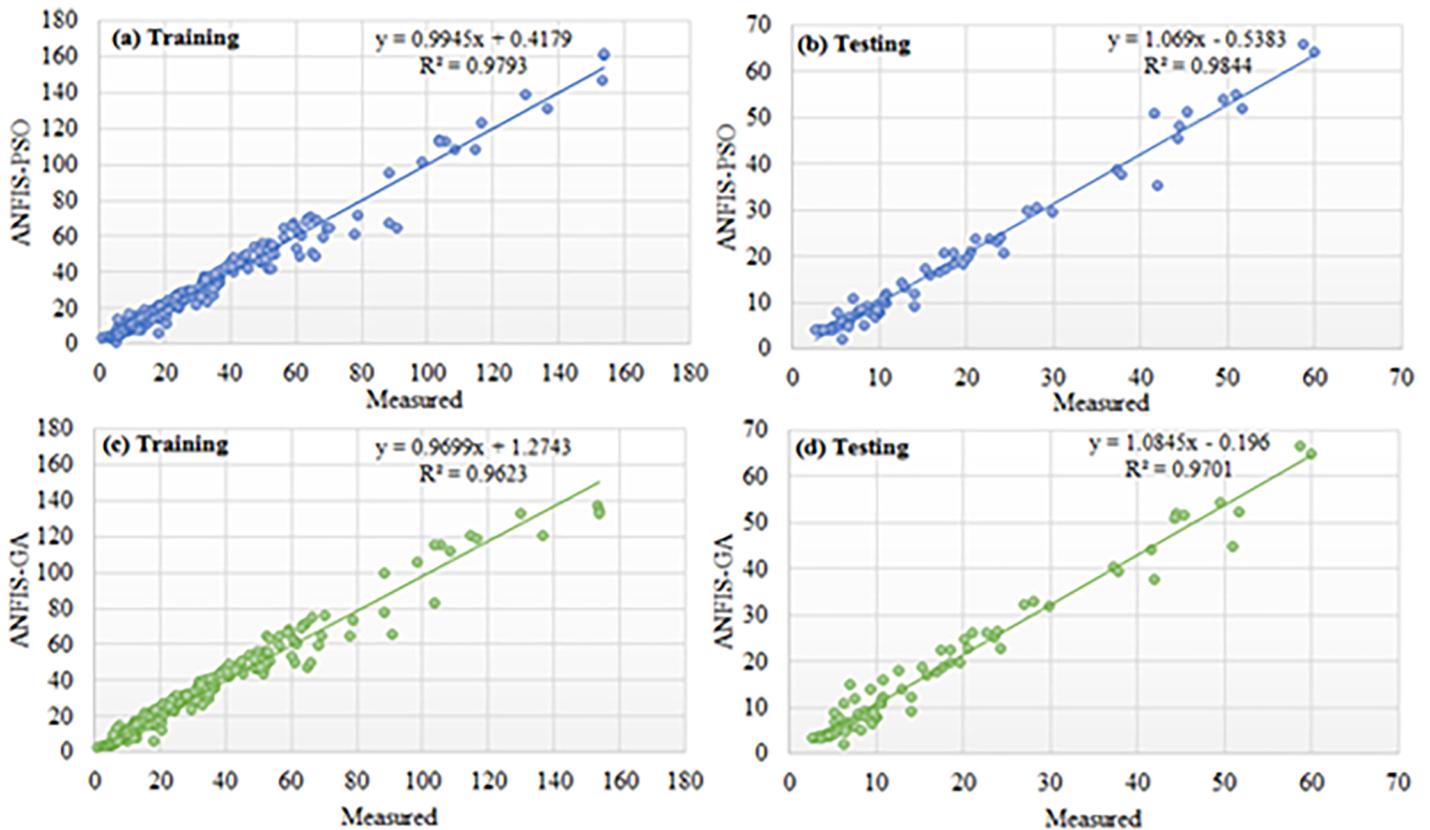


Fig 19. Regression plot of weekly WPD prediction for Pulau Langkawi when training and testing the ANFIS-PSO and ANFIS-GA prediction model.

