INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENERGIAS SUSTENTÁVEIS



AVALIAÇÃO DE APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS EM MÉTODOS DE MEDIÇÃO-CORRELAÇÃO-PREDIÇÃO

SARAN VIKASH VELUMANI

fevereiro de 2019

POLITÉCNICO
DO PORTO

Instituto Superior de Engenharia do Porto

EVALUATION OF ARTIFICIAL NEURAL NETWORKS APPLICATION TO MEASURE-CORRELATE-PREDICT METHODOLOGIES

Saran Vikash Velumani
1162156

2019

Instituto Superior de Engenharia do Porto
Departamento de Engenharia Mecânica

POLITÉCNICO
DO PORTO

EVALUATION OF ARTIFICIAL NEURAL NETWORKS APPLICATION TO MEASURE-CORRELATE-PREDICT METHODOLOGIES

Saran Vikash Velumani
1162156

Dissertação apresentada ao Instituto Superior de Engenharia do Porto para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Mecânica, realizada sob a orientação da Doutora Rosa Maria Barbosa Rodrigues Pilão.

2019
Instituto Superior de Engenharia do Porto
Departamento de Engenharia Mecânica

JÚRI

Presidente
Doutora Nídia de Sá Caetano
Professor Coordenador, ISEP

Orientador
Doutora Rosa Maria Barbosa Rodrigues Pilão
Professor Adjunto, ISEP

Co-orientador
Mestre Agostinho Paulo Ferreira Pinto
CEO, MEGAJOULE

Arguente
Professor Cláudio Domingos Martins Monteiro
Professor Auxiliar, FEUP

ACKNOWLEDGEMENTS

First, I would like to thank MEGAJOULE for giving me this great opportunity to pursue my project, through which I have gained significant experience and knowledge by working alongside experienced personnel.

It has all started in my course Co-Ordinator Nídia de Sá Caetano's office and I would like to thank her ideas, guidance and motivation that glide me through the tough situations. I would like to thank my thesis advisor Rosa Maria Barbosa Rodrigues Pilão at Instituto Superior de Engenharia do Porto. Whenever I ran into trouble spot or had a question about my research or writing, she was always there to help me out. She consistently allowed this paper to be my own work but steered me in the right direction whenever she thinks it is necessary.

I am very fortunate to get an opportunity to do an internship at MEGAJOULE. My guide Engº Paulo Ferreira Pinto is also the CEO of the company. Despite his busy schedule, whenever I knock the door, he always welcomes me with a big smile in his face and he is always open to discuss the topic to clarify my doubts. I would like to thank Engº Carlos Pinto, Engº Carlos Silva Santos, Engº Carlos Leal Alves, Engº Catarina da Silva Novais and all my colleagues at MEGAJOULE, you have helped me sail through this turbulence. I would like to give a special mention about Engº Carlos Pinto here, he has got a great helping tendency, which I appreciate a lot. All in all, you guys are a great team to work with!

I would like to thank Professor Jorge Manuel Fernandes dos Santos from Mathematical Department at Instituto Superior de Engenharia do Porto. Professor Santos helped me to understand the basics of this 'Blackbox' Neural network model and their basic functionality.

Finally, I must express my very profound gratitude to my parents, my brother, my sister in law and my cousin and my friends for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

KEYWORDS

Wind resource assessment, MCP, Artificial Neural Network, Short term correlation, Wind speed prediction.

ABSTRACT

In this study a single artificial neural network (ANN) model was developed to predict the short term mean hourly wind speed and wind direction at target sites using short term mean hourly reference wind data. Standard multi-layered, feed-forward, back-propagation neural networks with single hidden layer architecture was designed using neural network toolbox for MATLAB. The hidden layers and output layer of the network consist of tangent sigmoid transfer function (tansig) and linear transfer function (purelin) as an activation function. Five different sites from Japan, Saudi Arabia, Jordan, France and Russia with different terrain complexity, completely different weather conditions, and different correlation coefficient between reference and target sites were tested.

Single model was constructed, and two different approaches were experimented. Approach 1 made use of entire concurrent period dataset, the output values from the model was compared against the three methods: regression, matrix and neural network. Second approach was built on certain period of data and tested on unused data. The purpose behind the fabrication of this approach is to try and understand the neural network model.

The results of approach 1 was that the neural network model is able to statistically perform better than other methods and equally well in predicting wind direction sectors. The maximum mean absolute percentage error for NN MATLAB model was found to be 62.5% in Japan to 23.7% in France. The model suffers in predicting the lower wind speeds

which explains the distortion in wind frequency distribution and resulting in Power density deviation. The maximum deviation was -18.1% in Jordan and -7.9% in France.
The sites in Japan, Saudi Arabia, France and Russia were considered for approach 2. The results were interesting, in case of japan the first month was better than the last month result. Overall the performance of the model was better in case of France followed by Russia site. The maximum deviation of Power density was noticed in case of Japan's last month scenario -26.6% to minimum of about 3.2% in France and -5.2% was observed in case of Russia. In Saudi Arabia site, the only case where the concurrent period extends over a period of one year, the performance of the model was statistically good but suffers from same problem of previous cases. The deviation in power density was spotted around -21.4%.

Palavras-Chave

Avaliação de recursos eólicos, MCP, Artificial Neural Network , Correlação de curto prazo, Previsão de velocidade do vento

RESUMO

Neste estudo, foi desenvolvido um modelo de rede neural artificial (RNA) para prever a velocidade média do vento de curto prazo e a direção do vento em locais-alvo, usando dados de vento de referência de curto prazo. Foram projetadas redes neurais padrão multi-camadas, feed-forward, de propagação reversa com arquitetura de camada oculta única usando a caixa de ferramentas de rede neural para o MATLAB. As camadas ocultas e a camada de saída da rede consistem na função de transferência sigmóide tangente (tansig) e na função de transferência linear (purelin) como uma função de ativação. Foram testados cinco locais diferentes, Japão, Arábia Saudita, Jordânia, França e Rússia, com diferentes complexidades de terreno, condições climáticas completamente diferentes e diferentes coeficientes de correlação entre os locais de referência e os de destino.
Foram testadas duas abordagens diferentes com o modelo construído. Na abordagem foi usado todo o conjunto de dados do período concorrente e os valores de saída do modelo foram comparados com três métodos em estudo: regressão, matriz e rede

neural. A segunda abordagem foi construída usando apenas um determinado período de dados e o modelo foi testado em dados não utilizados. O objetivo desta segunda abordagem foi tentar entender o modelo de rede neural.

Os resultados obtidos com a abordagem 1 aplicada aos 5 sítios em estudo permitiram verificar que o modelo de rede neural desenvolvido se apresenta estatisticamente melhor do que os outros métodos testados. Verifica-se que é capaz de prever bem a direção do vento por setores. Foi obtido um erro percentual médio absoluto máximo com o modelo NN MATLAB de 62,5% no Japão e de 23,7% na França. O modelo desenvolvido apresenta uma limitação na previsão das velocidades de vento mais baixas, o que explica a distorção na distribuição da frequência do vento e resulta no desvio da densidade de potência. O desvio máximo obtido para a densidade de potência foi de -18,1% na Jordânia e de -7,9% na França.

Na abordagem 2 foram utilizados os dados do Japão, Arábia Saudita, França e Rússia. Os resultados foram interessantes. Verificou-se que no caso do Japão foi possível obter melhores resultados para o primeiro mês do que para o último mês. No geral, o desempenho do modelo foi melhor no caso da França, seguido pela Rússia. O desvio máximo da densidade de potência foi observado no caso do cenário do último mês do Japão -26,6% e foram observados desvios mínimos de cerca de 3,2% na França e -5,2% na Rússia. No site da Arábia Saudita, o único caso em que o período concorrente se estende por um período de um ano, o desempenho do modelo foi estatisticamente bom, verificando-se a mesma dificuldade de previsão de velocidades baixas. O desvio na densidade de potência foi de cerca de -21,4%.

LIST OF ABBREVIATIONS AND SYMBOLS

List of Abbreviations

AEP	Annual Energy Productions
ANN	Artificial Neural Network
AR	Auto Regressive
CFD	Computational Fluid Dyanamics
FFNN	Feed Forward Neural Network
IoA	Index of Agreement
LM	Levenberg Marquardt
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MATLAB	Matrix laboratory
MPLS	Multiple Principal Least Square
MS	Microsoft
MSE	Mean Square Error
NN	Neural Network
R	Correlation coefficient
RMSE	Root Mean Square Error
TWh	Tera Watt Hour(s)
VLT	Virtual Long Term
VRM	Variance Ratio Method
WAsP	*Wind* Atlas Analysis and Application Program

List of symbols

A	Swept Area / Weibull scale parameter
a_i	Actual data
k	Weibull shape parameter
e	Residual value
G	Sensitivity coefficient
P	Power
ρ	Density
P_i	Predicted value
u	Fitted mean wind speed
v	Velocity
W	Watts
z	Surface roughness value
θ	Wind veer
ΔX_i	Measuremnt error
σ_i	Uncertainty

INDEX OF FIGURES

INDEX OF TABLES

INDEX

INTRODUCTION

1 INTRODUCTION

Wind energy is clean, renewable, and one of the fastest growing alternative energy sources. One of the key objectives of 7th Environmental Action Programme is to "turn the European Union into a resource-efficient, green, and competitive low-carbon economy", also one of the horizontal priorities is to "help European union address the international environmental and climate challenges more effectively" by 2020 [1]. Due to these action plans the renewable energy sector is growing at a blistering rate also the rapid rate of fossil fuel depletion is a vital reason. The data from Portuguese association of renewable energy (APREN) reveals that as far as power consumption in 2017 is considered, 44% of total power requirement was satisfied by renewable sources, in which the wind energy sector leads other renewable competitors by producing 11.9 TWh while hydropower produces 7.3 TWh [2]. In global scale, the rise in renewables was fueled by unprecedented growth in China and United States of America, accounting for around 50% rise in power production. Electricity from renewable energy sources in European Union (EU) rose by 8% and by 6% in India and Japan respectively.

With increasing in wind power installations, it is very important to forecast the wind energy potential which eventually helps the grid management. Wind energy potential completely depends on wind speed at the target site. The assessment of wind resource is imperative, but the stochastic nature of wind makes it a perplexing criterion to assess. The veracity of the wind energy potential depends on number of factors and uncertainties. To encounter this problem, researchers have proposed the Measure-Correlate-Predict (MCP) methodology. This method practices short term wind data from target site and correlating it with concurrent period data from reference site arriving at a relation and with that relation extrapolating it for the long-term conditions inclusive of all uncertainties involved in the due course.

1.1 Context

The present work was carried out from an internship in partnership with the Megajoule. Megajoule is a private Portuguese company dedicated to energy consulting Renewable Energy, is a leader in the evaluation of wind resources in Portugal, one of the world's top wind energy markets. It presents a high range of services related to the evaluation of wind resources, namely the validation of sites for project implementation, studies of evaluation of wind potential, planning and conducting campaigns to measure the characteristics of the wind in a site and conducting audits and projects. In addition, services related to other renewable energy solar resource and biomass also provided. The scope of this work is related to examine and analyze the application of Artificial neural networks (ANN) in MCP methodologies. In the due course different methodologies from industrial software were also studied and tested and results of industrial model were compared against the designed model. This comparison has to take into account several factors such as the correlation coefficient between datasets, considered in the accomplishment of the studies. The relevance of this study to the company lies in the possibility of implementing the ANNs as one of the methods in MCP.

,

1.2 Objective

The aim of the research is to explore the possibility of implementing ANNs as one of the methods in MCP methodologies. Although there are lot of research papers on ANNs in similar applications but only few of them considering wind directions, the focus of this work was to build a single model that utilizes both the wind speed and wind direction and provides information about the same, as the methods in commercial software's does the same. The yielded result from the developed model in neural network is compared against the methods commonly used in wind industry software. To analyse wind speed and wind directions statistically and graphically and calibrate the power density estimate of different models (particularly designed model is compared against others).

1.3 Structure of the dissertation

This section presents the structure of the report, consisting of the following chapters:

The first chapter is an introductory chapter which talks about the contextualisation, objective and motivation of this research work.

The second chapter discusses about the wind resource assessment methodology which consists of different procedures and in-depth definition of first part of resource assessment, introduction to MCP and research paper associated with it.

Third chapter consists of the measure correlate predict methodologies, with three different methodologies used and introduction to ANN and state of art on ANN in similar applications also design of the NN model.

The fourth chapter talks about the site description, data availability and results and discussion.

The fifth chapter discusses about the conclusion and future work.

WIND RESOURCE ASSESSMENT METHODOLOGY

2 WIND RESOURCE ASSESSMENT METHODOLOGY

The characterization of the wind regime and evaluation of the wind potential of a given location follows a methodology with several stages. In figure 1 is a schematic representation of the different steps of this methodology.

One of these steps involves extrapolating the wind data from a local measurement campaign that is necessarily short (usually 1 year) for the long term defined as the average lifetime of a wind farm (typically 20 years). The relevance of this work is precisely at this stage where different correlation algorithms can be used between local measurement data and long-term reference data. An artificial neural network was parameterized, and its performance compared to different alternative MCP methodologies.

In the following sections the initial steps of this methodology, until the extrapolation phase of the local measurements for the long term, are detailed.

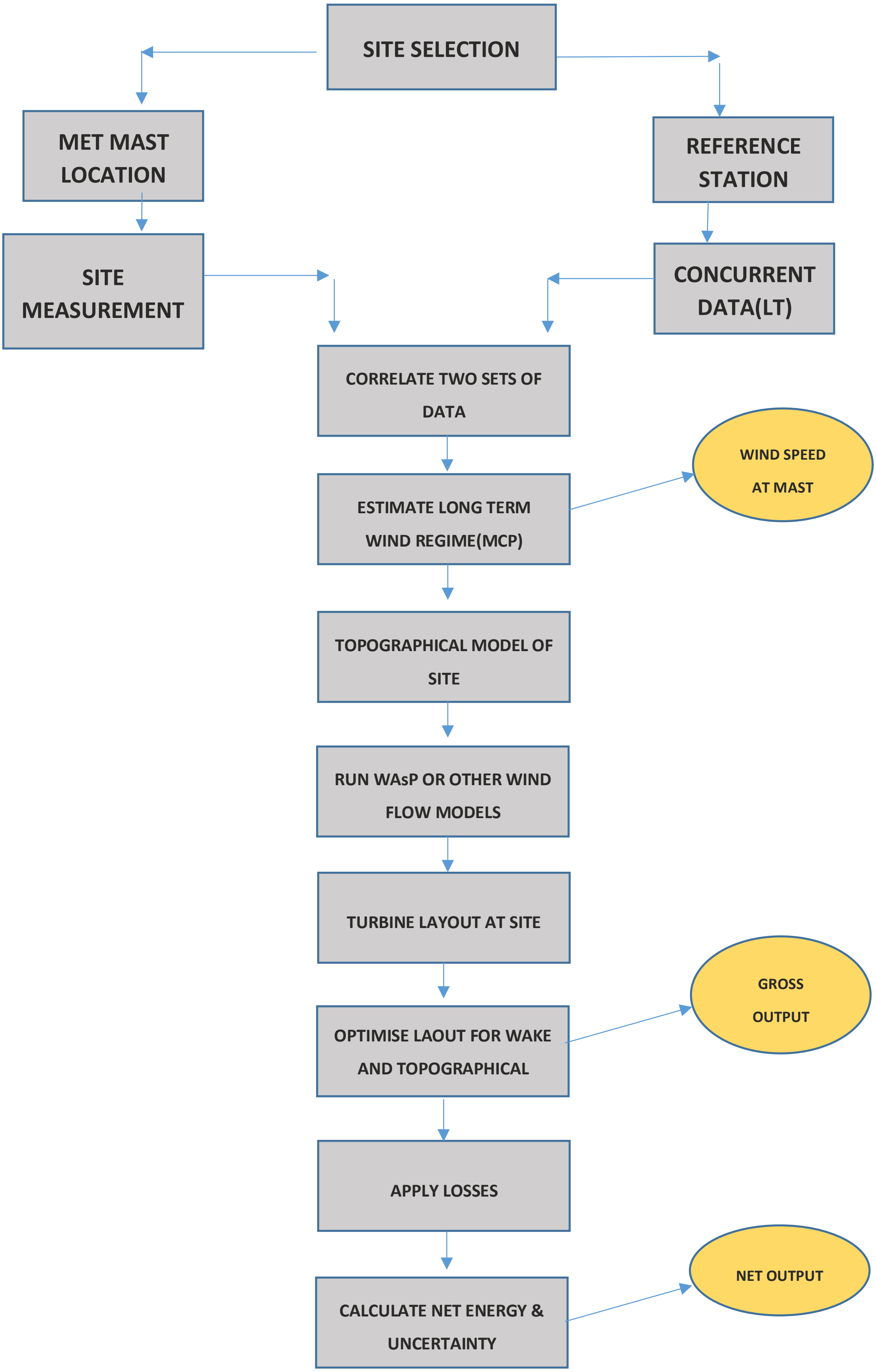


Figure 1 – An overview of wind energy production estimation process [3]

2.1 Identification and quality evaluation of the site

The term folklore is coined as oral history which is used as one of the methods in estimating the wind resource. The knowledge of the people about characteristics of wind is used as the base source which tends to have deviations and shortcomings but for the initial stage of assessment it is compromised [4]. As the natives of the place are living there for years, therefore it is taken as for granted that their knowledge of the surroundings and understanding of the local weather is good enough for the preliminary gathering of the data. This method is advised to be used only applicable when prospect of using all other method is not possible and critical verdicts cannot be made with this method [4]. By observing the vegetation and the biological indicators the wind climate of the place can also be assessed. High wind area is also located using a method called Biological Wind Prospecting. Trees have been used for hundreds of years as ecological indicators of wind direction, identification of location with strong continuous wind using wind deformed trees [5].

Site identification is the first stage in wind resource assessment to identify possible wind farm locations in whole. The location is further refined by considering reports from wind resource maps and publicly available data. Site identification provides information about feasible and non-feasible conditions which includes constructional possibilities, access to the location and ecological constraints. Based on these information's geographic information system (GIS) is established which plays a vital role in future decision-making process. Some of the pragmatic geographic data includes,

- Wind resource maps
- Constructions
- Topographical data
- Pipelines (oil and gas)
- Water bodies
- Licensing requirements
- Electrical transmission lines
- Airspace restrictions
- Highways/railways.

With help of all these geographical information's, appropriate candidate site is selected and can be applied to different spots in same location to have more candidate sites. As soon as geographic information system is designed, the competent location to plant measuring mast tower can be determined [6].

Territorial wind resource maps are effective starting point for identifying potential wind resources. With the help of mesoscale weather models, microscale wind flow models refined wind resource map is produced. Wind resource assessment investigation can be handled in a macrolevel involving vast regions over several thousand square kilometers. These mesoscale models can provide overall idea about high potential areas. Weather Research and Forecasting (WRF) is a numerical weather prediction system which can be used to build mesoscale model.

The proper location of measuring mast in intended site is of high importance as it will reduce uncertainty of wind resource at potential turbine locations. The mast should be in a similar or representativeness of topography to final target areas. The location of mast should be able to capture all possible seasonal variations and conditions that will be experienced by the turbines. The need for multiple masts is advised in case of complex sites.

As there is no international standard to follow, based on practices some assumptions were made regarding contour of the location. If the contour is flat with uniform surface roughness the site can be classified as simple. In this case maximum recommended distance between proposed turbine location and mast is around 5-8 kilometers. If the contour is subjected to series of low hills, single ridge perpendicular to predominant wind is classified as moderately complex and in this case the distance recommended is around 3-5 kilometers. If the contour possesses geometrically complex ridgelines, location with heavy forest cover then they are termed as very complex sites and recommended distance is 1-3 kilometers. These standards could be followed to reduce uncertainties in wind flow modelling.

Buildings, forests, rock outcroppings are considered as obstructions or disturbance that can negatively affect analysis of wind characteristics. The area of increased turbulence can rise to two times the height of obstacle in upward direction. The sensors should not be placed no shorter than 20 times the height of obstruction in the horizontal distance in the predominant direction of wind [7].

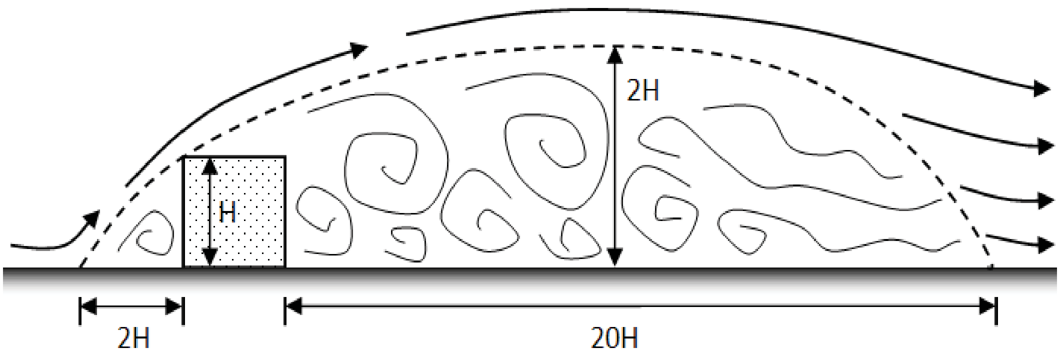


Figure 2 – Obstruction effects on air flow [7]

2.2 Local measurements

The aim of wind resource assessment is to precisely calibrate wind speed, wind direction, and air temperature and barometric pressure as well. Wind speed data is the decisive indicator of sites wind energy resource. The measurements should be made at multiple heights in order to capture all possible wind shear characteristics. The frequency of wind direction is critical in enhancing wind turbine layout in a wind farm. Air temperature measurements help us to compute air density values [6,7].