<https://doi.org/10.1371/journal.pone.0193772.g019>

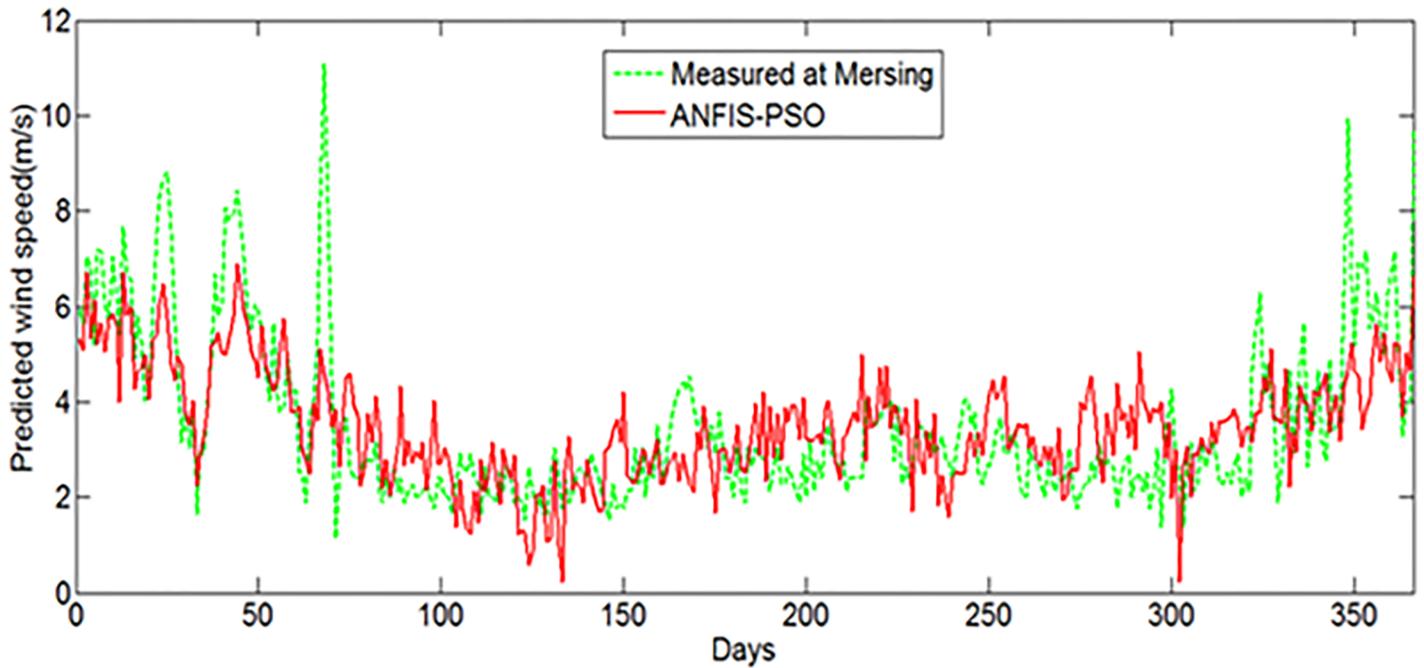


Fig 20. Extrapolation of daily average wind speed for Tioman Island using proposed ANFIS-PSO.