Wind is a result of differences in atmospheric pressures. The difference in atmospheric pressure is due to uneven surface heating of earth by sun resulting in high and low regimes of pressure results in generation of wind. Wind is the vector quantity possessing wind direction and speed. A proper wind measurement campaign reduces uncertainty in wind resource assessment. For wind speed measurements the basic sensors used are anemometers**.** The most conventional type of anemometer is the cup anemometer for horizontal wind speed survey. The cup always faces the wind, involving typical technique to inculcate aerodynamic structure which converts the pressure into rotational torque. The rotation of cup is proportional to the speed. Some of the other types of anemometers:

- 3-cup anemometer
- Photo electric anemometer
- Propeller anemometer

Vertical anemometers are used in complex sites to capture vertical wind speeds which might introduce turbulence in the location. Anemometers are typically mounted at heights between 30m to 60m. In order to compare data from different locations, all wind speeds should be extrapolated to hub height. Wind speed from one height can be accustomed to hub/required height by using the pattern of power law equation, [8]

$$\frac{V_h}{V_r} = \left(\frac{h_h}{h_r}\right)^{\alpha} \tag{1}$$

Where,

V_h is the velocity (m/s) at height h_h (m)

V_r is the velocity (m/s) at height h_r (m)

α is wind shear component

wind shear is defined as variation of wind over either vertical or horizontal distances. Also, along with power law, log law can also be used which is based on the principal of boundary layer flow that is given [8] by

$$\frac{V(z)}{V(z_r)} = \frac{\ln\left(\frac{Z}{Z_0}\right)}{\ln\left(\frac{Z_r}{Z_0}\right)} \tag{2}$$

Where Z and Z_r are the target and reference heights in meters respectively and V(z) and $V(z_r)$ are the corresponding wind speeds (m/s) and Z_0 is the surface roughness value depending on type of terrain (table 1).

Table 1 – Surface roughness value for different types of terrain [6]

Type of terrain	Roughness value
Snow surface	0.003
Fallow field	0.03
Crops	0.05
Hedges and minor buildings	0.1
Forests and woodlands	0.5
Centre of cities with tall buildings	3.0

The critical factors that should be considered while selecting an anemometer are

- Durability
- Operating ability
- Starting threshold
- Distance constant
- Response to Off-horizontal wind
- Sensor calibration

Wind direction measurements are of primary importance as they are used in modelling the spatial distribution of the wind resource across the proposed zone and for optimising the layout of wind turbines. Wind vanes are usually used in wind direction measurements. To capture the wind direction frequencies precisely it is recommended to have wind vanes installed at two different levels. Wind vanes should not be installed at the same height as that of anemometers as it would affect the wind speed readings. Air temperature is yet another crucial aspect which defines the wind farm's operating situation. The measurements of air temperature are very important because it defines the air density in the intended location.

The Power density available in the wind can be calculated by the following equation,

$$P = \frac{1}{2} \times \rho \times A \times v^3 \qquad (3)$$

Where, P is Power density (W/m^2)

ρ is density of air (kg/m^3)

A is swept area of rotor (m^2)

v^3 is velocity of air (m/s).

The power generated is directly proportional to density of atmospheric air, so it is of prime importance to measure precisely. Temperature readings should be conducted at several meters above ground level to avoid fluctuations due to surface heating.

In-order to study the wind resource and wind shear characteristics of the site, a tall mast of prerequisite height is needed with multiple measurement levels legitimately following the standards. In general towers are of two types tubular and lattice which are available in three different variants tilt up, telescoping and fixed. Tubular structures are used in measurement campaign where the requirement of tower height is less than 60m and lattice towers are used in conditions requiring tower height more than 60m. The height of the towers should be fixed in order to position the sensors (anemometers, vanes, temperature). The basic practice is that height of the tower should be at least 3/4th the hub height to represent all the possible conditions that will be encountered by the turbine. Two to three anemometer heights are required for a tower of 50m to 60m in height. If two anemometers are recommended for the tower, then the height ratio between top and bottom should be around 1.6. A set of temperature sensors in ideal conditions but at different heights should be located near the lower and upper measurement levels without interfering with the wind measurements. The local measurements should be the complete representativeness of the future turbine calibrations. The mast readings are usually measured for one year but there are practices where two consecutive year measurements are also used. The use of two consecutive year data prove to have less uncertainties than its one-year counterpart.

Data loggers can store data locally and most of them can transfer data to another location through cellular phones, radio frequency, satellite link etc. Data loggers should be able to store data values in a sequential format with corresponding time and date stamps and able to work on battery power. The measured parameters in data loggers should be sampled once every one or two seconds and recorded as standard deviations (wind speed and direction), averages (wind speed, direction temperature and pressure), maximum and minimum values (wind speed and temperature). Standard Deviation values are defined for one or two second intervals, averages are defined for ten-minute intervals and maximum and minimum values are defined for one or two second interval. The power source for datalogger systems are provided through AC power units or lead acid batteries can be used, also solar power is also an option to consider. Ground based remote sensing is also a technique which could be complemented with traditional method in complex topographies. SODAR (SOnic Detection And Ranging) and LIDAR (LIght Detection And Ranging) are used actively in wind measurements. SODAR works on acoustic pulses while LIDAR works on laser light. These ground-based devices can capture the wind profile clearly up to the height of 150m. Moreover, the cost of installation and operation is very less compared to towers and are easily relocated to nearby location in the wind farm to check for wind characteristics. Determination of true north is one of the vital process before installing the measuring mast [6].

As the wind passes downstream through the turbine blades reduces in speed, generating wakes which will eventually reduce power produced in the wind farm. In order to experience the minimum wake effects turbines should be spaced between 5 and 9 rotor diameters in the prevailing wind direction and between 3 and 5 rotor diameters in the direction perpendicular to the prevailing wind [6].

Once the wind resource measurements are compiled, they are transferred to data processing section to check for the quality of data and to validate the data. The data is inspected for integrity and representativeness also explicitly to detect invalid values in data record. The quality control is initiated to identify data gaps, sensor degradation, sensor icing, wiring damage, abnormal behavior in data samples and time stamps. Data validation is usually performed with software's like windPRO or self-built spreadsheets. The software's and spreadsheets provide graphical representations of the data where it is easy to spot the data gaps or invalid data. The common measurement parameter checks are performed on the data and they are range tests, relational tests, and trend tests. It is possible to fill in any gaps in the record with valid data from other sensors that are in similar conditions. A wind data measurement campaign for a candidate site is normally undertaken during the year prior to the proposed installation.

2.3 Long term evaluation

The next step in characterizing wind resource at target site is to relate the procured wind data with long term historical or metrological data. In order to accomplish better characterization of long-term wind conditions and to reduce the uncertainty in wind farm energy forecasts it is always appropriate to relate the short-term target site data with long term metrological or re-analysis data. There are several different Reanalysis data available but the most frequently used are NCAR (National Centre for Atmospheric Research), NCEP (National Centre for Environmental Prediction), MERRA (Modern Era Retrospective-analysis for Research Applications) and CSFR (Climate Forecast System Reanalysis). If such long-term data are not available, it is also possible to simulate a virtual long-term wind data (V LT). V LT is a time series of wind speed and wind direction along with some other metrological parameters as well.

2.3.1 Measure – Correlate – Predict (MCP)

Wind is highly variable on short and long-time scales, from minute to minute and year to year. Current practices require at least one year of measured wind data at the site of interest to yield a reliable resource assessment. Monitoring, however, are expensive and funding decisions may occur over longer time scales. Additionally, a single year's wind monitoring campaign cannot capture year to year variation in wind characteristics. To address these issues, researchers have developed the measure-correlate-predict (MCP) methodology. This method uses site-specific short-term monitoring data combined with correlated long-term historical data from a neighboring monitoring station to estimate the long-term resource at the candidate site. The assessment of the long-term wind resource at a potential site based on a relatively short-term on-site measurement campaign is a crucial task in the progress of a commercial wind turbine installation. The classic approach is based on the measure-correlate-predict (MCP) method where an analogous model between the target site wind data and the data obtained from a suitable reference site for about 10 to 20 years is built from concurrent records (figure 3). Thus, built MCP model enables us to predict the long-term wind characteristics at target site. MCP methods tend to afford more accuracy than physical modelling methods [10]. It is usually very complicated to establish a relationship between two sites as the wind speed and direction tends to vary with time and distance, the contour of the terrain, large scale and small-scale weather patterns, obstacles (trees, buildings etc.) and atmospheric stability. The role of MCP in wind resource assessment is formidable [11], as most of the wind analysts who have carried out measurements to estimate wind characteristics have used some means of MCP method to predict long term wind regime. The MCP technique combined with flow models such as WAsP or CFD models used in wind industry in enhancing the estimation of long-term power production of the wind farm.

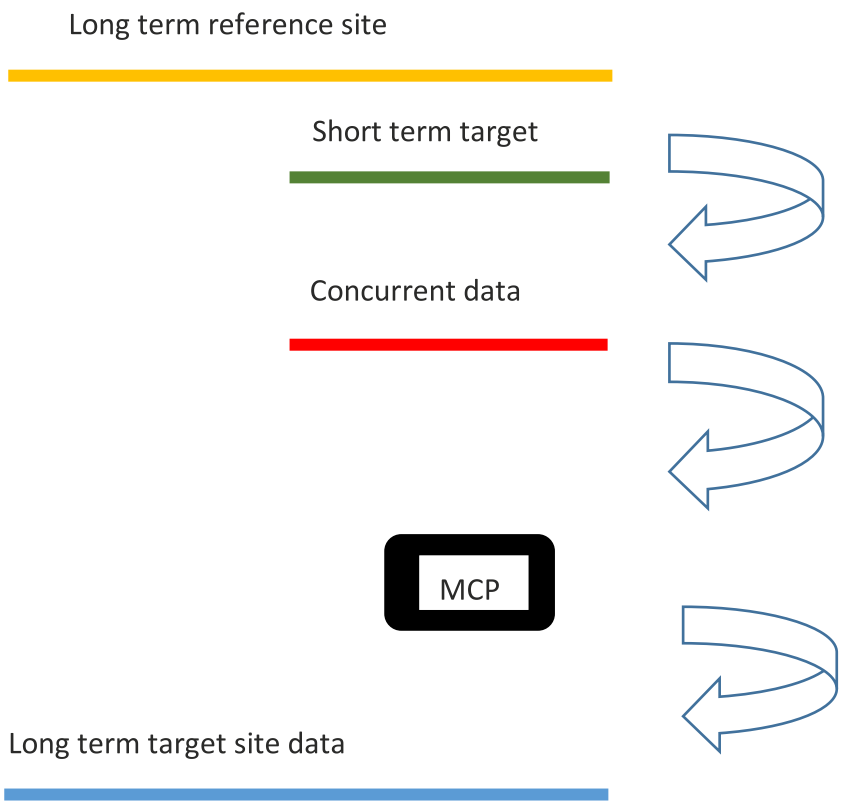


Figure 3 – Pictorial representation of MCP procedure

MCP is the general term which encompasses numerous methods starting with simple linear regression, polynomial regression, methods of ratios, matrix method, Weibull method, wind index MCP, variance ratio method, multiple principal least square method, bivariate Weibull approach, support vector method, artificial neural network, Monte Carlo wind speed simulation. These methods perform independently from each other and have their own advantages and limitations.

There are lot researches happening in MCP methods and lot of nuances introduced lately. The degree of correlation between the short-term target site data and concurrent reference data is directly proportional to the quality of prediction or forecasting the long-term data but according to a method called Multiple Principal Least Square (MPLS) [11] has an advantage of using in low correlation between target and reference site data. In his paper author compared MPLS with Variance Ratio method and the results proved that MPLS performs much better than variance ratio.

Bivariate Weibull approach was a new approach [12], as it is assumed generally wind speed follows Weibull distribution author thinks that this approach would have stronger theoretical background than usually used regression methods. Regression method was compared against bivariate Weibull approach using a twelve-month measurement period and results were not promising. The conclusion was that while using short measurement periods regression method is more appropriate than bivariate Weibull approach.

The fundamental difference between wide range of MCP methods lie basically in the type of relationship established between wind data recorded in target site and wind data recorded concurrently at one or more nearby reference stations with the history of long-term data [13]. In this review author analyses wide range of MCP methods including methods from early 1940’s to methods in current automatic learning techniques, the report covers over 150 international journals on wind energy. The general description of linear, non-linear and probabilistic approach was given along with the limitations of using MCP methods. The methods working with multiple reference stations was also coined along with single reference stations. The author points out some of the requirements to be followed by reference station measurement procedure and most important is that the concurrent period over which wind data are available for reference station and target station should be able to cover all possible seasonal variations.

In order to compare the MCP algorithms, Rogers et al. [14] recommends using some performance metrics like correct mean wind speed, correct wind speed distribution, correct annual energy production at the target site assuming sample power curve of turbine. The different MCP methods proposed in various literatures which includes linear regression model, model utilizing distributions of ratios of wind speeds, vector regression method and finally method based on ratio of standard deviation of two data have been taken under study, the performance of these methods analyzed using a common set of data. By comparing the conclusion that regression model tends to provide biased estimations when compared with method of ratios and distribution of ratios.

The attempt to combine different MCP algorithms [15] to ensure that they could catch variations in wind speed as well as wind directions.so called ‘hybrid’ method uses data from multiple reference stations and the wind data was branched into different sectors according to wind direction. The hybrid approach was tested on four MCP methodologies linear regression, variance ratio, artificial neural network, support vector machine. The author uses three sets of performance metrics to evaluate his 'hybrid' model. The first set of metrics used to analyze statistical parameters like mean wind speed, wind speed variance, RMSE (Root Mean Square Error) and MAE (Mean Absolute Error), second set of metrics involve evaluation of distribution of long-term wind speed where Weibull and multivariate and multimodal wind distributions were used, and final set of metrics analyzed the wind farm performances. The conclusion was that best hybrid MCP algorithm was greatly determined by length of correlation period and hybrid strategy using multiple reference stations can predict the long-term wind speed distributions more accurately.

Three new MCP models namely Weibull regression, Simple linear regression with probability density function, Weibull regression with probability density function for long term prediction of wind speed along with two already existing methods (simple linear regression, variance ratio) were suggested [16]. The performance metrics like mean, standard deviation, Weibull scale and shape factor and energy density were analyzed. The conclusion was that Weibull regression using probability density function provides accurate prediction for all relevant metrics in consideration.

A new matrix method for predicting long term wind roses with MCP was suggested by Woods [17]. The measured wind speed at target site and concurrent reference speed are split into groups according to the value of direction sector measurement (12 sectors used). The final number count matrix for one year's hourly data was tabulated. This method reveals the fact that assumption made in linear regression MCP that sector distribution at reference site gives good representation of measurement site fails. The author concludes that matrix method is superior to traditional MCP methods because the representation of angular data makes it particularly applicable to prediction of wind roses in complex sites.

MEASURE-CORRELATE-PREDICT METHODS

3 MEASURE-CORRELATE-PREDICT METHODS

3.1 Linear regression method

The regression method used is an improved version over a classic linear regression analysis in which distribution of residuals are also taken under considerations. In specific cases it is possible to use polynomial regression also, eventually demanding complexity of the case.

The general linear regression equation is

$$Y = f(x) + e \tag{4}$$

Where Y is the dependant variable (wind speed at target site (m/s))

x is the independent variable (wind speed at reference site (m/s))

f(x) is the function of regression model

e is residual value.

There is a special case involved in linear regression where fit is forced through origin, in that case the fit is poorer than one that arrive normally, for which the regression equation will be

$$Y = f(x) \tag{5}$$

The improved wind distribution curve is obtained if residuals were included in regression, also if residuals are not considered there is a chance of 10% error in wind energy prediction. The regression model is a combination of wind speeds and wind veer where, the wind veer is the difference between wind direction at the measurement site and the reference site [18].

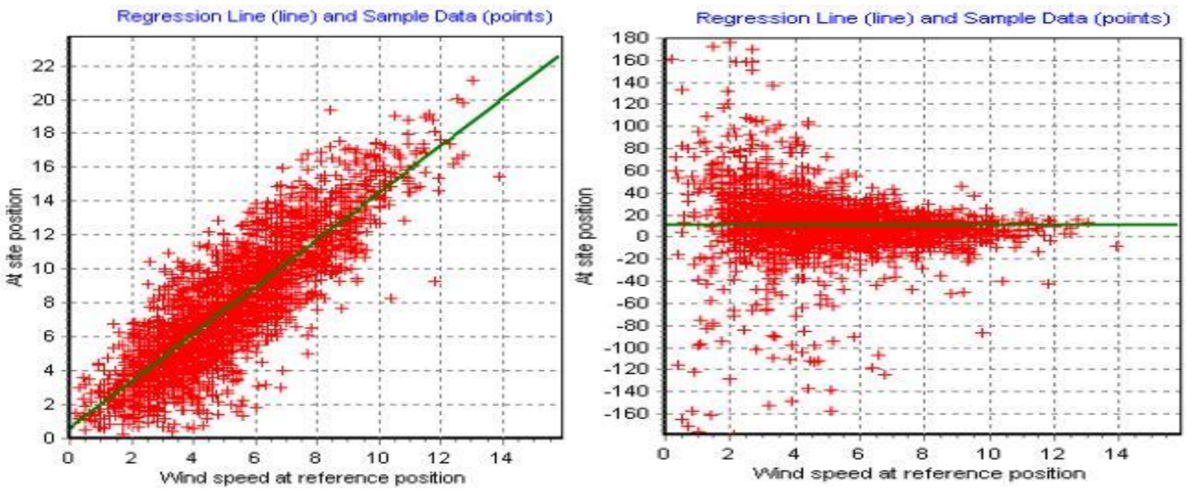


Figure 4 – Wind speed fit (left) and wind veer (right) [18]

3.2 Matrix method

The Matrix method utilizes the binning technique by combining wind speed and wind direction. The relation between target and reference wind speed is calculated individually on each bin for each sector which helps in simulation of wind speed characteristics in target site.

The concurrent period of wind data is used to calculate the set of transfer functions which will be used to estimate wind speed and wind directions at target site from reference site. Matrix method tries to capture the variations in wind speed and wind directions with the help of joint distribution.

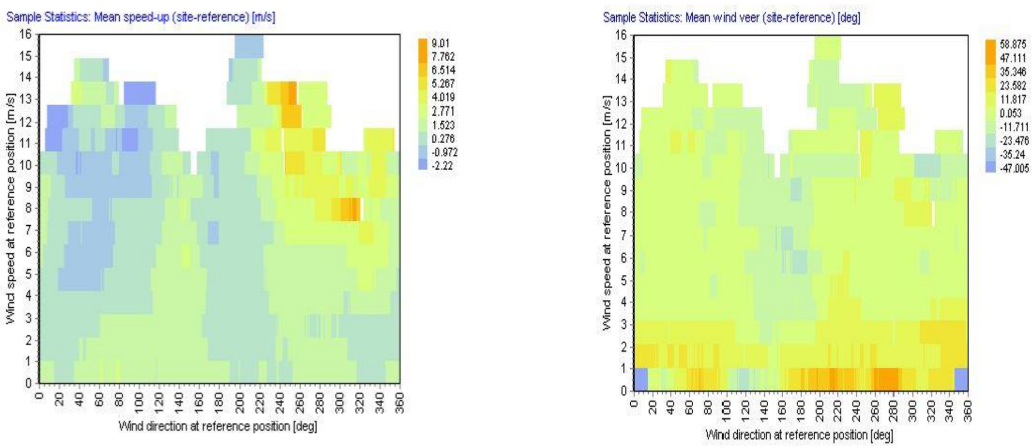


Figure 5 – Mean wind speed up (left) and mean wind veer (right) [18]

The model is based on the joint distribution of the measured wind speed-ups and wind veers. Thus, for each measured sample we must measure pairs of the two variables.

$$\Delta u = u_{site} - u_{ref} \quad (6)$$

$$\Delta\theta = \theta_{site} - \theta_{ref} \quad (7)$$

Where,

Δu is the wind speed-up

u_{site} is the wind speed at the site position

$u_{reference}$ is the wind speed at the reference position

$\Delta\theta$ is the wind veer

θ_{site} is the wind veer at the site position

$\theta_{reference}$ is the wind veer at the reference position.

Wind speed and wind direction at reference site are combinedly used as the function to estimate the wind speed and wind direction at target site with the help of correction factors [18].

The wind speed at target site could be estimated from the mean speed up graph, and to estimate the wind direction mean wind veer graph could be used. The correction factor should be added to corresponding location to achieve the estimations on wind speed and wind direction in reference with the equations.

3.3 Artificial neural network method

An artificial neural network (ANN) is a system consisting of densely interconnected simple processing elements which can perform large simultaneous computations for data processing [19]. The ANNs are motivated by the functioning of biological neuron and follows the similar pattern in modern computing. The first artificial neuron was introduced by Walter Pitts and Warren McCulloch. The standard ANN comprises of an input layer, hidden layer, and the output layer all of them are interconnected by different weights. As a reason of these interconnections, the ANN possesses a powerful computational power to learn from examples and generalize the solution to a wide range of problems. Usually input layer does not perform any calculations and thus for a network the hidden and output layer neurons are counted. Network topology or network architecture is defined as the arrangement of neurons in the hidden and output layers [20]. An example of feed forward network with R inputs and s hidden neurons and s hidden layer is presented (figure 6).

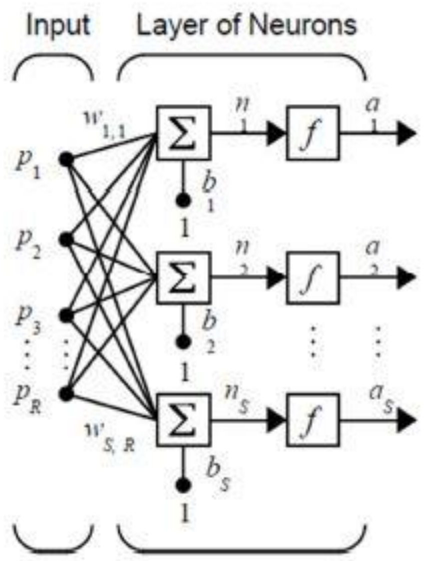


Figure 6 – Feedforward network with R inputs, s hidden neurons and s outputs [20]

The architecture of ANNs can be defined as the arrangement of neurons in the layer(s). Basically, the complexity of the problem necessitates the use of network with more than one layer called as multi-layer neural network [20]. ANNs are modelled to solve wide range of problems, in order to do specific functions several types of architectures have been designed they are Multi layered feed forward network with Back propagation algorithm, Hopfield network, Boltzmann machine.

In the modern world the neural networks exhibit wider range of application possibilities ranging from business, engineering, insurance sectors and in some cases predict the risks of bankruptcy and stock markets as well. some of the interesting applications include auto pilot enhancement, flight path simulation in aerospace industry and weapon control and target tracking in defence sector [21].

The basic process in a neural network involves mapping of random input vector into a corresponding random output vector considering that there is no relationship between two data sets. The mapping process is accomplished by assigning input vectors with connection weights and biases which transmit the information to the next neuron, initially weights are assigned randomly and subsequently stabilised by training the network. The dot product of the vector of weights and the input vectors are summed at the summing junction before they are fed into activation function of the network [22] also the weights are adjusted iteratively by “learning through example” strategy rather than classical “programming” methods.

Neural networks in general can learn and generalize situations to produce valid solutions. Neural networks trained by Back propagation algorithm are most common in science and engineering applications.