<https://doi.org/10.1371/journal.pone.0193772.g020>

Conclusion

The wind energy potential assessment is very important for independent power producer and governmental organization to determine how efficiently wind power can be extracted from a certain location. The wind power density (WPD) is the key assessment parameter in wind

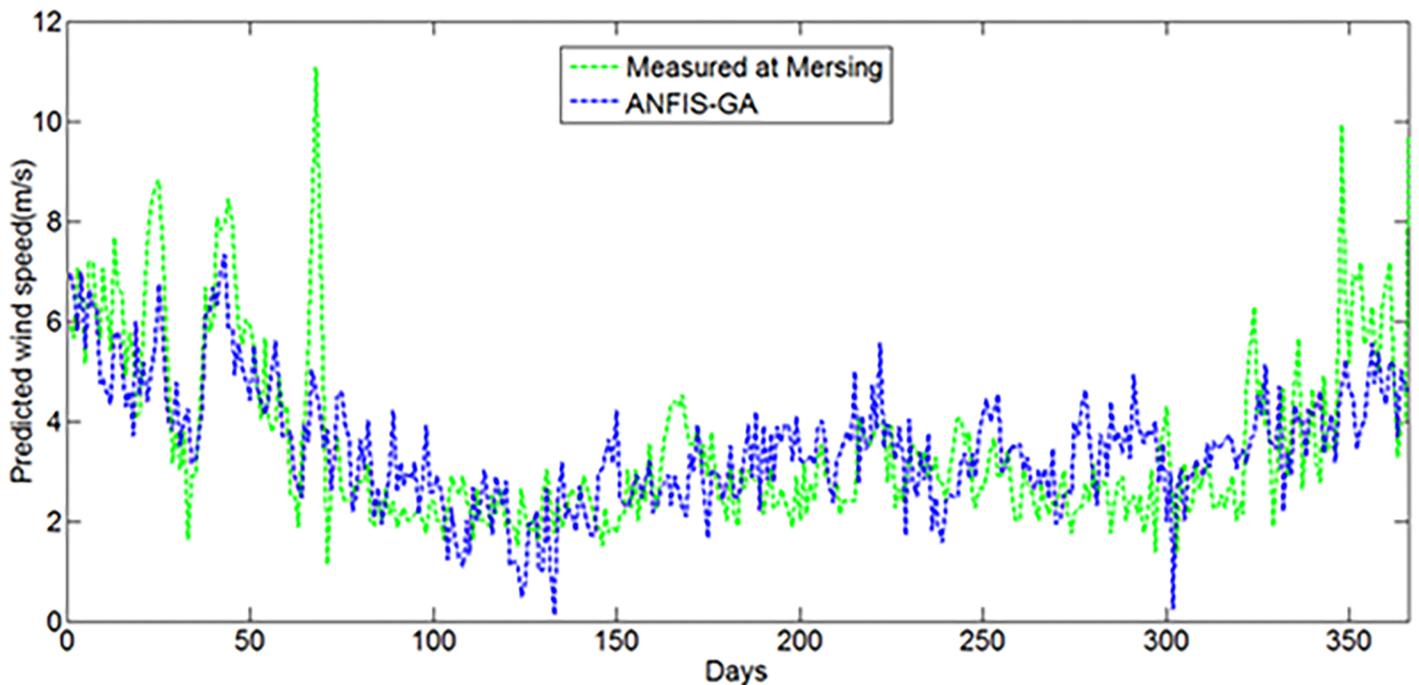


Fig 21. Extrapolation of daily average wind speed for Tioman Island using proposed ANFIS-GA.

<https://doi.org/10.1371/journal.pone.0193772.g021>

potentiality analysis. Therefore, an efficient soft computing technique based on ANFIS-PSO, ANFIS-GA, ANFIS-DE and standalone ANFIS prediction models were developed in this paper to predict long-term (monthly and weekly) average wind power density of four different locations in Malaysia. The choice of the ANFIS technique was made due to its simplicity, reliability as well as its efficient computational capability; its ease of adaptability to optimization and other adaptive techniques, and its adaptability in handling complex parameters. The most significant advantage of hybrid ANFIS is that PSO/GA/DE tune the membership functions of the ANFIS model to ensure minimum error. The prediction models were trained and tested using wind speed data collected from meteorological stations of the underlying locations and measured wind power density. Moreover, different training and testing data size were applied to the prediction models to obtain best data size that provides a minimal error. The first 80% of data used for training and remaining 20% data for testing provide the optimal error in WPD prediction. Based on the result from best data size, there is no model that performed uniformly superior to other for all locations in both training and testing stages. Overall, ANFIS-PSO and ANFIS-GA out-performed ANFIS standalone and ANFIS-DE. Therefore, the results and analysis confirmed that the proposed hybrid ANFIS, especially ANFIS-PSO and ANFIS-GA have the excellent capability to predict the WPD with higher accuracy and precision. Other soft computing techniques applicable to wind speed and power density prediction for other parts of the world can be developed and compare with hybrid ANFIS in the further study.