The training process aids in achieving optimal set of weights which eventually reduces error by gradient descent methods. The basic principle behind the back-propagation algorithm is that error obtained in prediction from one layer is used as an input to next layer with an idea of minimising the error. Feed forward neural networks are most common type of networks which are widely used due to their advantages such as they are data driven, able to map the inputs to the outputs without need of any assumptions during the designing of model, they are good approximators of dynamic and non-linear problems. Moreover, Feed forward neural networks are easy to implement [23].

The most exciting and crucial phase of the study is the construction of the model. There are number of parameters that have influence on the performance of the model. Some of them are activation function, size of hidden layer, learning algorithm, network weight initialisation and number of iterations.

The transfer/activation function is a constantly increasing function that transforms the sum of weighted signals to generate the final output. some of the activation functions are linear function, sigmoid activation function and the sigmoid family consists of logistic activation function and hyperbolic tangent function (figure 7).

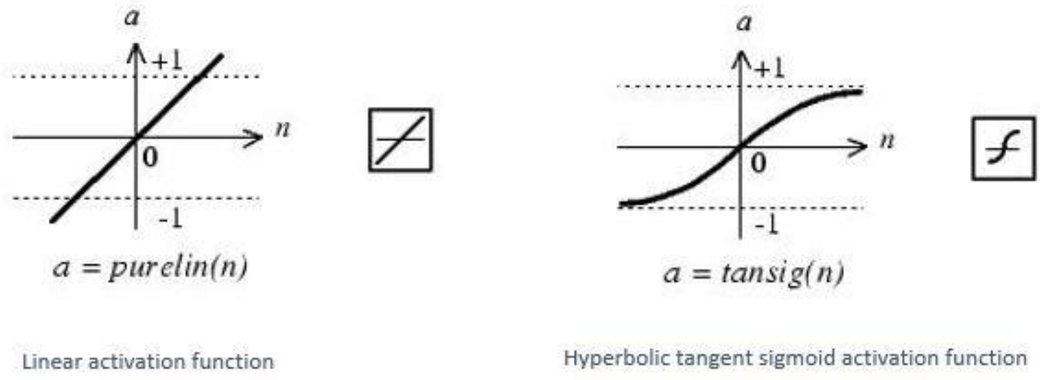


Figure 7 – ANN activation functions

The number of inputs and outputs naturally decide the number of neurons in the input and output layers. The real task is to determine the size of the hidden layer also number of neurons(H) in them, as well. Deciding the size of hidden layer is a challenging part in network modelling, for instance if we consider a small H, then the network might be unable to capture all the hidden information about the data resulting in linear estimate of the output while larger H might be time consuming to train and show good results during training and validation but substantial error during testing phase resulting in overfitting. To get the optimal size of the hidden layer, the best way is to iteratively adjust the size while measuring the error during the neural network testing [19, 24].

A network training function is a general learning function, these functions repeatedly apply a set of input vectors to a network and update the network every time until certain stop criteria are fulfilled. Some of the network training functions are Levenberg Marquart algorithm, Scaled Conjugate Gradient, Bayesian Regularisation algorithm etc. The stop criteria can be of a maximum number of iterations, a minimum gradient of error, etc. [25].

The larger networks can perform well with complex functions, but overfitting the data is the common problem. Overfitting occurs when the network can be very good at modelling the training dataset but perform poorly when a new data set is introduced, this can be experienced by increasing the error in testing data while error values continues to decrease in training dataset. 'Early stopping' technique keeps an eye on validation error and stops the algorithm when validation error increases for multiple iterations [20].

Epochs are set as one of the training parameters and are important in gauging the training time taken by a neural network to reach convergence and to set the goal that determines the extent to which the network should be trained [26].

Several error measurement techniques can be applied to assess the relationship between the two sets of data, e.g. the mean square error (MSE), the sum square error (SSE), the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), coefficient of correlation (R^2) or simply taking the coefficient of correlation (R) between the two sets of data.

3.3.1 State of Art on ANN in similar applications

The possibility of using neural networks was examined [9], to make the predictions of long-term energy yield at target wind farm site. The neural network approach is compared with standard MCP algorithm. The author conducted experiments on seven data sets which are from different wind conditions and locations in all these cases the neural network approach works at least as good as MCP and in some cases substantially better. The conclusion is that amongst all the architecture used multi-layer perceptron's trained using the combination of backpropagation followed by Quasi-Newton training yields better result.

Singh [27], tried to design and develop an ANN based system for predicting wind energy generated by wind turbines. The architecture used by proposed neural network model is a feed forward network supervised with back propagation algorithm. The input parameter for proposed model includes wind speed, relative humidity and generation hours with generated energy by turbine as output parameter. Mean Square Error (MSE) and Mean Absolute Error (MAE) were used as performance metrics to analyse the accuracy. MATLAB platform was used to build the network. The author concludes that this model proves to be an efficient and beneficial tool for estimation of wind energy.

The study has been conducted by Mahamad et al. [28], extensively to make use of ANN in predicting solar energy potential in Malaysia. The author uses MS (Microsoft) Excel platform in establishing neural network model. The ten-year metrological data was considered, the input parameters for the neural network model includes month, minimum temperature, maximum temperature, average temperature, sunshine ratio and relative humidity with one hidden layer while solar radiation as output parameter. MAPE is used as performance metric to compare the result between proposed and existing model. The result of this study clearly show that ANN based model for solar radiation is more precise than other conventional models.

Sheppard [29], had carried out extensive research on prospect of materialising ANN in wind speed predictions and eventually wind power predictions. Three different methods were appropriated into considerations variance ratio, Mortimer's method, ANN. The impact of averaging interval on correlation coefficient was also studied along with estimation of MCP uncertainties. In normal cases the averaging interval will lead to high accuracy in wind sped predictions but have a disproportionate relationship with Wind Power Density (WPD). Although, ANN have capability to predict wind speed in tough conditions, the variance ratio method was recommended by author for general use whose datasets possess good correlation coefficient (>0.8).

Mohandes et al. [30], proposes neural network model for predicting mean monthly wind speeds and mean daily wind speeds. ANN's are compared against auto regressive (AR) model. RMSE was used as performance metrics, regarding neural network author checks for various possibilities of network topology keeping number of input and output neurons constant and varying number of hidden units. The ANN model performs better than its counterpart in predicting the mean monthly wind speed as well as mean daily wind speeds.

Carta et al. [31], throws light over benefits of adopting feature selection methods in ANN model using multi-layer perceptron (MLP) structure to predict mean hourly wind speeds at target site. The two types of feature selection were considered the are filter approach (FA) and wrapper approach (WA) and author uses full dataset without any filters. Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), Index of Agreement (IOA) were considered as performance metrics to compare the results. Multi-layered feedforward network with single hidden layer and network is activated through sigmoid activation function to achieve wind speed as output. A customised software called WEKA

(Waikato Environment for Knowledge Analysis) was used to frame neural network. Filter approach appears to be more ambitious also requires less computational ability, but wrapper approach tends to be favourable in complex topographical conditions.

Quadrado et al. [32], intends to estimate average wind speed for the following hour. Although, the idea used in this paper is little different from our concept, the way author used neural network toolbox in MATLAB gave some idea of using it. Multi-layer perceptron with different number of hidden neurons were tested and backpropagation algorithm was used. MSE was used as performance metrics. For the future improvement to the model author proposes to use other metrological data like atmospheric pressure and atmospheric temperature along with wind speed.

Lopez et al. [33] analysed the effect of adding wind direction as one of the input parameters in neural network approach. Ten-minute interval of wind speed and wind direction were used. In a complex terrain it is quite difficult to assess the wind characteristics, but neural network can be considered in such conditions while traditional methods may fail. The neural network used in this study is a multilayer perceptron with a single hidden layer using supervised learning. Bayesian regularisation was used as training algorithm. There was more than one reference site involved, in order to choose the number of sites three different conditions were tested, sites located at vertices of triangle, sites located in prevailing wind direction, sites having good linear correlation with target site. Two different types of network design were proposed for each of the three possible conditions, the first type is using wind speed data from all sites along with wind direction from one site having good linear correlation with target site and second type is to use only wind speed data. Based on the results authors conclude that network using data from single reference station considering both wind speed and wind direction data outperforms all the other network designs and including wind direction as input parameter can reduce Root Mean Square (RMS) by 23%.

Valezquez et al. [34] made a comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kW h produced by a wind turbine at a candidate site. The network utilises multilayer perceptron topologies (MLP), the design of architecture involves single layer of hidden neurons. The single layer architecture was able to convincingly approximate any continuous transformations. Backpropagation algorithm activated by sigmoid function was used as training algorithm and Levenberg-Marquardt algorithm was used in minimisation of mean-square error made during the training process. Also restraining the number of hidden layers to one reduced training times. The wind data comprising of mean hourly wind speed and wind direction of reference site are used as input signals as the study involves more than one reference site data, number of neurons in the input layer is twice the number of reference site used while single output neuron corresponding to mean hourly wind speed at target site. Two different scenarios for artificial neural network were considered, the model using single reference station and model using two reference stations but either of the cases the linear MCP method uses only one reference stations. The number of neurons in the hidden layer is assumed to be the same for all the cases thus having this idea, number of neurons in the hidden layer was increased in two's until validation shows improvement. Finally, the number of neurons in the hidden layer was set to 15 as the validation failed to improve. The short-term wind data series of reference and target site were divided into three random subsets of training, validation and test data at the proportion of 60%, 20% and 20% respectively. Tests were performed with the help of algorithm known as Trainlm from MATLAB toolbox of neural networks. The model using two reference stations usually outperformed the model using single reference while single reference station model using wind speed and wind direction as input parameters gave reduced MAPEs in 87% of the cases in comparison with variance ratio method (VRM) when estimating long term mean hourly wind speeds at target sites. The conclusion from this report is that amongst all scenarios analysed MAPE of energy obtained using ANNs was less than that of linear MCP model.

Fadare [35] proposed an ANN approach in modelling of solar energy potential. The main aim of this study is analysing the possibility of implementing ANN to model non-linear relationship between solar radiation and other metrological parameters. Multi-layered perceptron (MLP) models were adopted, encompassing feedforward back-propagation network with different topologies designed using MATLAB neural network toolbox. The network consists of input layer, hidden layer and output layer, the seven input parameters namely latitude, longitude, altitude, month, mean temperature, mean sunlight duration, relative humidity serves the input layer corresponding to global solar radiation as the single output neuron. Single and double hidden layer topologies were tested by varying number of hidden neurons from 5 to 15. The input layer possesses no transfer function while hidden layer neurons were powered by tangent sigmoid function and output layer neuron activated by linear function. Levenberg Marquardt (LM) and Scaled Conjugate Gradient (SCG) were used as the training algorithm for the network. The input and the target datasets were randomly divided into three subsets of training, validation and testing at the proportion of 50%, 25%, 25% respectively. The performance of the networks with different topologies were measured based on the correlation coefficient (R) between the target value and the actual value. The outcome of this study indicate that the ANN based model is meticulous for prediction of solar radiation. The ANN model is reliable and can be used in forecasting solar radiation for any region provided that meteorological and geographical data (latitude, longitude, altitude, month of the year, mean sunshine duration, mean temperature, and relative humidity) are available.

Bilgil et al. [36] applied ANN in predicting wind speeds at target stations using more than one reference stations data. Multi-layered perceptron (MLP) models were adopted, consisting of an input layer, two hidden layer and an output layer. The mean monthly wind speeds from the reference stations along with their respective months were used as input parameters, the mean monthly wind speeds of the target stations were estimated in the output layer. The reference station is selected based on criteria which is defined by high correlation coefficient between the reference and target station. Resilient propagation (RP) algorithm was used as learning theorem. The input layer possesses no transfer function while hidden layer neurons were powered by logistic sigmoid transfer function and output layer neuron activated by linear transfer function. The performance of the network is defined by two metrics, mean absolute percentage error (MAPE) and correlation coefficient (R) between the target values and the output values of the network. The MAPE values were found to be varied from 4.49% to 14.13% from actual value for all the stations used. In overall ANN tool looks to be a promising tool in predicting wind speeds at target stations, also major advantage of ANN model is that it can predict the wind speed at target station satisfactorily until unless the wind speed data of reference stations are available disregarding the need for topographical and other climatological data.

Sagbas et al. [37] analysed the feasibility of using ANN in estimating wind speeds. Multilayer perceptron (MLP) model was established, with an input layer and double hidden layer and an output layer. Mean monthly relative humidity, mean monthly atmospheric pressure and mean monthly atmospheric temperature were assumed as the input parameters to the network while single output neuron corresponds to mean monthly wind speed of considered target site. Feed forward network with backpropagation algorithm was used and no activation transfer function was used for the input layer and logarithmic sigmoid function for hidden layers and linear function for output layer were used. Mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) were used as performance metrics to evaluate the deviation from the actual value. The MAPE was found to be 8,694% for one of the meteorological stations and the best result was found to be 4,038% for other meteorological station. Thus, the proposed neural network-based approach is accurate enough in the prediction of the wind speed output.

Fadare [38] fabricated a model to capture the wind speed profile using ANN. A wind speed profile model will be of great guidance for the siting and sizing of wind power plant installations. Multi-layered perceptron (MLP) models were adopted, encompassing feedforward network with back-propagation algorithm with different topologies designed using MATLAB neural network toolbox. The network comprised of three layers, the geographical parameters of the sites specifically latitude, longitude, and altitude of the station and the corresponding month so all together these four parameters are considered as input to the network and monthly mean wind speed as single output neuron. single and double hidden layer topologies were used and randomly into three subsets training, validation, and testing datasets. The performance of the different networks with multiple configurations and different training algorithms were analysed based on the correlation coefficient (R) between the predicted and the measured values in order to sort out the network with optimum prediction capacity.

The conclusions of this research show that ANN based model is a promising tool to consider and has performed satisfactorily in regards with prediction of wind speed profile.

3.4 Design of artificial neural network model

ANN models are inspired by the biological neural system, with capability to learn, store and recall information based on a given training dataset. They are 'black-box' modelling technique capable of performing non-linear mapping of a multidimensional input space onto another multidimensional output space without the knowledge and dynamics of the relationship between the input and output spaces. The construction of the model involves Multilayer Perceptron architecture (MLP). Multi layered feed forward backpropagation algorithm with different layouts were designed using NN toolbox version 9.0.0.341360(2016a) for MATLAB. The network consists of three layers: input layer, hidden layer and output layer. Three input parameters were considered for the model, mean hourly wind speeds and wind direction as sine and cosine vectors of the angle and similarly three output neurons in output layer corresponds to wind speed and sine and cosine vectors of wind direction respectively.

Different networks with single hidden layer topologies were used and number of neurons varied from 5 to 30 neurons at interval of five neurons to improve generalisation capability of network [35].The input layer involves no transfer function, while hidden layer utilises hyperbolic tangent sigmoid (tansig) function as an activation function and output layer use linear (purelin) function. To avoid overfitting 'early stopping' technique was used, also the method improves generalisation on network performance which is available in default for all feedforward backpropagation network type in neural network fitting toolbox. The general layout of FFNN is presented (figure 8).

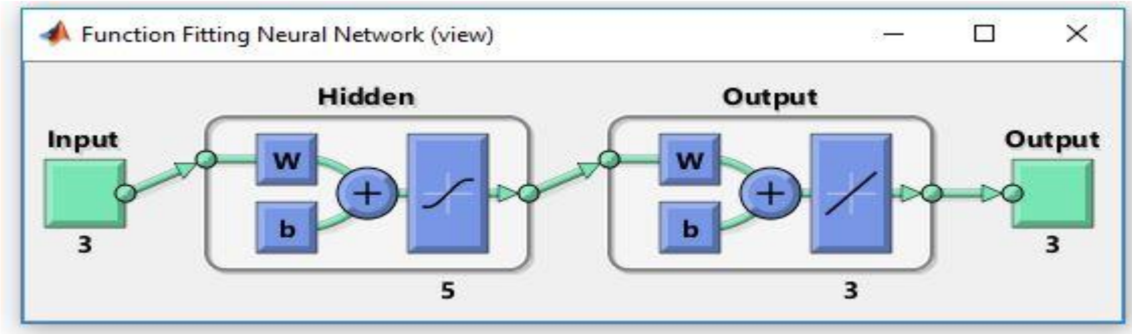


Figure 8 – FFNN with 3 inputs, 1 hidden layer with 5 neurons and 3 outputs (credit MATLAB)

Levenberg Marquardt [34] was used as learning algorithm for the training of the network. The networks were designed individually for each site. In order to avoid overfitting of the data and to improve overall generalisation of network, the datasets were divided randomly into three subsets: training dataset, validation dataset, testing dataset using dividerand procedure. The training data consists of 60% of entire dataset and 20% of data for validation and 20% for testing process separately. The results of training dataset are used for evaluating gradients and revising network weights and biases. The validation data subset was used to check the progress of the ANN training, optimising their parameters. The test data remains unused in either the training or validation period which is finally used to calculate error percentage of the developed network i.e., to calibrate how well network performs when unused set of data is introduced to it. The maximum number of 1000 iterations were used and minimum gradient of 10^-6 was used in training process. The performance of the network was assessed based on correlation coefficient(R) and mean square error (MSE) calculated from output and the target values in order to determine the network with optimum performance. Finally, the best performing network was trained with 90% of the data and 5% to validate and the remaining 5% to test the final performance of the model.

There were two types of approaches tested. Approach 1 utilises one-year concurrent wind data (mean hourly wind speed and wind direction as sine and cosine vectors) from reference and target sites to build the model and relation is established, then the built neural network function was again applied on same one-year reference dataset resulting in new predicted one-year target dataset which was compared against the actual measured one-year dataset (figure 9). The purpose of second approach is to examine the performance of network when a completely new set of data is introduced into the model which is trained with different sets of data. This study is only made to understand the neural network model more precisely. Different sets of data are tried i.e.., training with 10 months of data and testing on 2-month new data, training with data from intermediate 10 months and testing with first and last month, training with 9 consecutive month data and testing on 3 consecutive months.

The concept behind building of second approach is to understand the possibility of using functional relationship built on certain period of months which have different wind characteristics when compared with other period of months. The wind characteristics in winter is different from summer, i.e. the wind speed will be higher during winter seasons while less wind speeds could be experienced during summer seasons. Thus, model is built out of very interest to know about the performance and behaviour of the network. (figure 10). There was one specific case where the concurrent period of the reference and target sites were about 1 year and 11 months. In this case the last one year of wind data from reference and target sites were used to train the network and remaining 10 months of reference site's wind data was used to test the network's performance, resulting in new predicted target site 10-month data which was compared against the actual 10-month measured wind data (figure 10). The approach 2 is experimented on all sites except Jordan where only six months period of concurrent period is only available.

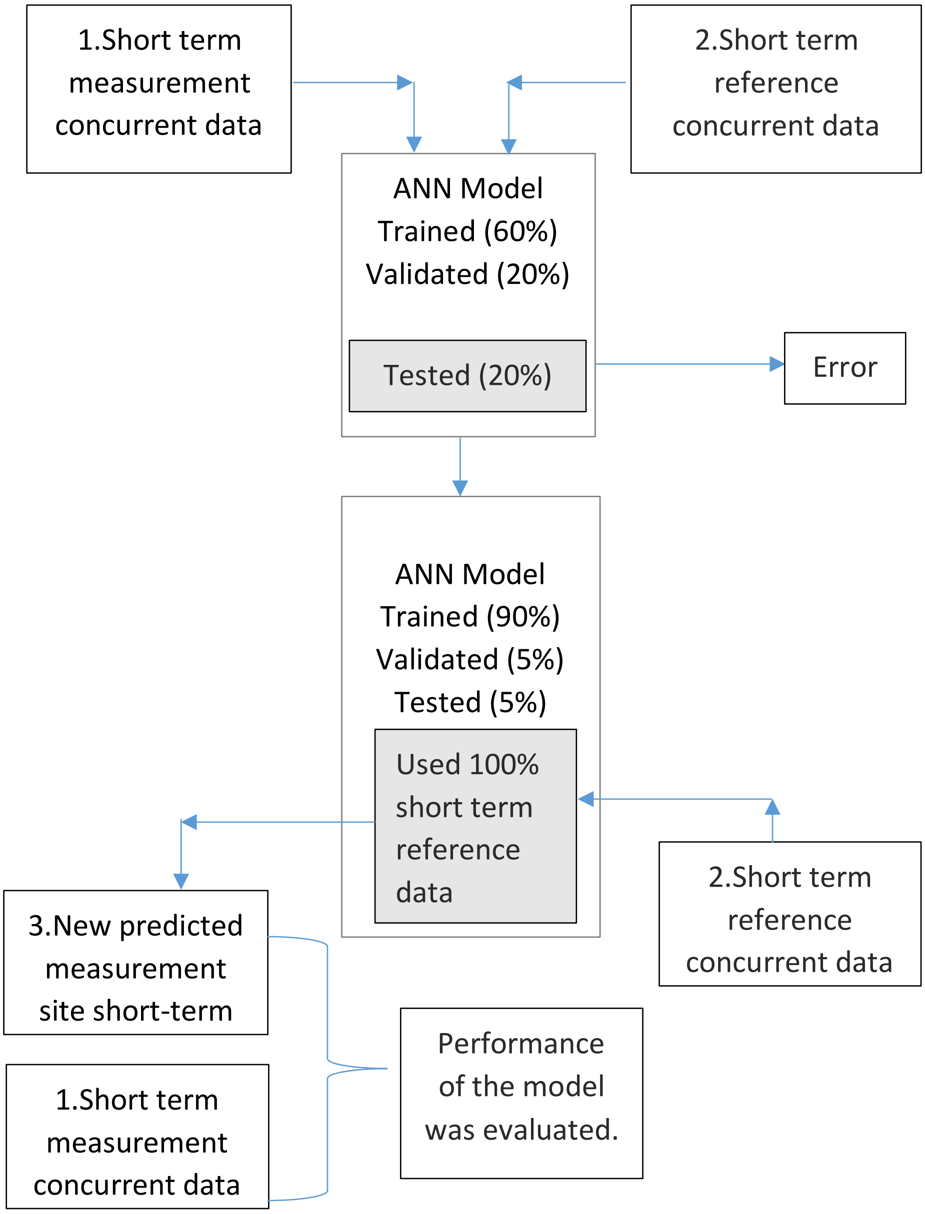


Figure 9 – ANN Approach 1

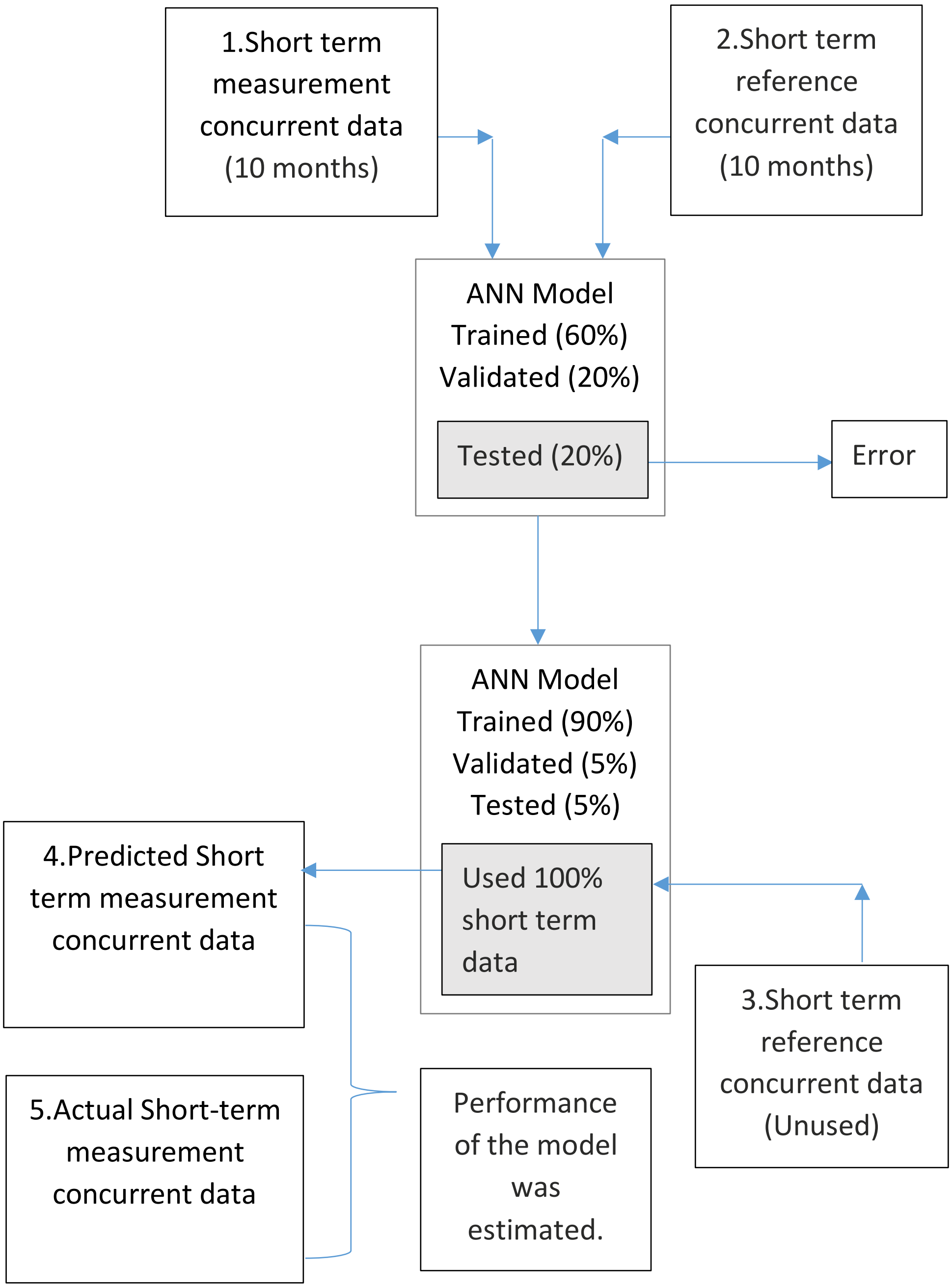


Figure 10 – ANN approach 2

RESULTS AND DISCUSSION

4.1 Sites description and data availability

4.2 ANN model performance assessment

4.3 Comparison of different correlation methodologies

4.4 Statistical error analysis

4.5 Approach 2 analysis

4 RESULTS AND DISCUSSION

4.1 Sites description and data availability

Wind data used for MCP methodologies should generally be able to cover all possible seasonal variations because the relationship established between reference site and measurement site will be used to estimate long term wind characteristics at the target site. The stochastic nature of the wind in different seasons for example wind speeds will be high in winter and low wind speeds are experienced in summer, making it important that dataset covering at-least one full year should be used, in some cases if the complexity of the terrain is high it is suggested to take two consecutive years for measuring the wind characteristics.

4.1.1 Japan

The mast is sited in a flat area near the sea coast, with increasing terrain complexity towards east. The land cover has been essentially characterized as rice fields and farmland, over the flat area, while forestry is present over the complex terrain areas. The mean hourly wind data (wind speed and wind directions) are available for a period of one concurrent year for reference and measuring mast locations.

4.1.2 Saudi Arabia

The mast is sited in a flat plateau with moderate terrain complexity, particularly at the edges of the plateau. The land cover has been essentially characterized as desert with rare vegetation. The measurements in Saudi covers the period of one year and eleven months, thus wind speed and wind directions are available for both reference and mast locations.

4.1.3 Jordan

The mast is sited in smooth undulated terrain with low complexity. The land cover has been essentially characterized as desert with rare vegetation. The measurement campaign in Jordan is the shortest among all the cases covering only six months, thus only six wind data is available for estimating MCP methods.

4.1.4 France

The mast is sited in a flat plain. The land cover has been essentially characterized as farmland, with scattered settlements. The mean hourly wind speeds and wind directions are available for the period of one year for both reference and measurement mast locations.

4.1.5 Russia

The mast is sited in a flat plain. The land cover has been essentially characterized as low vegetation, mixed with farmland. The measurement campaign in Russia also took place for one year which entails mean hourly wind speeds and wind directions.

4.2 ANN model performance assessment

The performance of the designed network was assessed based on correlation coefficient (R) between target and output dataset and mean square error (MSE). The datasets used for the different method is presented in table 2 and the results for the optimised neural network model is presented in table 3.

Table 2 – Description of data used for different approaches

Site	Data availability (wind speed and wind directions)	Data for approach 1	Data for approach 2 training	Data for approach 2 testing
1. Japan	01/03/2015 to 29/02/2015	01/03/2015 to 29/02/2015	01/04/2015 to 31/01/2016	01/03/15-31/03/15 and 01/02/16-29/02/16
2. Saudi Arabia	02/06/2015 to 30/04/2017	01/05/2016 to 30/04/2017	03/04/16-30/04/17	02/06/15-02/04/16
3. Jordan	01/05/2017 to 31/10/2017	01/05/2017 to 31/10/2017	------	------
4. France	01/11/2016 to 31/10/2017	01/11/2016 to 31/10/2017	01/01/2017 to 31/10/2017	01/11/16 to 31/12/16
5. Russia	01/09/2013 to 31/08/2014	01/09/2013 to 31/08/2014-	01/12/2013 to 31/08/2014	01/09/13 to 30/11/13

Table 3 – Performance of the optimized network

Type of model	Name of the site.	Training algorithm	No. of neurons in single hidden layer	Training dataset MSE	Training dataset R-value	Test dataset MSE	Test dataset R-value
Approach 1	JAPAN	LEVENBERG-MARQUARDT	20	1.614	0.932	1.580	0.924
	SAUDIARABIA		5	1.761	0.938	1.536	0.946
	JORDAN		10	1.433	0.944	1.170	0.954
	FRANCE		25	0.449	0.973	**0.434**	0.975
	RUSSIA		10	1.240	0.947	1.131	0.955
Approach 2	JAPAN		25	1.535	0.928	1.313	0.936
	SAUDIARABIA		10	1.711	0.939	1.648	0.946
	FRANCE		25	0.461	0.972	**0.451**	0.973
	RUSSIA		15	1.276	0.954	1.218	0.953

The network performances are calibrated in terms of the Mean Square Error (MSE), and correlation coefficient (R) between the target and the output values. The error values of final optimised network for training and testing datasets are shown in table 3. Based on the training dataset, the MSE and coefficient (R) ranged from 0.461–1.761 and 0.928–0.973, respectively, while corresponding ranges of 0.434–1.648 and 0.924–0.975, respectively were obtained on the testing dataset (table 3).

From this study, it is realized that ANN is a convenient method to apply for the prediction of the wind speed and wind direction.

4.3 Comparison of different correlation methodologies

The results from Approach 1 are presented in the section below and are compared to the results obtained from regression, matrix, neural network approach from commercial software. In-order to understand the results better number of graphs and charts were used. For wind speed, time series graphs, scatter plot graphs built using MS excel were showcased and wind directions wind rose plots are generated from WAsP climate analyst 3.1 software (freeware) and to check the wind speed distribution MATLAB distribution fitting toolbox was used.

The three parameters are used to analyse the wind speed statistics and to see the convergence between the measured and the output values were root mean square error (RMSE), mean absolute percentage error (MAPE) [36], index of agreement (IoA) [34] where each of them is defined as

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)\right)^2} \qquad (8)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} abs\left(\frac{p_i - a_i}{a_i}\right) * 100 \qquad (9)$$

$$IoA = 1 - \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}\left(abs(p_i - \bar{a}) + abs(a_i - \bar{a})\right)^2} \; ; \qquad 0 \le IoA \le 1 \tag{10}$$

respectively, in equations (8)-(10), n is the number/length of data analysed, ai is the actual or measured or target data, pi is the predicted data or the output data and ā mean of actual or measured data.

The frequency distribution at the site is estimated by

$$f(u) = \frac{k}{A}\left(\frac{u}{A}\right)^{k-1} exp\left(-\left(\frac{u}{A}\right)^{K}\right) \tag{11}$$

Where A is Weibull scale parameter (m/s)

K is Weibull shape parameter

u is fitted mean velocity (m/s).

The wind rose is the classical method of graphically representing wind conditions, direction and speed at a specific location over a period and wind data is then sorted by wind direction. The wind direction data is divided into twelve sectors, 30°sector each, in preparation for plotting a circular graph. The wind rose provides the information about the frequency of winds blowing from directions. The length of the rose around the circle is directly related to the frequency of time that the wind blows in a specific direction.

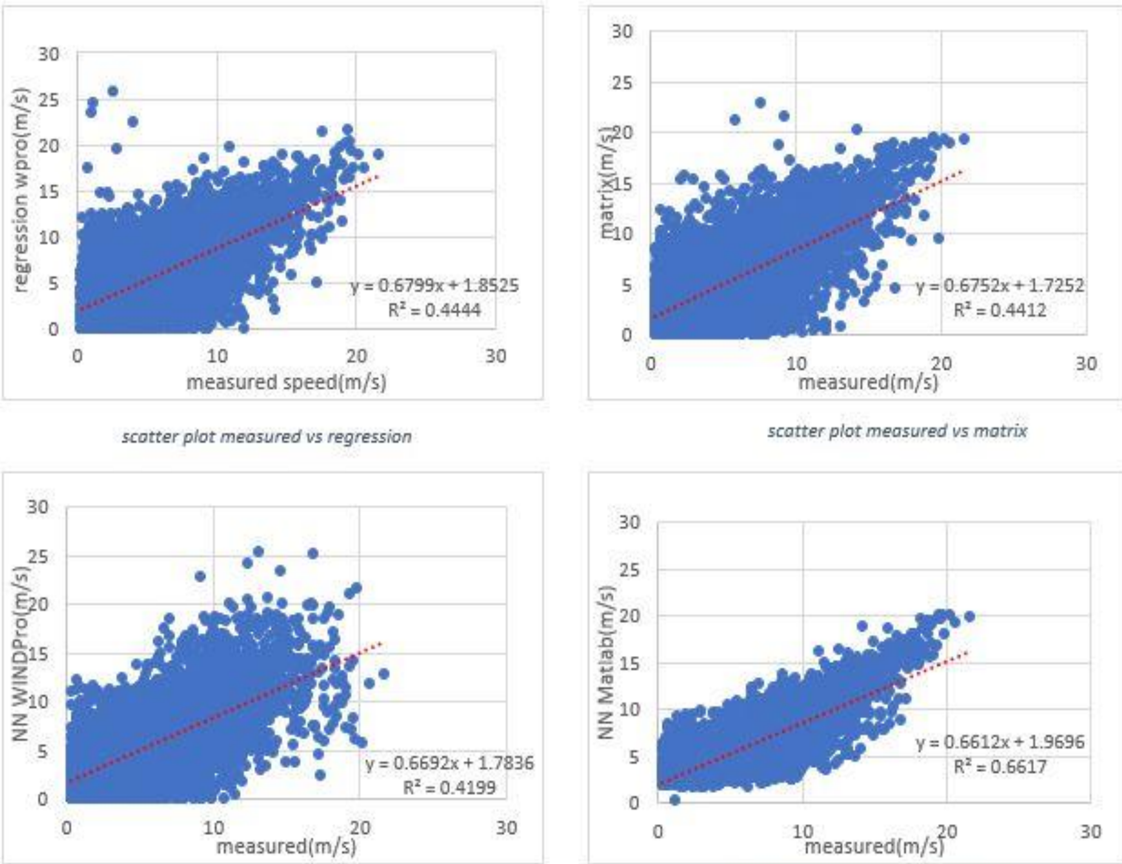
The results for each site are presented below.

4.3.1 Japan site

The results of the scatter plot for Regression, Matrix, commercial Neural network and developed Neural network from MATLAB were presented in figure 11. The site is classified as complex due to its topographical structures.

4.3.1.1 Wind speed analysis

The scatter plot depicts the nature and correlation between the data. The results of the scatter plot reveal that the neural network from MATLAB method has more linear relationship in comparison with other methods. The slope values of all the methods are close to each other with regression method having slightly high value. Among the other three methods regression method tend to perform better in terms of linearity. For example, in (figure 12), the developed ANN approach captures the variations in majority of time periods but in some cases, they tend to generalise the prediction leaving out the extreme values.



scatter plot measured vs regression

scatter plot measured vs matrix

scatter plot measured vs NN WINDPro

scatter plot measured vs NN MATLAB

Figure 11 – Scatter plot for wind speeds (Japan site)

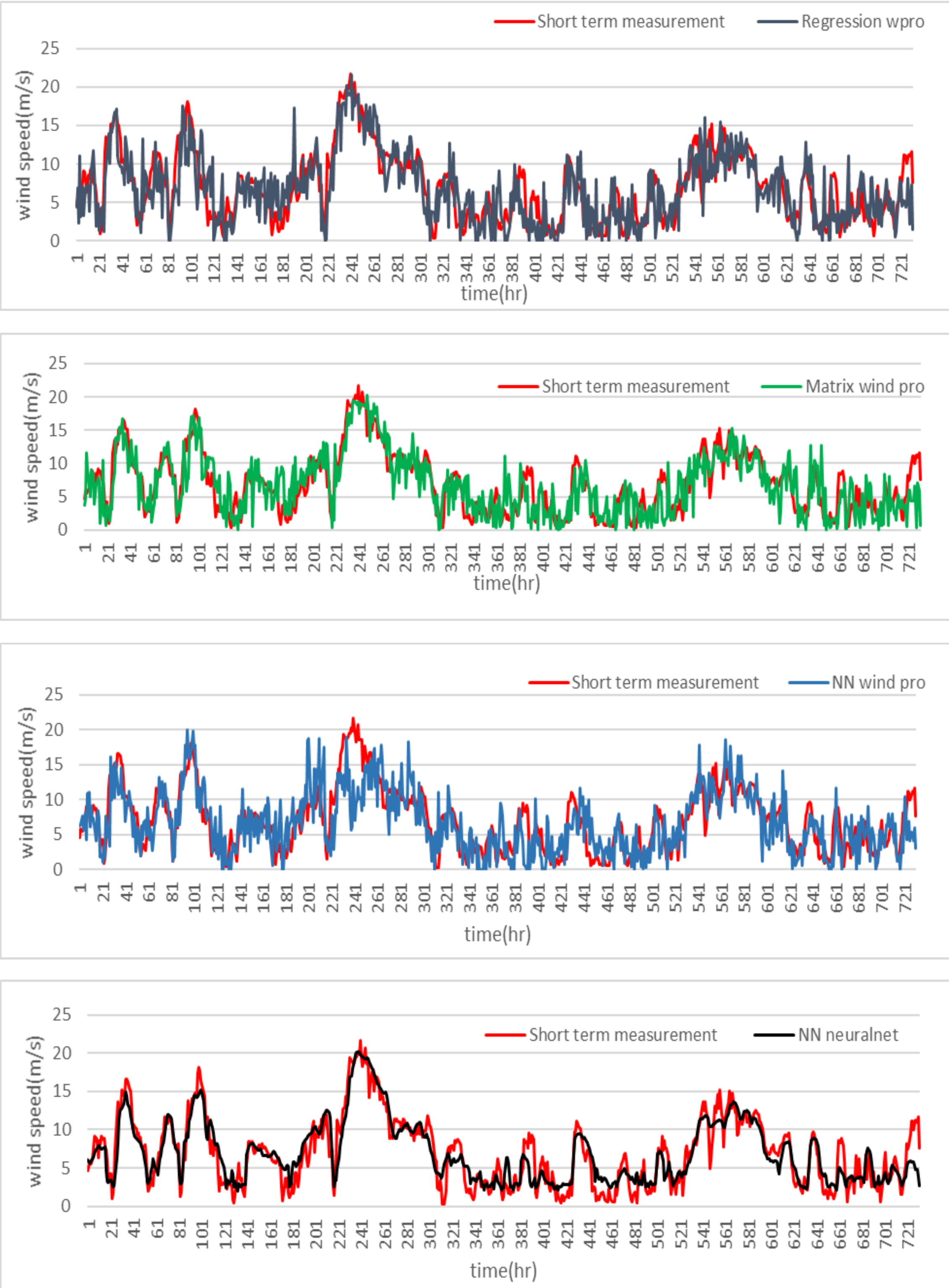
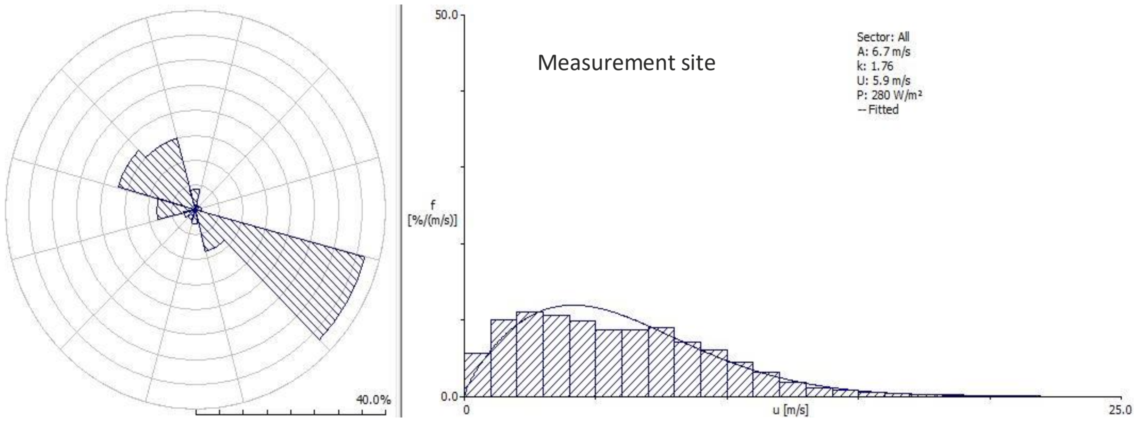
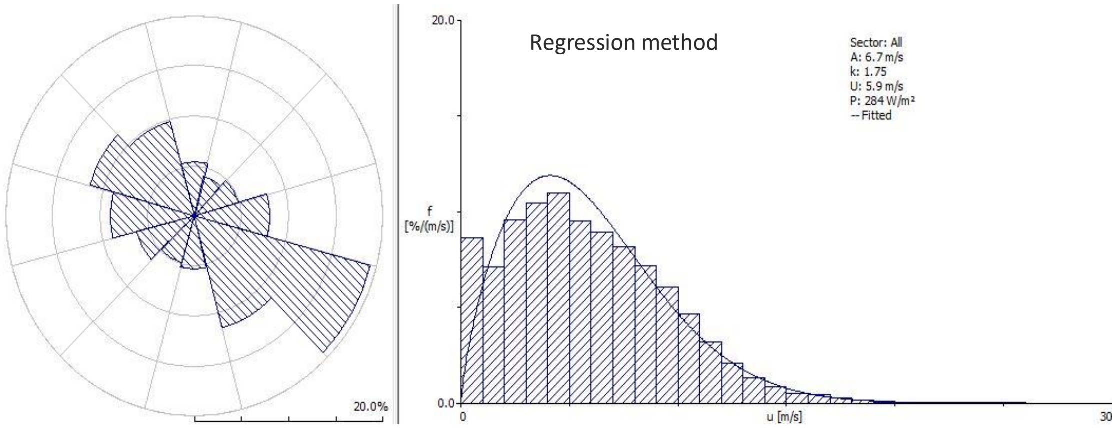


Figure 12 – Time series of with speeds comparison (Japan site)

4.3.1.2 Wind direction analysis

Wind rose representation is used to analyse the wind direction predictions from different methods and compared against measured data. The wind speed distribution curve is also presented along with the wind rose.