Supporting information

S1 Dataset. Supporting Information.rar.
(RAR)

Acknowledgments

The authors would like to acknowledge the financial support received from the University of Malaya, Malaysia, through Frontier Research Grant No. FG007-17AFR and Innovative Technology Grant No. RP043B-17AET.

Author Contributions

Conceptualization: Monowar Hossain.

Data curation: Monowar Hossain.

Formal analysis: Monowar Hossain, Laith M. Halabi.

Funding acquisition: Saad Mekhilef, Mehdi Seyedmahmoudian, Ben Horan, Alex Stojcevski.

Methodology: Monowar Hossain, Firdaus Afifi.

Supervision: Saad Mekhilef.

Writing – original draft: Monowar Hossain.

Writing – review & editing: Monowar Hossain, Saad Mekhilef, Laith M. Halabi, Lanre Olatomiwa, Mehdi Seyedmahmoudian, Ben Horan, Alex Stojcevski.

References

1. Hossain M., Mekhilef S., Danesh M., Olatomiwa L., and Shamshirband S., "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems," *Journal of Cleaner Production*, vol. 167, pp. 395–405, 2017/11/20/ 2017.

2. Hu X., Zou Y., and Yang Y., "Greener plug-in hybrid electric vehicles incorporating renewable energy and rapid system optimization," *Energy*, vol. 111, pp. 971–980, 2016/09/15/ 2016.
3. Hossain M., Mekhilef S., and Olatomiwa L., "Performance evaluation of a stand-alone PV-wind-diesel-battery hybrid system feasible for a large resort center in South China Sea, Malaysia," *Sustainable Cities and Society*, vol. 28, pp. 358–366, 1// 2017.
4. Zhang L., Hu X., Wang Z., Sun F., and Dorrell D. G., "Experimental impedance investigation of an ultra-capacitor at different conditions for electric vehicle applications," *Journal of Power Sources*, vol. 287, pp. 129–138, 2015/08/01/ 2015.
5. Zhang L., Hu X., Wang Z., Sun F., and Dorrell D. G., "A review of supercapacitor modeling, estimation, and applications: A control/management perspective," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1868–1878, 2018/01/01/ 2018.
6. Global wind report on annual market update 2015. Accessed on 15.09.2016. <http://www.gwec.net/publications/global-wind-report-2/global-wind-report-2014-annual-market-update/>.
7. Azad H. B., Mekhilef S., and Ganapathy V. G., "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Transactions on Sustainable Energy*, vol. 5, pp. 546–553, 2014.
8. Huffaker R. and Bittelli M., "A Nonlinear Dynamics Approach for Incorporating Wind-Speed Patterns into Wind-Power Project Evaluation," *PloS one*, vol. 10, p. e0115123, 2015. <https://doi.org/10.1371/journal.pone.0115123> PMID: 25617767
9. Rodríguez O., Jesús A., Jaramillo O. A., and Martínez M., "Wind Power Error Estimation in Resource Assessments," *PloS one*, vol. 10, p. e0124830, 2015. <https://doi.org/10.1371/journal.pone.0124830> PMID: 26000444