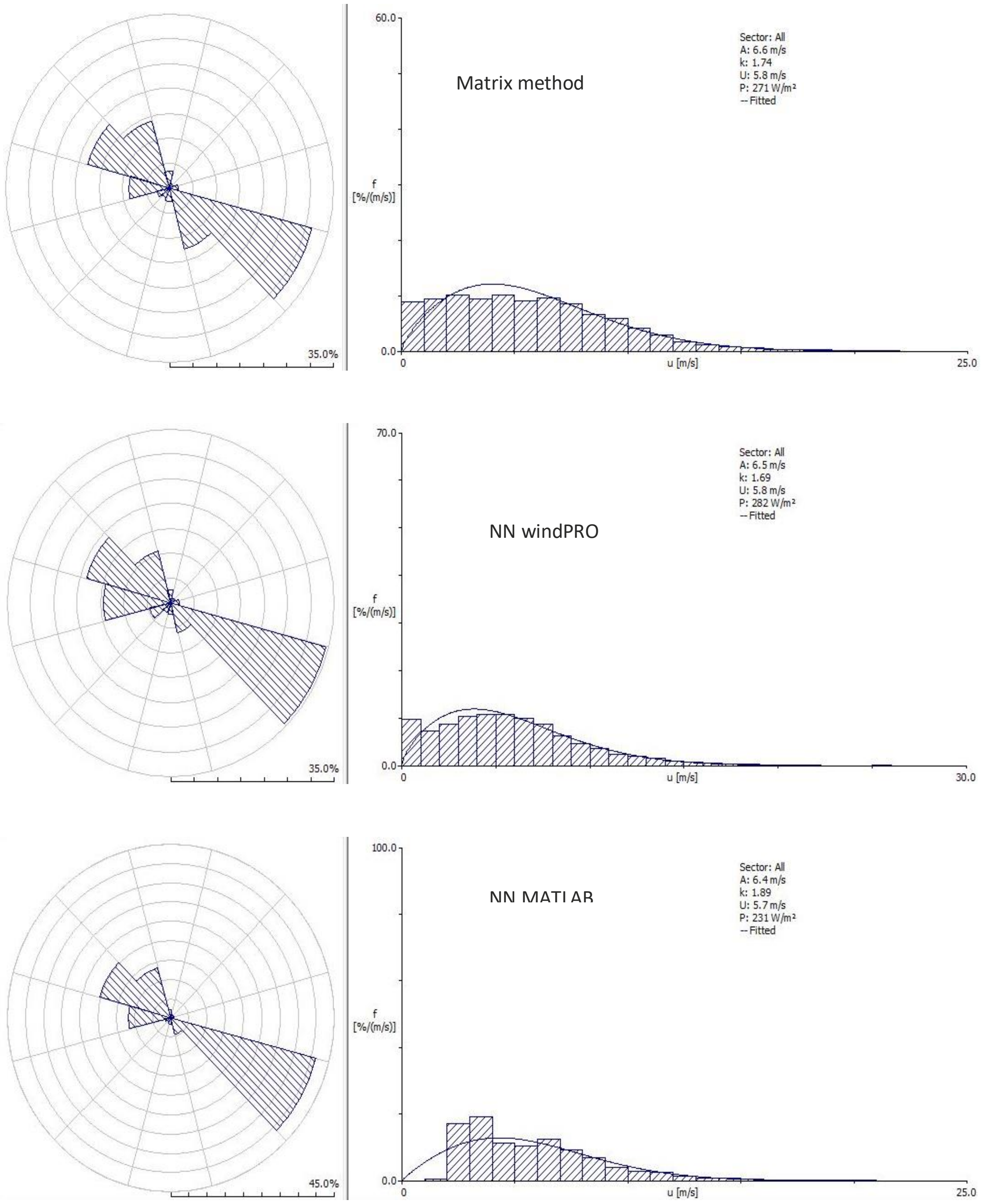


Figure 13 – Wind speeds frequency rose and histogram comparison (Japan site)

The rose graph for measurement site indicates that for the year in consideration most of the wind is recorded in sector 5 followed by sector 11 and sector 12 respectively (table 4).

Table 4 – Results from WAsP wind analysis software (Japan site)

Method	Major sector(s) of wind	A(m/s)	k	Fitted mean velocity(u) (m/s) In all sectors	Available power density (W/m^2) In all sectors
Measurement data	5^{th},11^{th},12^{th}	6.7	1.76	5.9	280
Regression	5^{th},11^{th},12^{th}	6.7	1.75	5.9	284
Matrix	5^{th},11^{th},12^{th}	6.6	1.74	5.8	271
NN WindPRO	5^{th},11^{th},12^{th}	6.5	1.69	5.8	282
NN MATLAB	5^{th},11^{th},12^{th}	6.4	1.89	5.7	231

Lower k values correspond to wider wind speed distributions, which means that winds tend to vary across a wide variety of speeds. Higher k values correspond to narrower distributions of wind speed, which means that wind speeds tend to remain within a narrow range. A is the parameter of the Weibull scale in m/s; a measure of the characteristic wind speed of the distribution. A parameter directly corresponds to the fitted mean wind speed u.

In comparison with the other methods and measurement data the proposed ANN method contributes highest k value, indicating the existence of narrow wind speed distributions (table 4). The inferences from figure 14 also clearly explains the more frequencies of occurrence in certain wind speeds. The results of table 4 indicates that regression method outperforms all the other methods in this site. The mean velocity (u) of developed method experiences detrimental deviation of 3.4% from the measured value.

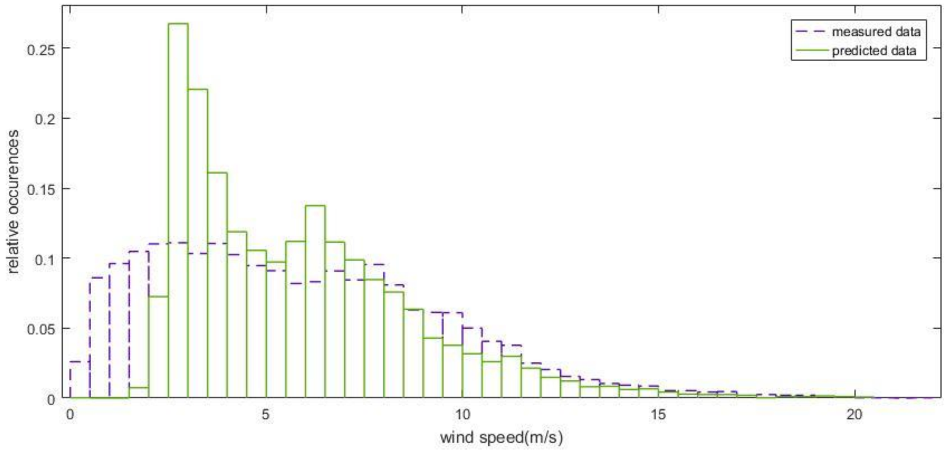


Figure 14 – Wind speed histogram comparison (Japan site)

4.3.2 Saudi Arabia site

The results of Saudi are presented in the order of wind speed and wind direction. The terrain is classified as moderately complex.

4.3.2.1 Wind speed analysis

The results of the scatter plot (figure 15) reveal that the designed neural network model has more linear relationship in comparison with other methods. Among the other three methods regression method tend to perform better in terms of linearity. The methods from commercial software generally try to capture all the variations resulting in lot of outliers, but the proposed model tries to force the outliers into the more generalised mean value, resulting in more linearity. For example, this is evident from figure 16 designed model fails to predict the lower wind speeds, although it can generalise the pattern of measured wind speed. The inferences from slope value show that proposed model possess the lower value which means that dataset deviate from ideal slope more in comparison with other methods. The time series prediction of wind speeds is presented following the structure of regression, matrix, NN from industrial software, NN developed model respectively (figure 16).

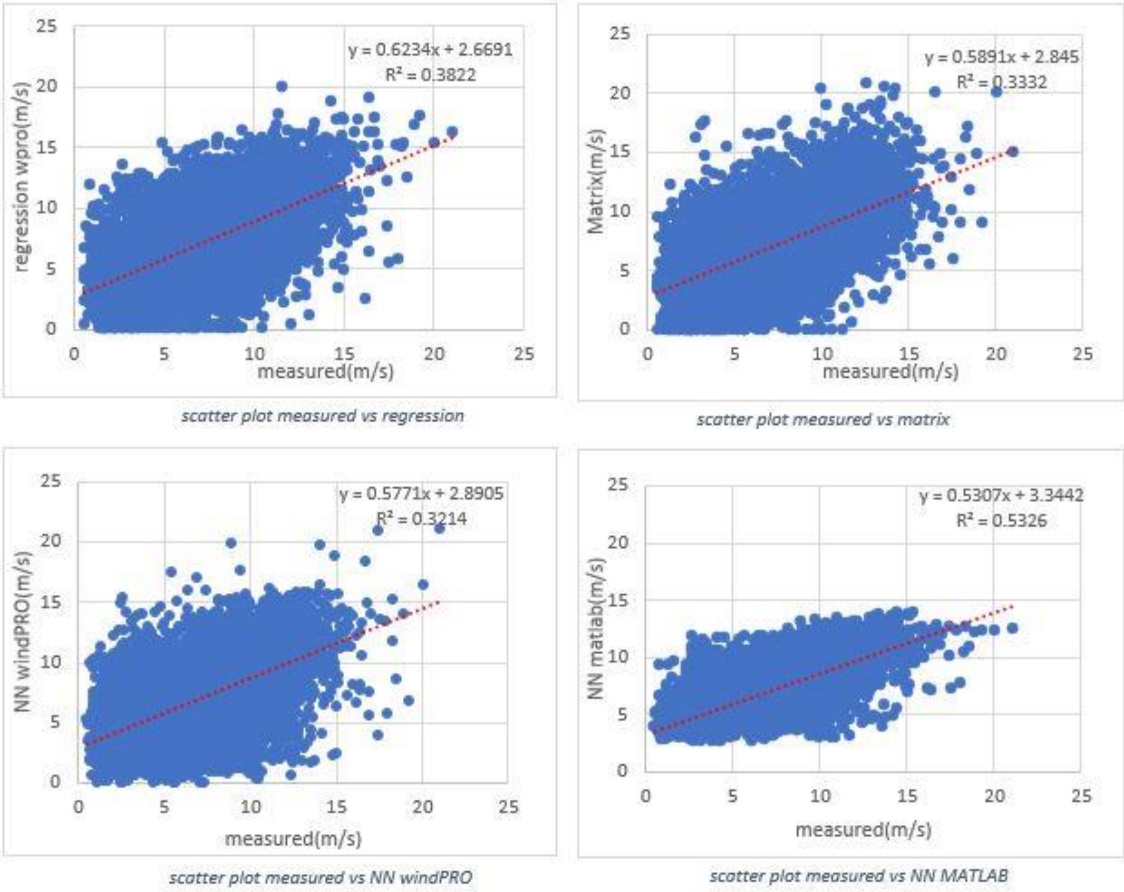
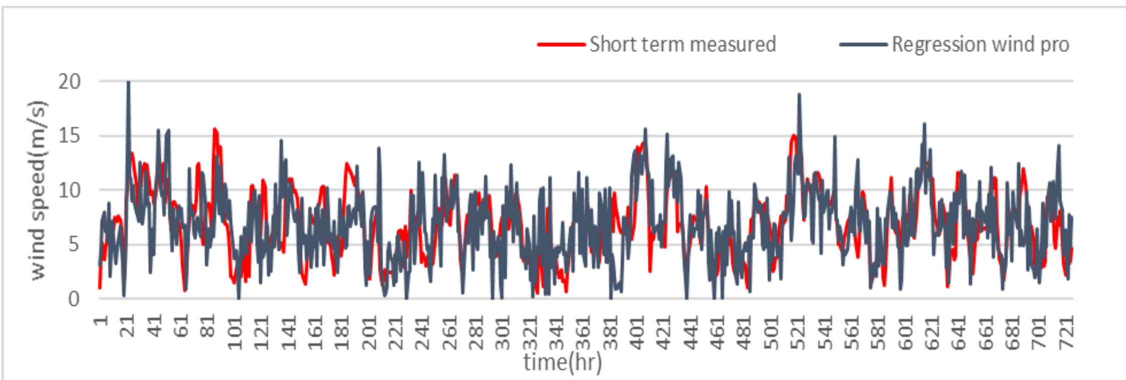


Figure 15 – Scatter plot for wind speeds (Saudi Arabia site)



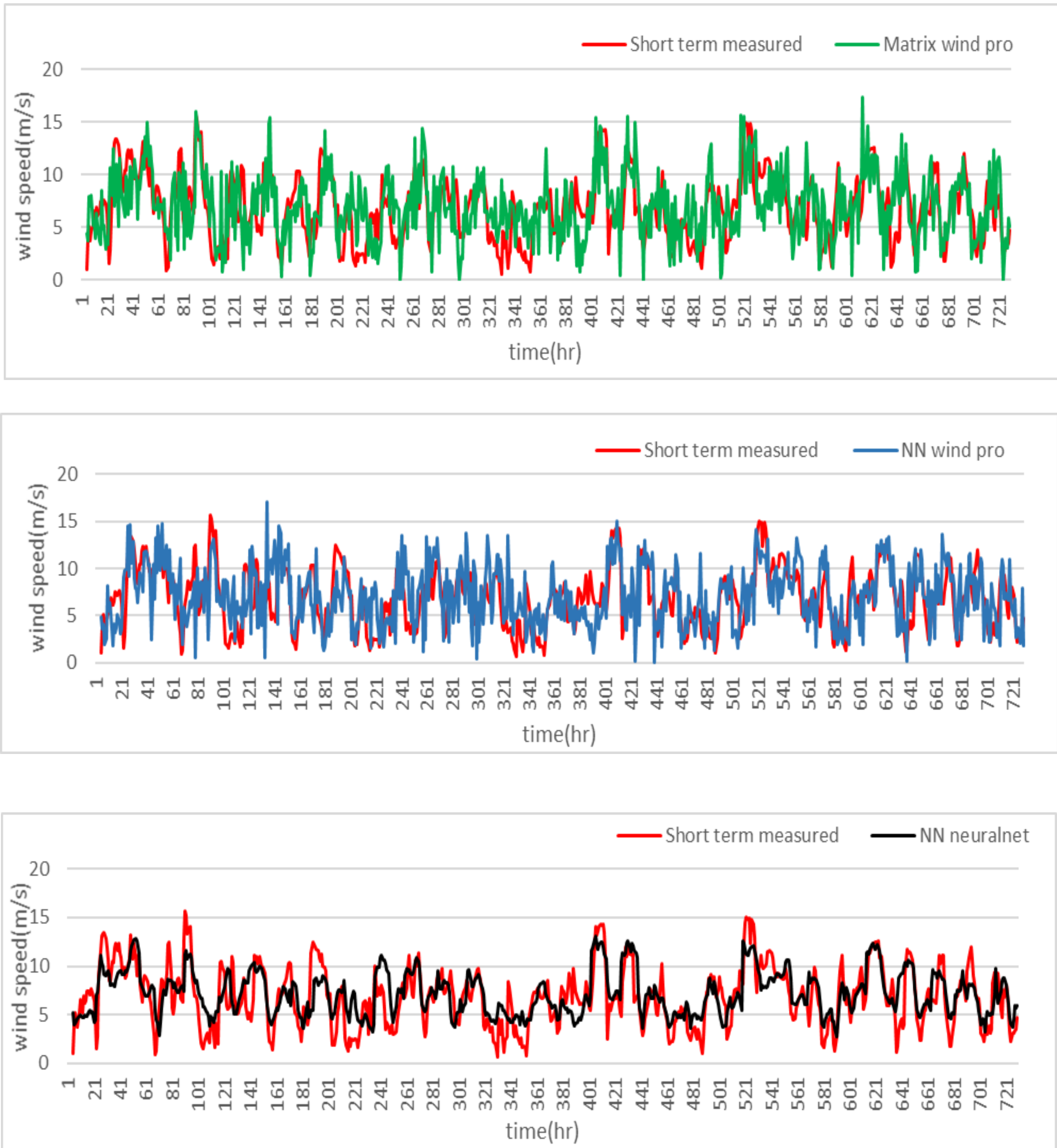
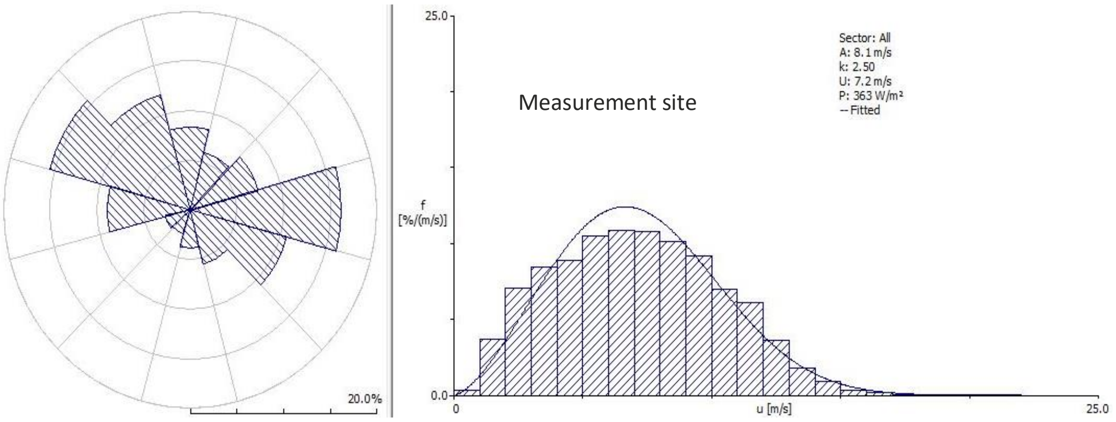
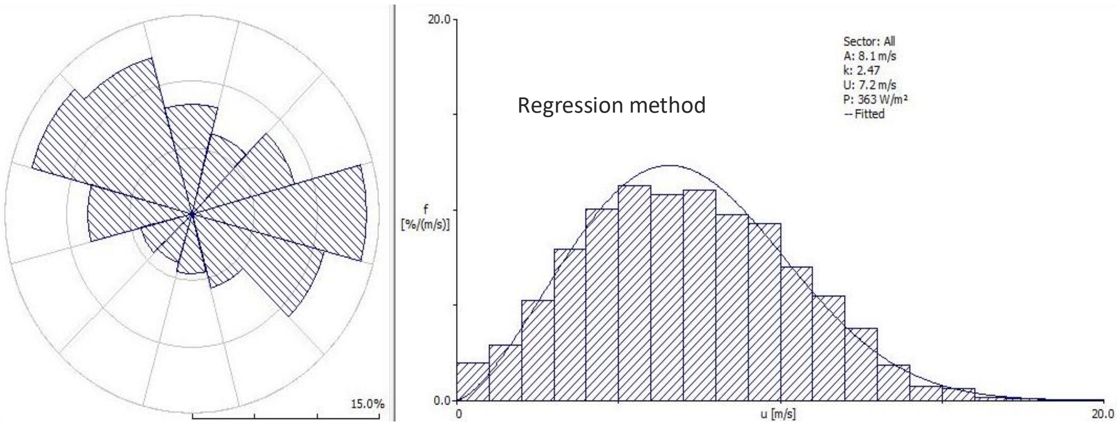


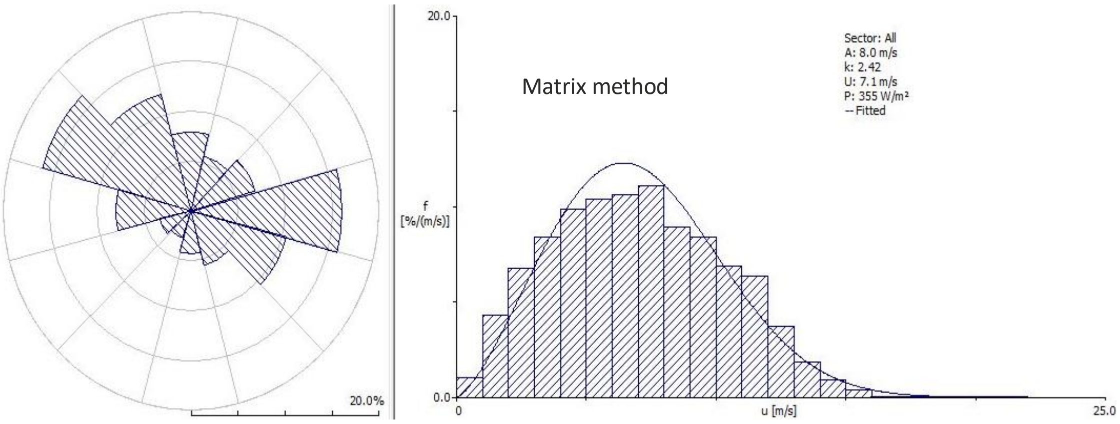
Figure 16 – Time series of with speeds comparison (Saudi Arabia site)

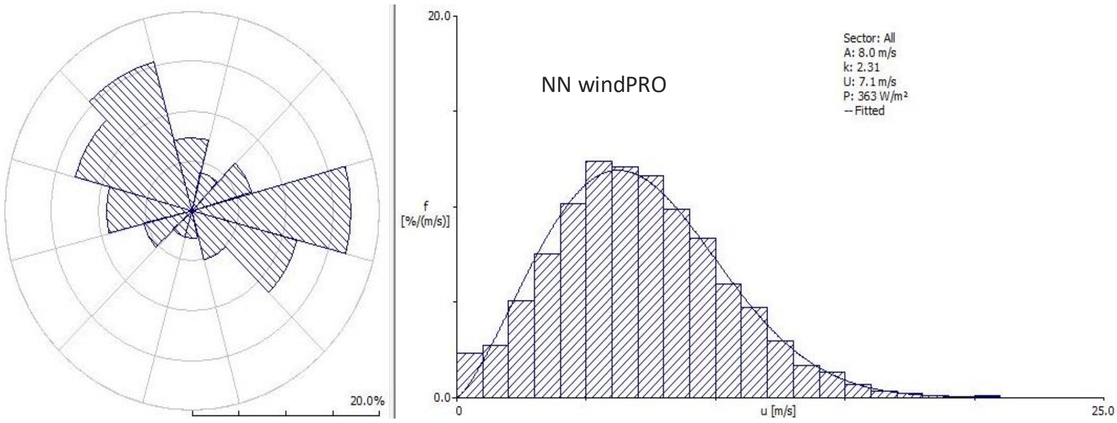
4.3.2.2 Wind direction analysis

Wind rose representation is used to analyse the wind direction predictions from different methods and compared against measured data. The wind speed frequency distribution curve is also presented along with the wind rose (figure 17).


20.0%
25.0
0.0
f
[%/(m/s)]
0
u [m/s]
25.0
Measurement site
Sector: All
A: 8.1 m/s
k: 2.50
U: 7.2 m/s
P: 363 W/m²
-- Fitted


15.0%
20.0
0.0
f
[%/(m/s)]
0
u [m/s]
20.0
Regression method
Sector: All
A: 8.1 m/s
k: 2.47
U: 7.2 m/s
P: 363 W/m²
-- Fitted


20.0%
20.0
0.0
f
[%/(m/s)]
0
u [m/s]
25.0
Matrix method
Sector: All
A: 8.0 m/s
k: 2.42
U: 7.1 m/s
P: 355 W/m²
-- Fitted


20.0%
20.0
0.0
f
[%/(m/s)]
0
u [m/s]
25.0
NN windPRO
Sector: All
A: 8.0 m/s
k: 2.31
U: 7.1 m/s
P: 363 W/m²
-- Fitted

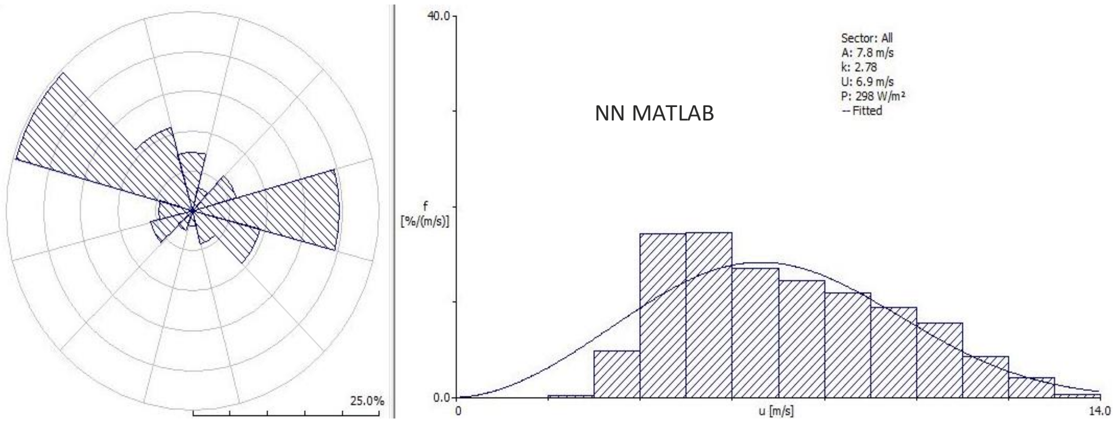


Figure 17 – Wind speeds frequency rose and histogram comparison (Saudi Arabia site)

The results from wind rose graphs are help us to understand performance of each method more elaborately. The rose graph for measurement site indicates that for the year in consideration most of the wind is recorded in sector 4 followed by sector 11 and sector 12 respectively (table 5).

Table 5 – Results from WAsP wind analysis software (Saudi Arabia site)

Method	**Major sector(s) of wind**	**A(m/s)**	**k**	**Fitted mean velocity(u) (m/s) In all sectors**	**Available power density (W/m^2) In all sectors**
Measurement data	4^{th},11^{th},12^{th}	8.1	2.5	7.2	363
Regression	4^{th},11^{th},12^{th}	8.1	2.47	7.2	363
Matrix	4^{th},11^{th},12^{th}	8.0	2.42	7.1	255
NN WindPRO	4^{th},11^{th},12^{th}	8.0	2.31	7.1	363
NN MATLAB	4^{th},11^{th},12^{th}	7.8	2.78	6.9	298

The distribution of the curve can be statistically explained with the value of k, the form factor k for developed model has the higher value when compared with other methods. The comparison of frequency distribution curves between measured and designed model is presented in figure 18.

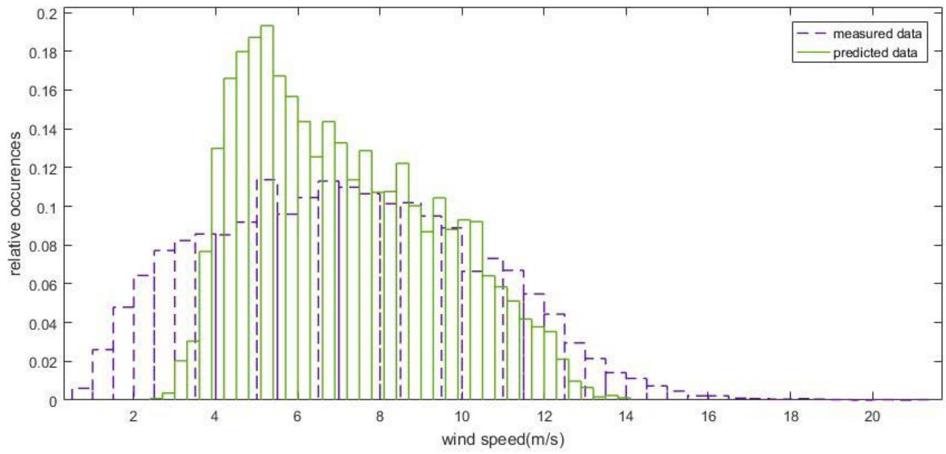


Figure 18 – Wind speed histogram comparison (Saudi Arabia site)

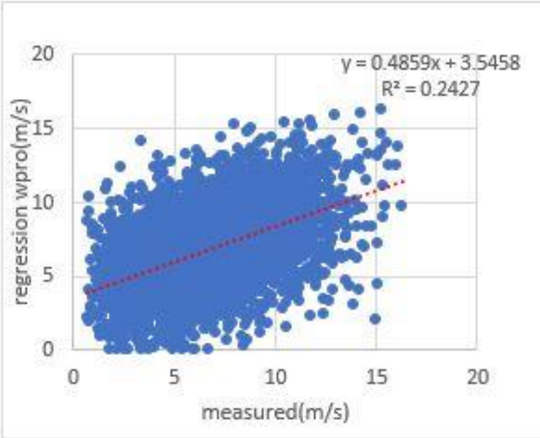
The frequency of occurrences of wind speed are concentrated in a narrow bandwidth. The regression method performs better overall. The fitted mean velocity (u) of NN MATLAB method encounters a negative deviation of 4.2% from the measured value.

4.3.3 Jordan site

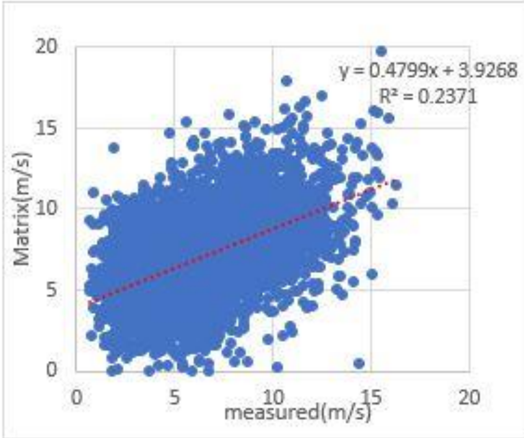
The analysis of MCP methodologies in Jordan involves only six-month data. The site is characterised as low complex. The wind speed and wind direction analysis are presented below.

4.3.3.1 Wind speed analysis

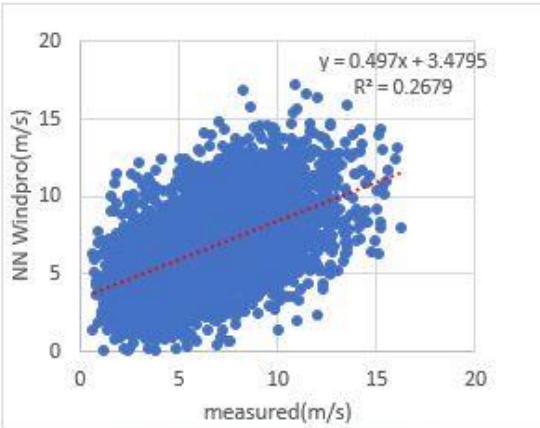
The results of the scatter plot reveal that the neural network from developed model has more linear relationship in comparison with other methods. The slope value appears to be minimum. This inferences state that the developed NN model force the data close to the trendline at the cost of moving entire data away from unity slope value (figure 19). Among the other three methods NN in commercial method tend to perform better in terms of linearity. From figure 20 it is evident that the designed NN model can predict the pattern but experiences lots of discrepancies.



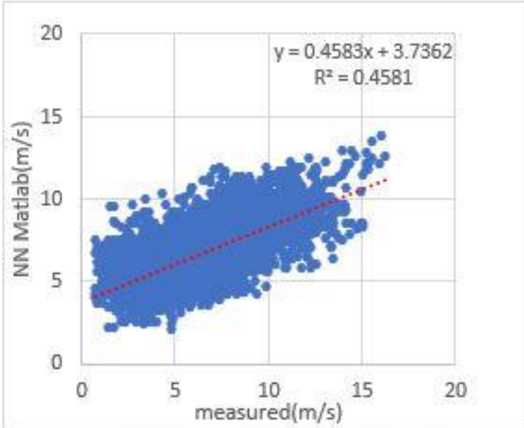
scatter plot measured vs regression



scatter plot measured vs matrix

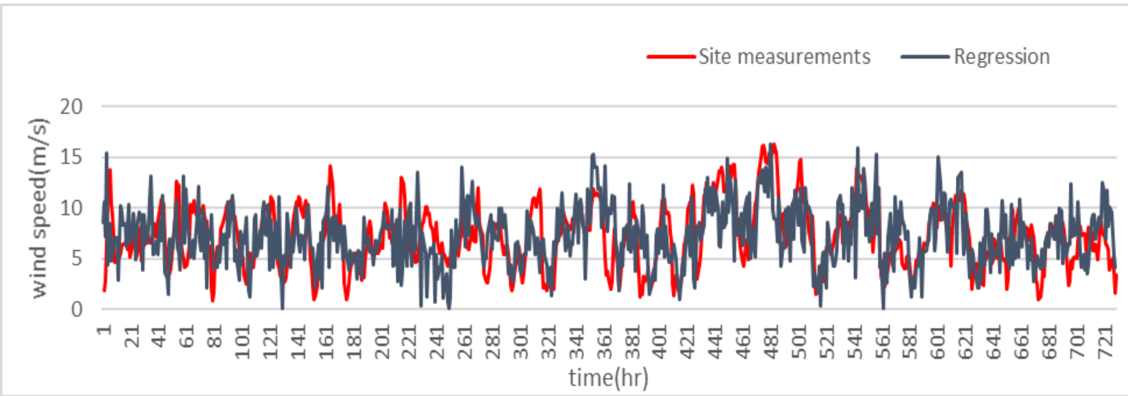


scatter plot measured vs NN windPRO



scatter plot measured vs NN MATLAB

Figure 19 – Scatter plot for wind speeds (Jordan site)



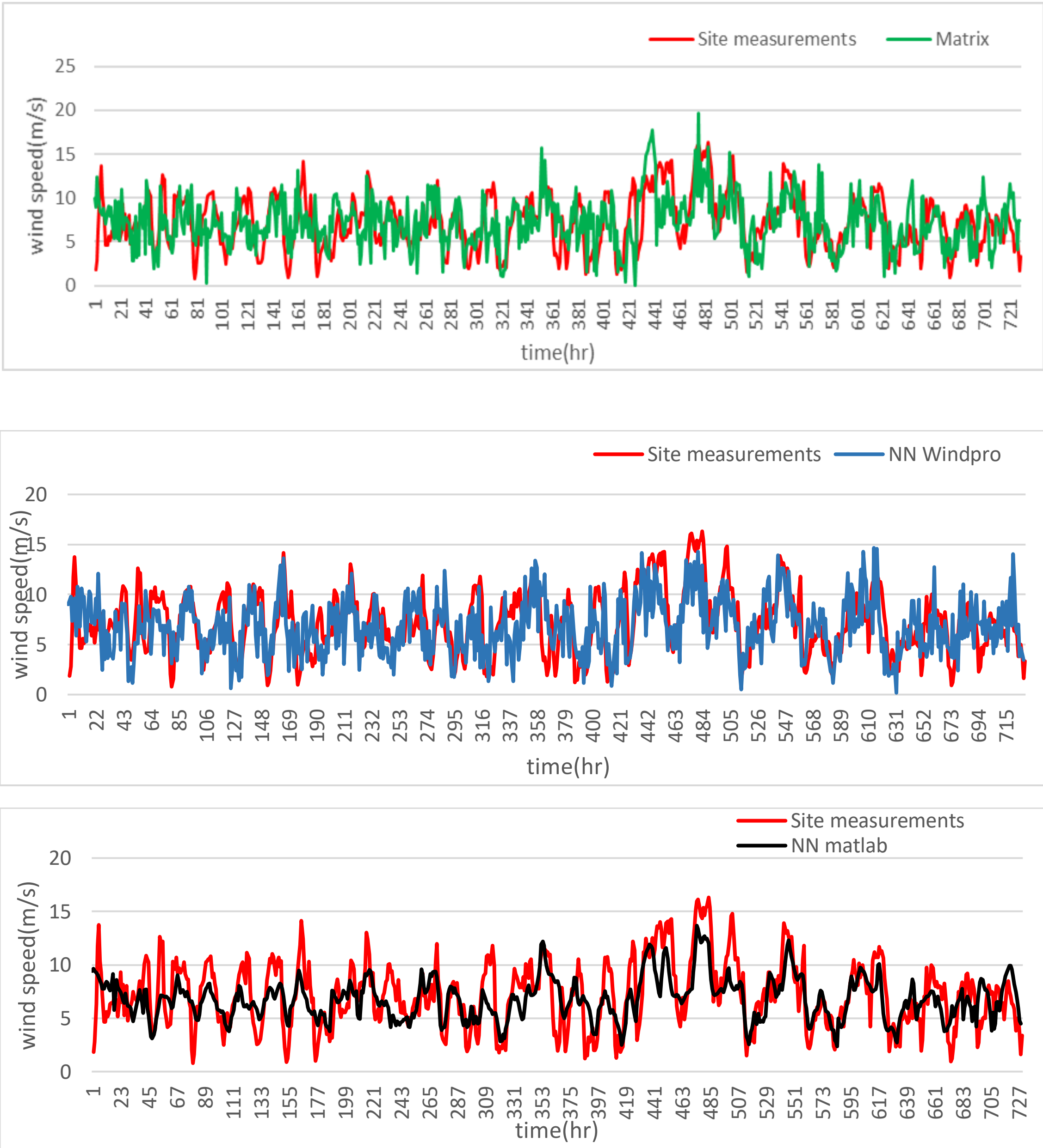
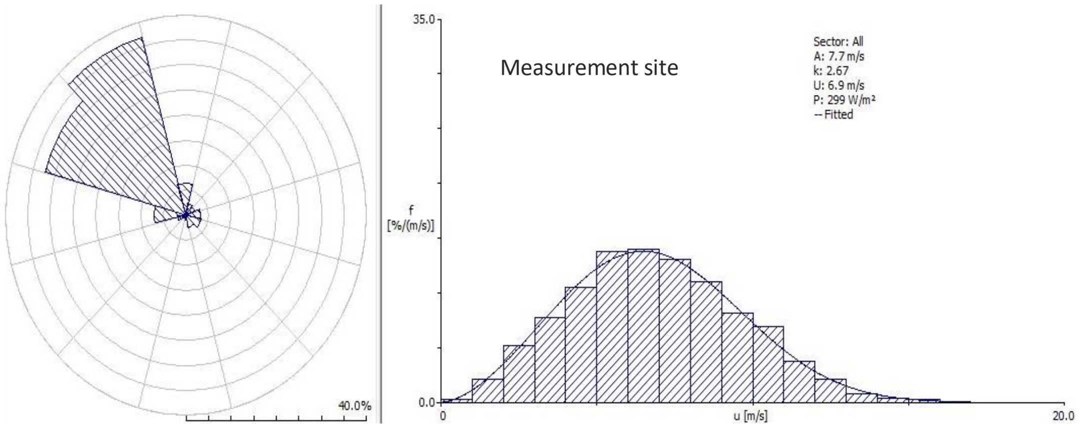
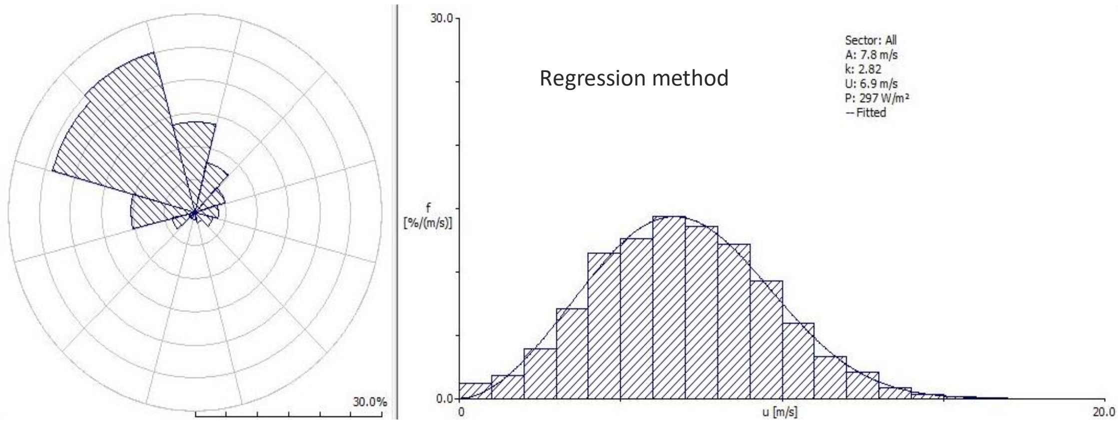


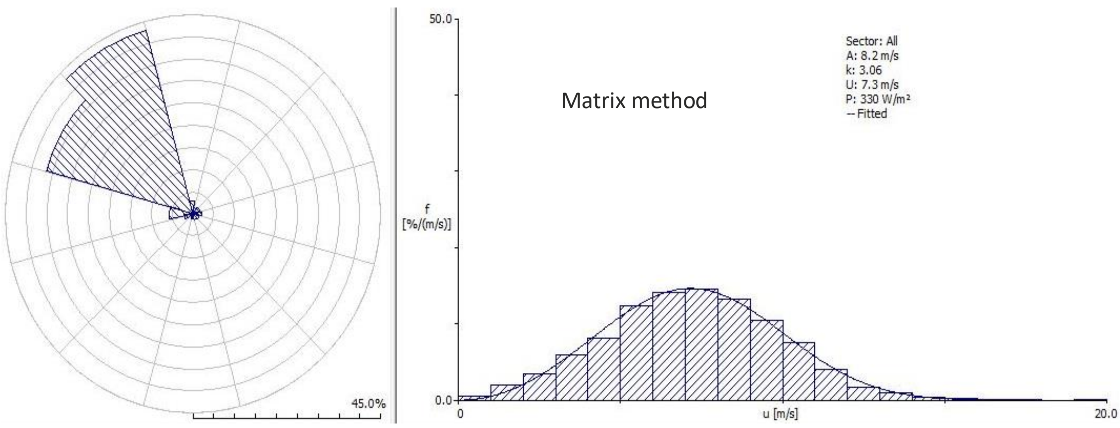
Figure 20 – Time series of with speeds comparison (Jordan site)

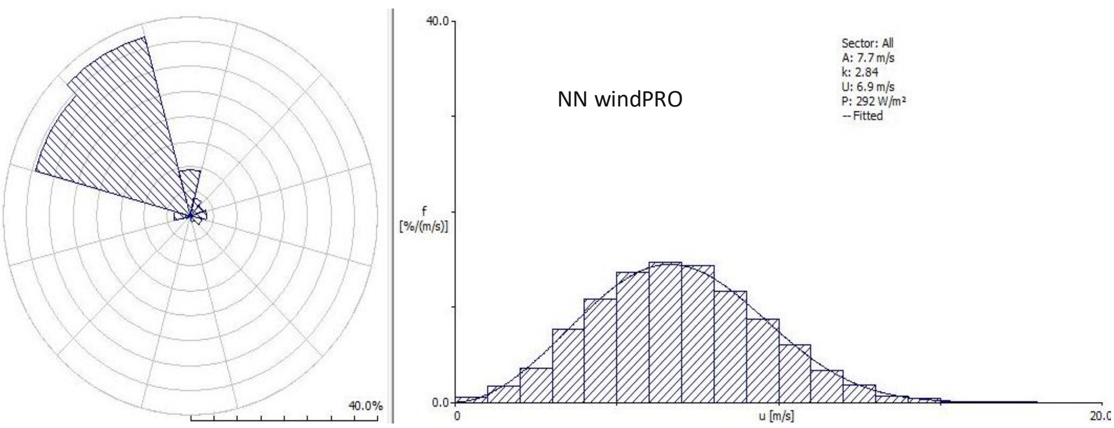
4.3.3.2 Wind direction analysis

Wind rose representation is used to analyse the wind direction predictions from different methods and measured data (figure 21). The wind speed distribution curve is also presented along with the wind rose.


40.0%
35.0
Measurement site
Sector: All
A: 7.7 m/s
k: 2.67
U: 6.9 m/s
P: 299 W/m²
-- Fitted
f
[%/(m/s)]
0.0
0
u [m/s]
20.0


30.0%
30.0
Regression method
Sector: All
A: 7.8 m/s
k: 2.82
U: 6.9 m/s
P: 297 W/m²
-- Fitted
f
[%/(m/s)]
0.0
0
u [m/s]
20.0


45.0%
50.0
Matrix method
Sector: All
A: 8.2 m/s
k: 3.06
U: 7.3 m/s
P: 330 W/m²
-- Fitted
f
[%/(m/s)]
0.0
0
u [m/s]
20.0


40.0%
40.0
NN windPRO
Sector: All
A: 7.7 m/s
k: 2.84
U: 6.9 m/s
P: 292 W/m²
-- Fitted
f
[%/(m/s)]
0.0
0
u [m/s]

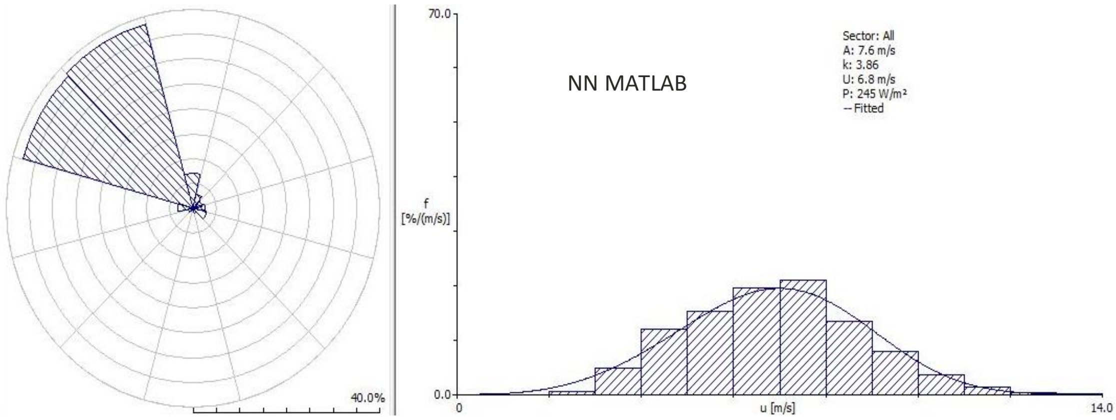


Figure 21 – Wind speeds frequency rose and histogram comparison (Jordan site)

The rose graph for measurement site indicates that for the year in consideration most of the wind is recorded in sector 5 followed by sector 11 and sector 12 respectively. (table 6)

Table 6 – Results from WAsP wind analysis software (Jordan site)

Method	Major sector(s) of wind	A(m/s)	k	Fitted mean velocity(u) (m/s) In all sectors	Available power density (W/m^2) In all sectors
Measurement data	11th,12th	7.7	2.67	6.9	299
Regression	11th,12th	7.8	2.82	6.9	297
Matrix	11th,12th	8.2	3.06	7.3	330
NN WindPRO	11th,12th	7.7	2.84	6.9	292
NN MATLAB	11th,12th	7.6	3.86	6.8	245

Although the proposed NN approach clearly estimate the major wind flow sectors. The model faces severe problem in case of wind distribution. The deviation in distribution can be rooted down to k value, k values (table 6) of matrix method and designed method that deviate more from the measurement data. The fitted mean velocity (u) of matrix method is 5.5% higher than measured value and that of developed model is 1.5% less than measured were observed. Thus, resulting in more deviation in Power density as well. In this case considered regression method performs better.