10. Aagreh Y. and Al-Ghazawi A., "Feasibility of utilizing renewable energy systems for a small hotel in Ajloun city, Jordan," *Applied Energy*, vol. 103, pp. 25–31, 3// 2013.
11. Bekele G. and Tadesse G., "Feasibility study of small Hydro/PV/Wind hybrid system for off-grid rural electrification in Ethiopia," *Applied Energy*, vol. 97, pp. 5–15, 9// 2012.
12. Islam M. R., Saidur R., and Rahim N. A., "Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function," *Energy*, vol. 36, pp. 985–992, 2// 2011.
13. Mohammadi K., Alavi O., Mostafaeipour A., Goudarzi N., and Jailivand M., "Assessing different parameters estimation methods of Weibull distribution to compute wind power density," *Energy Conversion and Management*, vol. 108, pp. 322–335, 1/15/ 2016.
14. Bozkurt Ö. Ö., Biricik G., and Tayşi Z. C., "Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market," *PloS one*, vol. 12, p. e0175915, 2017. <https://doi.org/10.1371/journal.pone.0175915> PMID: 28426739
15. Cao J., Cui H., Shi H., and Jiao L., "Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce," *PloS one*, vol. 11, p. e0157551, 2016. <https://doi.org/10.1371/journal.pone.0157551> PMID: 27304987
16. Kahraman H. T., Ayaz M., Colak I., and Bayindir R., "Applying the Meta-heuristic Prediction Algorithm for Modeling Power Density in Wind Power Plant," in 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016, pp. 447–451.
17. Shamshirband S., Petković D., Hashim R., and Motamedi S., "Adaptive neuro-fuzzy methodology for noise assessment of wind turbine," *PloS one*, vol. 9, p. e103414, 2014. <https://doi.org/10.1371/journal.pone.0103414> PMID: 25075621
18. Sun C., Hu X., Moura S. J., and Sun F., "Velocity Predictors for Predictive Energy Management in Hybrid Electric Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 1197–1204, 2015.
19. Sun C., Sun F., and Moura S. J., "Nonlinear predictive energy management of residential buildings with photovoltaics & batteries," *Journal of Power Sources*, vol. 325, pp. 723–731, 2016/09/01/ 2016.
20. Wang Z., Ma J., and Zhang L., "State-of-Health Estimation for Lithium-Ion Batteries Based on the Multi-Island Genetic Algorithm and the Gaussian Process Regression," *IEEE Access*, vol. 5, pp. 21286–21295, 2017.
21. Okumus I. and Dinler A., "Current status of wind energy forecasting and a hybrid method for hourly predictions," *Energy Conversion and Management*, vol. 123, pp. 362–371, 2016.
22. Peng H., Liu F., and Yang X., "A hybrid strategy of short term wind power prediction," *Renewable Energy*, vol. 50, pp. 590–595, 2// 2013.
23. Pousinho H. M. I., Mendes V. M. F., and Catalão J. P. S., "A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal," *Energy Conversion and Management*, vol. 52, pp. 397–402, 1// 2011.