The distribution of wind speeds can be understood better from figure 22, the neglection of lower wind speeds is the reason for distorted fit.

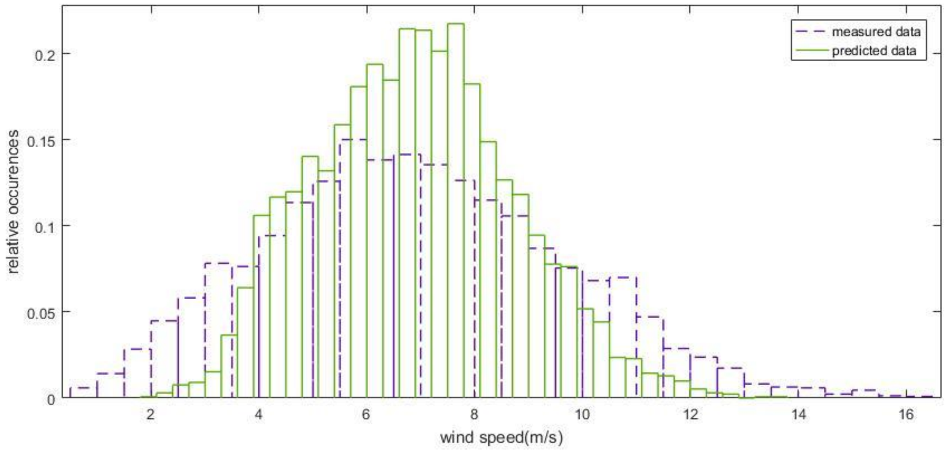


Figure 22 – Wind speed histogram comparison (Jordan site)

4.3.4 France site

The terrain of the site is characterised as flat.

4.3.4.1 Wind speed analysis

The scatter plot results for France is presented in figure 23.

France is one of the sites where every method performs equally well. The scatter plot results reveal that developed method performs better in estimating coefficient of determination. The slope values of all the methods exhibit same kind of behaviour (figure 23). Among the other three methods regression method tend to perform better in terms of linearity. The example of wind speed variations over a period of one month in mean hourly average is presented in figure 24.

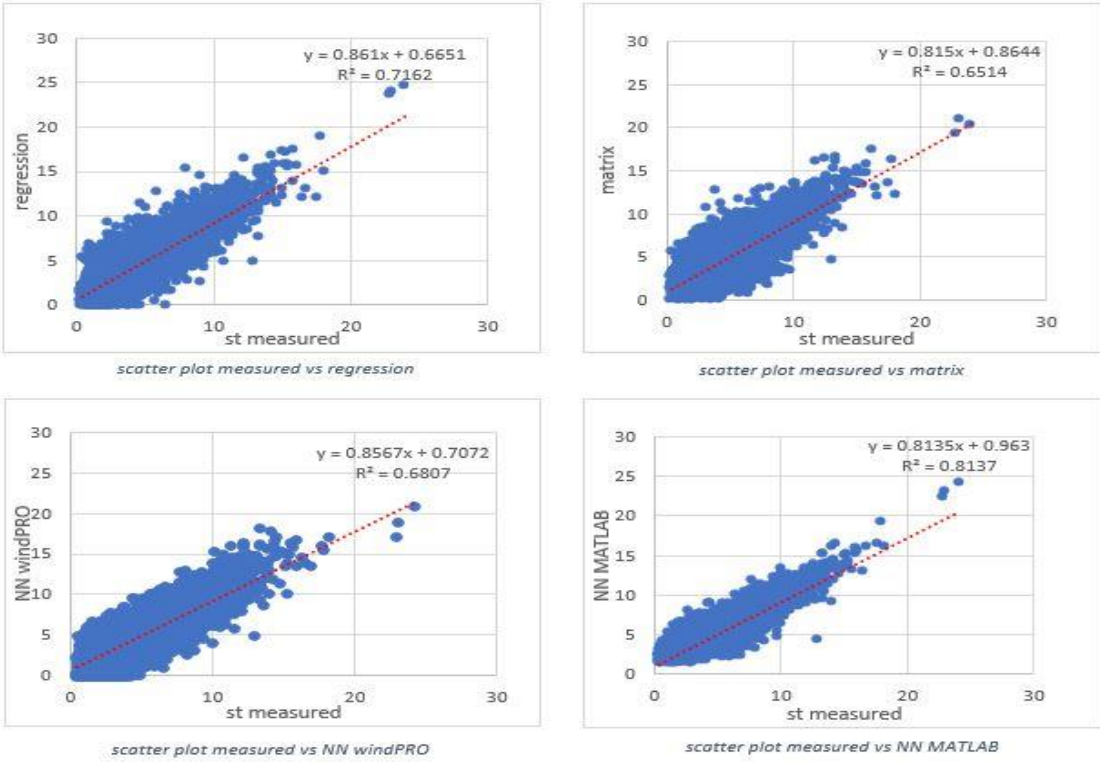
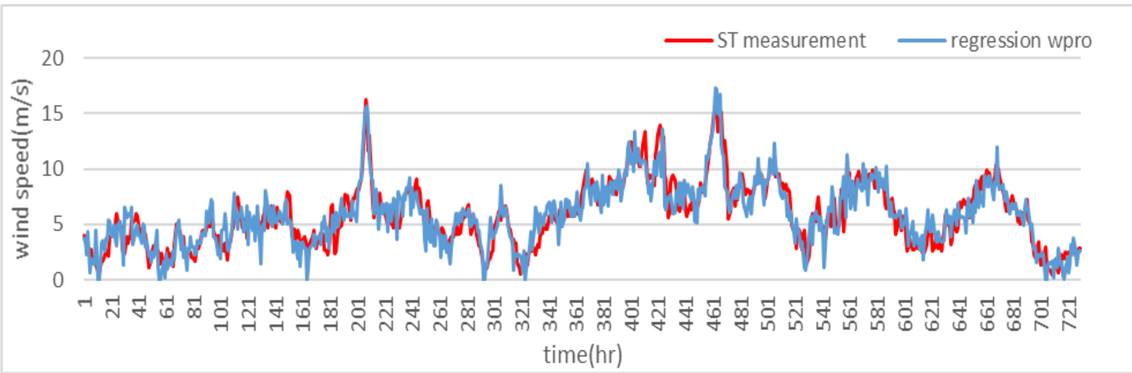
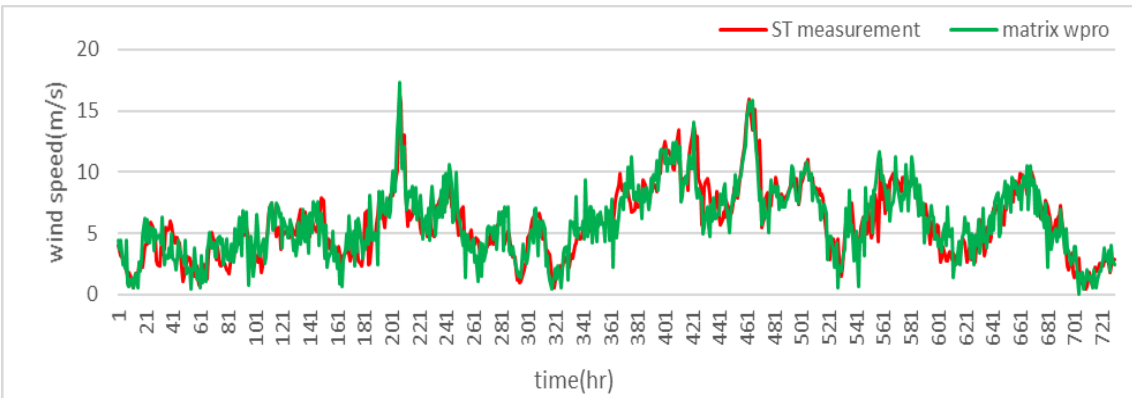


Figure 23 – Scatter plot for wind speeds (France site)





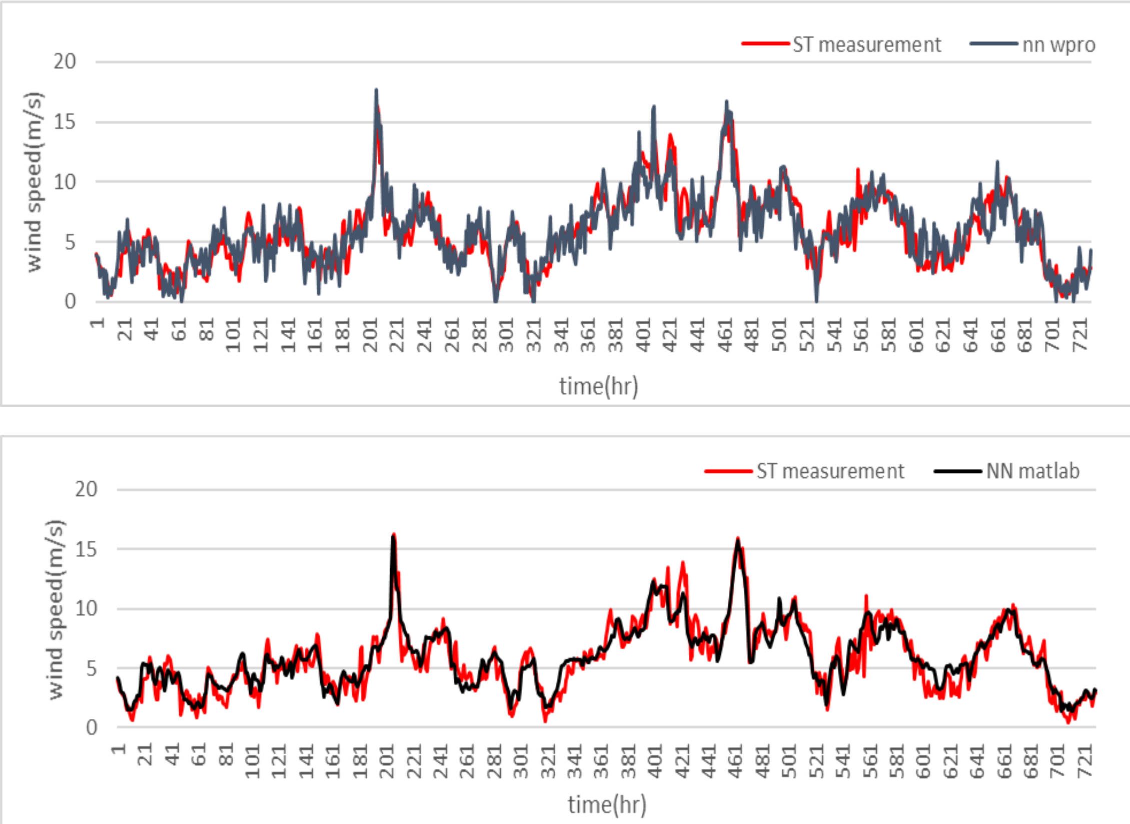
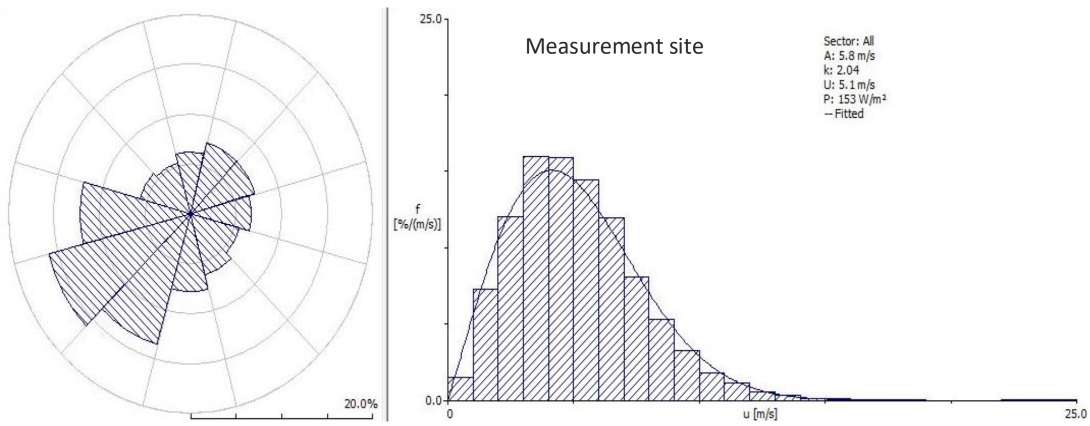


Figure 24 – Time series of with speeds comparison (France site)

4.3.4.2 Wind direction analysis

Wind rose representation is used to analyse the wind direction predictions from different methods are compared against measured data (figure 25). The wind speed distribution curve is also presented along with the wind rose.



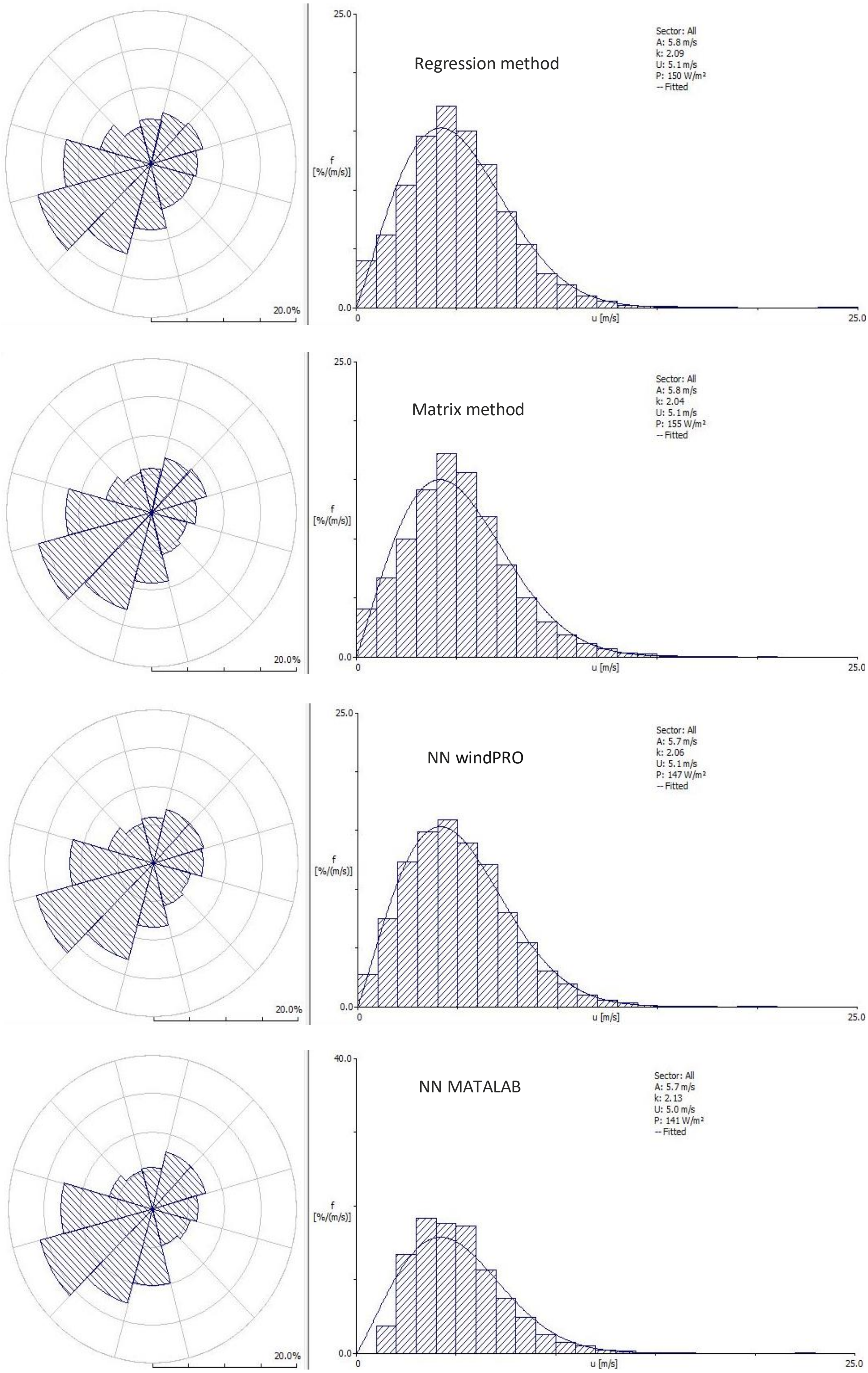


Figure 25 – Wind speeds frequency rose and histogram comparison (France site)

The rose graph for measurement site (figure 25) indicates that for the year in consideration most of the wind is recorded in sector 8 followed by sector 9 and sector 10 respectively.

Table 7 – Results from WAsP wind analysis software (France site)

Method	Major sector(s) of wind	A(m/s)	k	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m^2) In all sectors
Measurement data	8th,9th,10th	5.8	2.04	5.1	153
Regression	8th,9th,10th	5.8	2.09	5.1	150
Matrix	8th,9th,10th	5.8	2.04	5.1	155
NN WindPRO	8th,9th,10th	5.7	2.06	5.1	147
NN MATLAB	8th,9th,10th	5.7	2.13	5.0	141

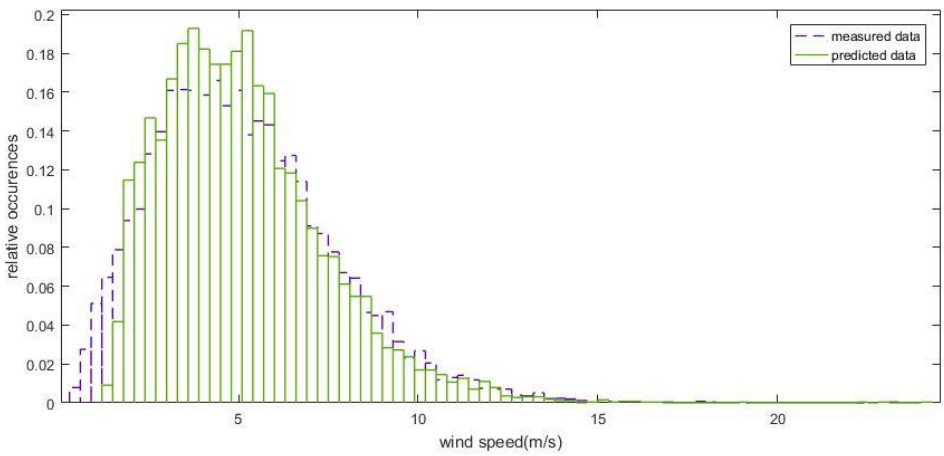


Figure 26 – Wind speed histogram comparison (France site)

The shape of distribution is directly dependant on the value of k, the designed model holds a small deviation from the measured value also the mean velocity shows 2%

deviation (table 7). The distribution of wind speeds can be understood better from figure 26, the wind distribution fit appears to be better.

4.3.5 Russia site

The terrain of the site is characterised as flat.

4.3.5.1 Wind speed analysis

The scatter plot reveals the nature and correlation between the data (figure 27). The results of the scatter plot reveal that the neural network from MATLAB method has more linear relationship in comparison with other methods. Among the other three methods regression method tend to perform better in terms of linearity. The possible reason for more linearity in developed model would be that data are more generalised this could be viewed in figure 28 as well.