24. Gnana Sheela K. and Deepa S. N., "Neural network based hybrid computing model for wind speed prediction," *Neurocomputing*, vol. 122, pp. 425–429, 12/25/ 2013.
25. Ramasamy P., Chandel S. S., and Yadav A. K., "Wind speed prediction in the mountainous region of India using an artificial neural network model," *Renewable Energy*, vol. 80, pp. 338–347, 8// 2015.
26. Haque A. U., Mandal P., Nehrir H. M., Bhuiya A., and Baker R., "A hybrid intelligent framework for wind power forecasting engine," in *Electrical Power and Energy Conference (EPEC), 2014 IEEE*, 2014, pp. 184–189.
27. Mohammadi K., Shamshirband S., Yee P. L., Petković D., Zamani M., and Ch S., "Predicting the wind power density based upon extreme learning machine," *Energy*, vol. 86, pp. 232–239, 6/15/ 2015.
28. Shamshirband S., Keivani A., Mohammadi K., Lee M., Hamid S. H. A., and Petkovic D., "Assessing the proficiency of adaptive neuro-fuzzy system to estimate wind power density: Case study of Aligoodarz, Iran," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 429–435, 6// 2016.
29. Haque A. U., Nehrir M. H., and Mandal P., "A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting," *IEEE Transactions on Power Systems*, vol. 29, pp. 1663–1672, 2014.
30. Jung J. and Broadwater R. P., "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762–777, 2014/03/01/ 2014.
31. Justus C. G., "Winds and wind system performance," Research supported by the National Science Foundation and Energy Research and Development Administration. Philadelphia, Pa., Franklin Institute Press, 1978. 120 p., vol. 1, 1978.
32. Ramadan H. S., "Wind energy farm sizing and resource assessment for optimal energy yield in Sinai Peninsula, Egypt," *Journal of Cleaner Production*, vol. 161, pp. 1283–1293, 2017/09/10/ 2017.
33. Mohammadi K. and Mostafaeipour A., "Using different methods for comprehensive study of wind turbine utilization in Zarrineh, Iran," *Energy Conversion and Management*, vol. 65, pp. 463–470, 1// 2013.
34. Olatomiwa L., Mekhilef S., and Ohunakin O. S., "Hybrid renewable power supply for rural health clinics (RHC) in six geo-political zones of Nigeria," *Sustainable Energy Technologies and Assessments*, vol. 13, pp. 1–12, 2// 2016.
35. Jowder F. A., "Wind power analysis and site matching of wind turbine generators in Kingdom of Bahrain," *Applied Energy*, vol. 86, pp. 538–545, 2009.
36. Ohunakin O., Adaramola M., and Oyewola O., "Wind energy evaluation for electricity generation using WECS in seven selected locations in Nigeria," *Applied Energy*, vol. 88, pp. 3197–3206, 2011.
37. Akdağ S. A. and Dinler A., "A new method to estimate Weibull parameters for wind energy applications," *Energy conversion and management*, vol. 50, pp. 1761–1766, 2009.
38. Jang J.-S., "ANFIS: adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, pp. 665–685, 1993.
39. Thakagi H. and Sugeno M., "Derivation of control rules from human operator's control action," *Proc of the IFAC Sysmp. on Fuzzy Information, Knowledge Representation and Decision Analysis*, pp. 55–60, 1983.
40. Eberhart R. C. and Kennedy J., "A new optimizer using particle swarm theory," in *Proceedings of the sixth international symposium on micro machine and human science*, 1995, pp. 39–43.
41. Bashir Z. and El-Hawary M., "Applying wavelets to short-term load forecasting using PSO-based neural networks," *IEEE transactions on power systems*, vol. 24, pp. 20–27, 2009.
42. Shoorehdeli M. A., Teshnehlab M., Sedigh A. K., and Khanesar M. A., "Identification using ANFIS with intelligent hybrid stable learning algorithm approaches and stability analysis of training methods," *Applied Soft Computing*, vol. 9, pp. 833–850, 2009.
43. Afifi F., Anuar N. B., Shamshirband S., and Choo K.-K. R., "DyHAP: Dynamic Hybrid ANFIS-PSO Approach for Predicting Mobile Malware," *PloS one*, vol. 11, p. e0162627, 2016. <https://doi.org/10.1371/journal.pone.0162627> PMID: 27611312
44. Juang C.-F., *Combination of Particle Swarm and Ant Colony Optimization Algorithms for Fuzzy Systems Design*: INTECH Open Access Publisher, 2010.
45. Mitchell M., *An introduction to genetic algorithms*: MIT press, 1998.
46. I. FIRDAUS AFIFI MD, "BIO-INSPIRED COMPUTING PARADIGMS IN PREDICTING MOBILE MALWARE," M.Eng.Sc, FSKTM, UNIVERSITY OF MALAYA, 2017.
47. Storn R. and Price K., "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, pp. 341–359, 1997.
48. Hegerty B., Hung C., & Kasprak K., "A Comparative Study on Differential Evolution and Genetic Algorithms for Some Combinatorial Problems," presented at the Mexican International Conference on Artificial Intelligence, 2009.