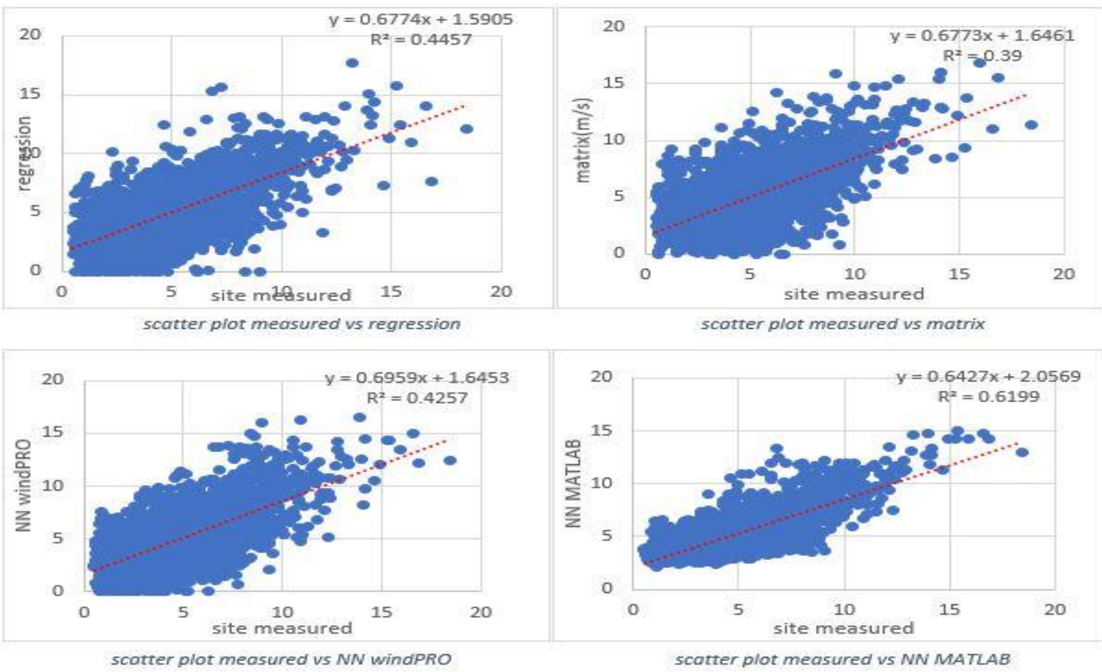


Figure 27 – Scatter plot for wind speeds (Russia site)

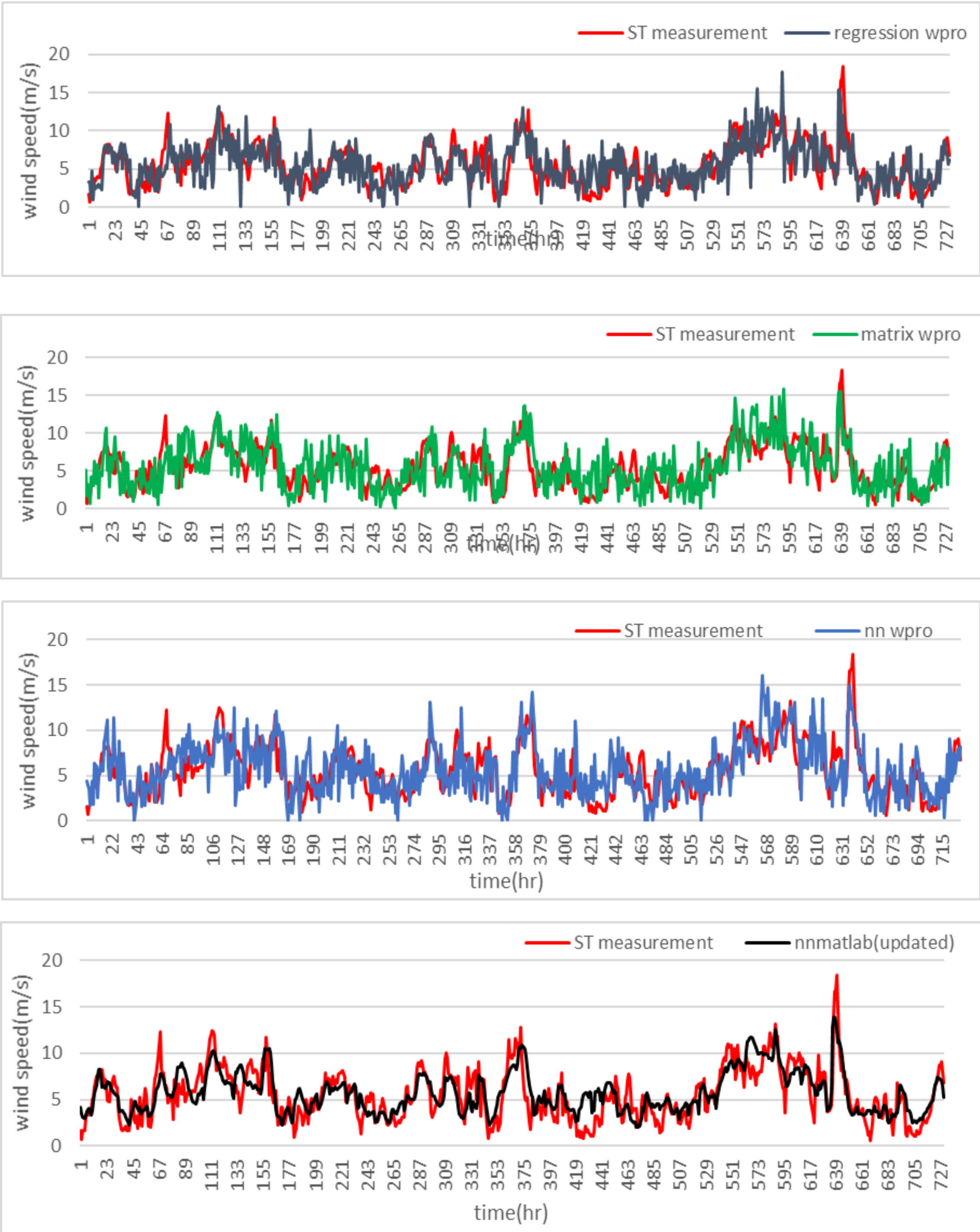
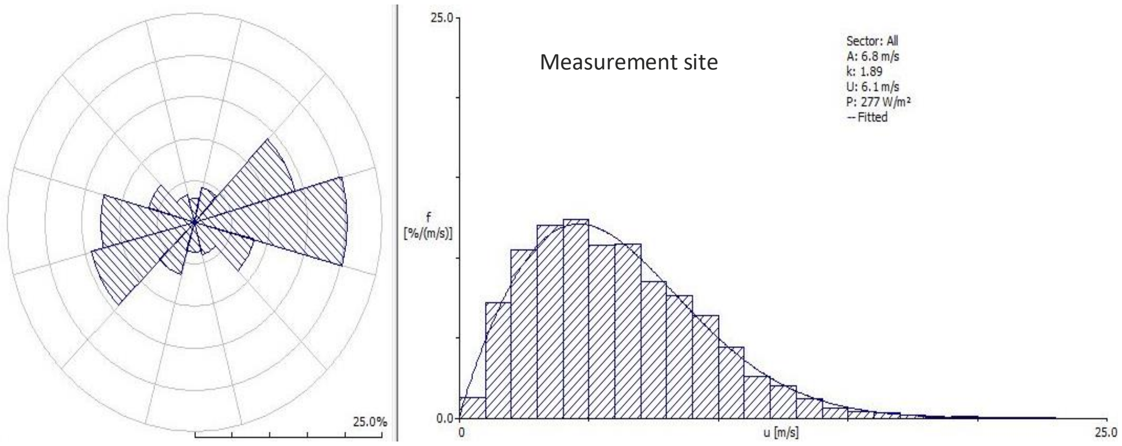
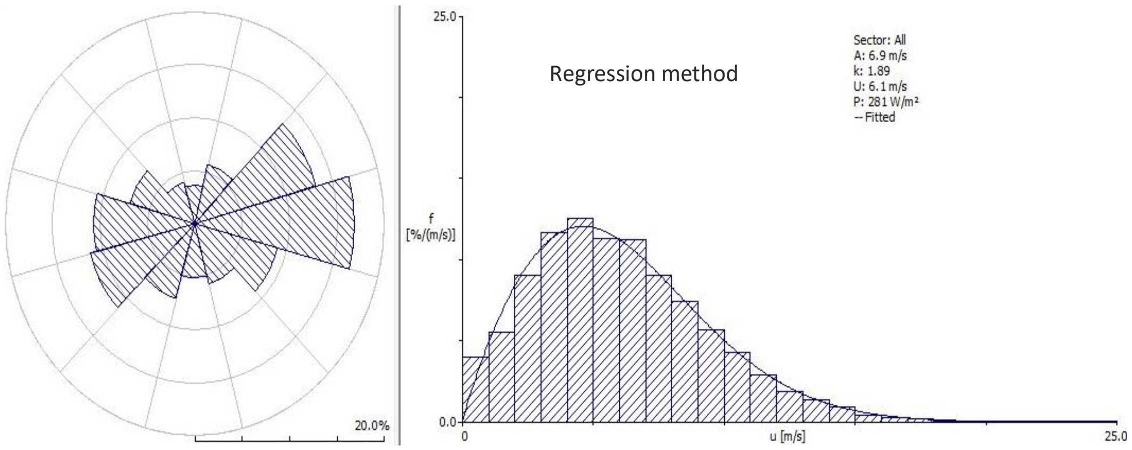


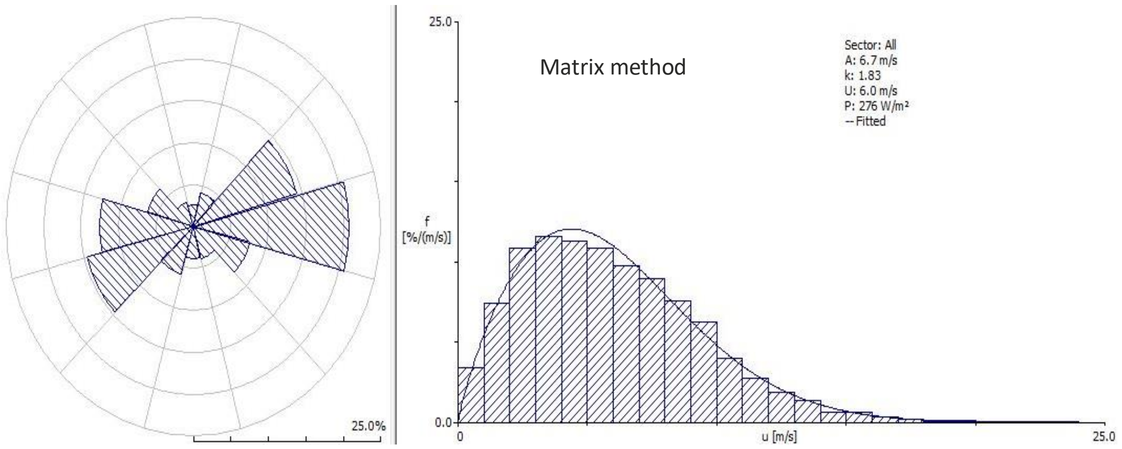
Figure 28 – Time series of with speeds comparison (Russia site)

4.3.5.2 Wind direction analysis

Wind rose representation is used to analyse the wind direction predictions from different methods and compared against measured data. The wind speed distribution curve is also presented along with the wind rose (figure 29).


Measurement site
Sector: All
A: 6.8 m/s
k: 1.89
U: 6.1 m/s
P: 277 W/m²
-- Fitted
25.0
f
[%/(m/s)]
0.0
0
u [m/s]
25.0
25.0%


Regression method
Sector: All
A: 6.9 m/s
k: 1.89
U: 6.1 m/s
P: 281 W/m²
-- Fitted
25.0
f
[%/(m/s)]
0.0
0
u [m/s]
25.0
20.0%


Matrix method
Sector: All
A: 6.7 m/s
k: 1.83
U: 6.0 m/s
P: 276 W/m²
-- Fitted
25.0
f
[%/(m/s)]
0.0
0
u [m/s]
25.0
25.0%

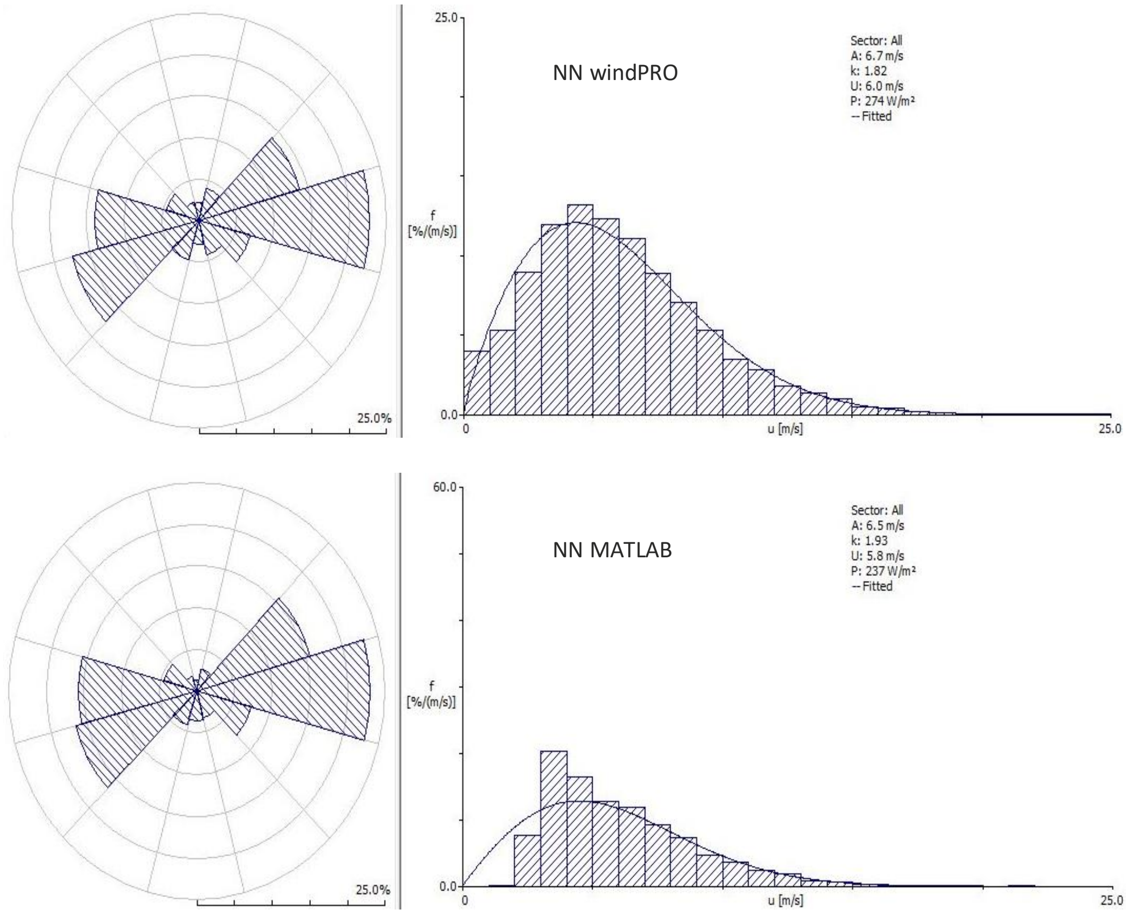


Figure 29 – Wind speeds frequency rose and histogram comparison (Russia site)

The results from wind rose graphs are help us to understand the nature of wind direction prediction of each method more elaborately. The rose graph for measurement site indicates that for the year in consideration most of the wind is recorded in sector 3 followed by sector 4, sector 9 and sector 10 (table 8).

Table 8 – Results from WAsP wind analysis software (Russia site)

Method	Major sector(s) of wind	A(m/s)	K	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m²) In all sectors
Measurement data	3rd,4th,9th,10th	6.8	1.89	5.9	277
Regression	3rd,4th,9th,10th	6.9	1.89	5.9	281
Matrix	3rd,4th,9th,10th	6.7	1.83	5.8	276
NN WindPRO	3rd,4th,9th,10th	6.7	1.82	5.8	284
NN MATLAB	3rd,4th,9th,10th	6.5	1.93	5.7	237

The wind distribution is predominantly affected by k shape parameter. The more value of k then it is more likely that wind speeds get trapped in a particular bin. Although the NN approach clearly estimate the major wind flow sectors, the fitted mean velocity also shows 3.4% deviation from measurement data.

The frequency distribution of wind speeds between measured and predicted can be understood better from figure 30.

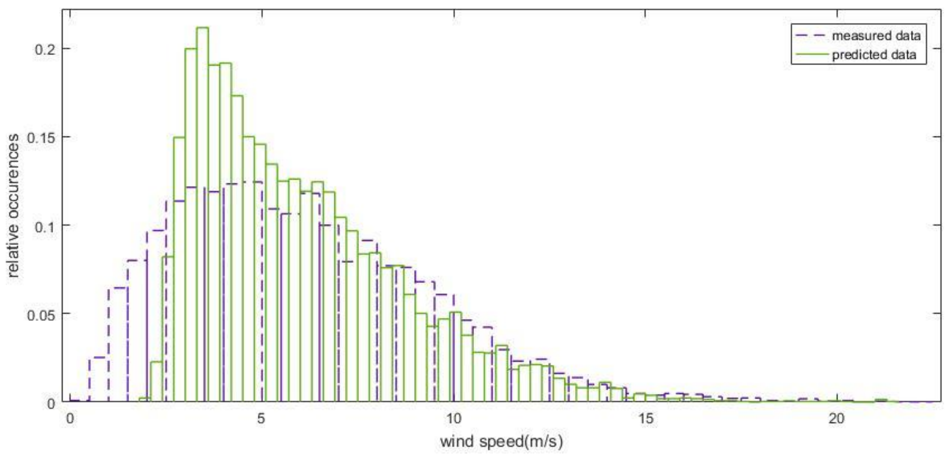


Figure 30 – Wind speed histogram comparison (Russia site)

4.4 Statistical error analysis

The statistical results for all the sites in approach 1 are presented in figure 31, 32, 33 in the order of root mean square error (RMSE), mean absolute percentage error (MAPE), index of agreement (IoA).

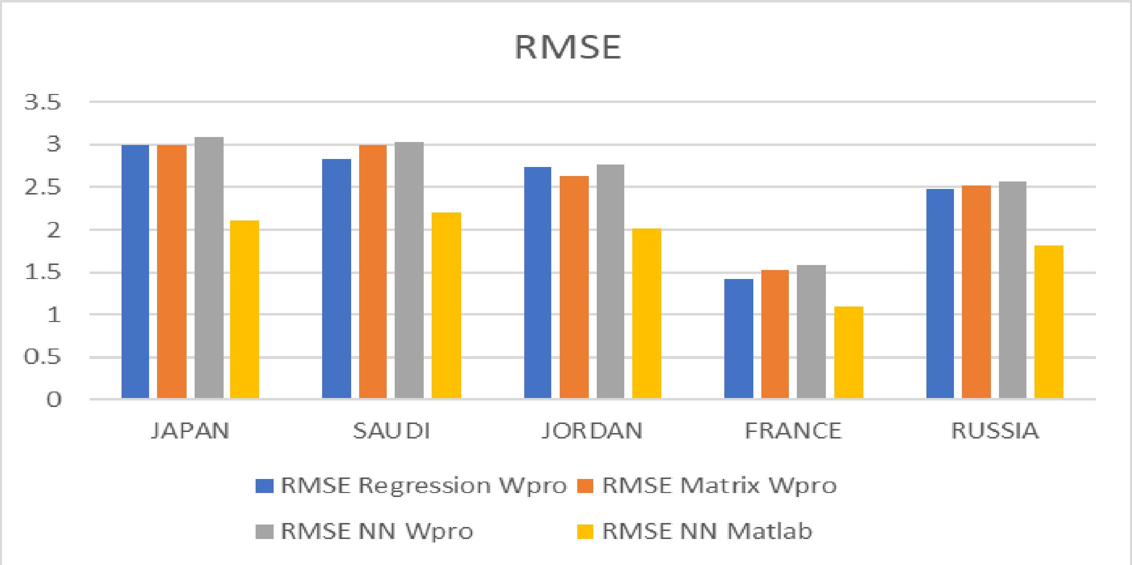


Figure 31 – Different methods RMSE results for all sites

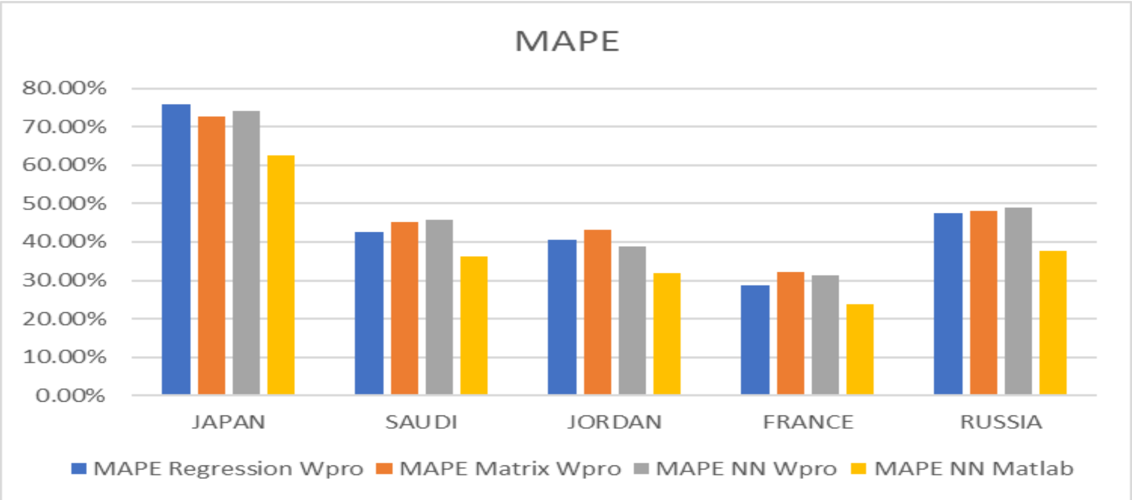


Figure 32 – Different methods MAPE results for all sites

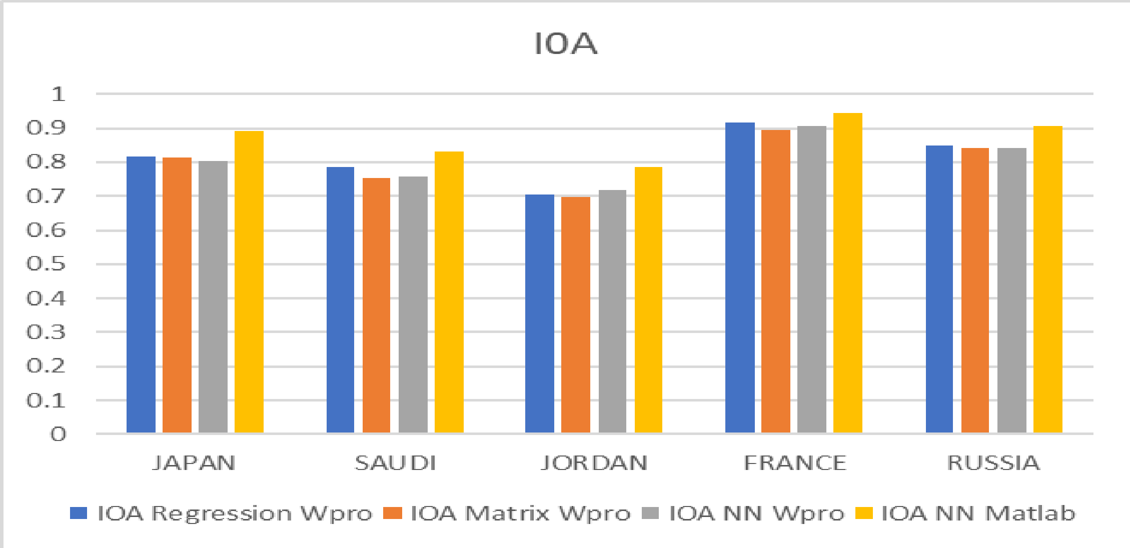


Figure 33 – Different methods IoA results for all sites

The statistical error values were calculated in accordance with equations (8), (9), (10) respectively and power density was estimated in reference with equation (11). Based on RMSE analysis for all the sites proposed model estimates the lowest error and this pattern is followed in MAPE error as well, the minimum value of MAPE is found to be 23.7% in France and 62.5% in case of Japan. The IoA parameter defines how well the predicted value agrees with the actual value in the scale of 0 to 1 while 0 indicating no agreement and 1 indicating maximum agreement. In all the cases IoA value estimated from designed model comes out better.

As far as statistical errors are considered in all the sites neural network approach from MATLAB works better, this is also evident from the time series graphs and scatter plot graphs in all sites. It is possible to see from the hourly time series graphs that MCP methods of industrial software develops number of peaks and gorges around the measured data thus resulting in more error values. Although designed model can catch all the variations, the problem arises in calibrating the extreme values of wind speeds both minimum and maximum. The possible reason for this kind of behaviour could be explained as more generalisation of data. The occurrence of minimum and maximum wind speeds does not occur in cluster this might make the network unable to catch those values. Further investigation should be made on this behaviour of the network. when considering wind directions, the network seems to predict all the major sectors with higher frequencies of wind very promisingly. In all the sites it was noticed that the power density estimated was under predicted, ranging from minimum of 7.9% in France to 18.1% in Jordan (table 9). The correlation (r) between reference site data and short-term measurement data has a great impact in the performance of the NN MATLAB model. (table 9)

Table 9 – Correlation coefficient between reference and measured site data and Power density deviations

Sites	correlation coefficient(r)	Deviation in PD estimate (regression method in %)	Deviation in PD estimate (matrix method in %)	Deviation in PD estimate (NN industrial in %)	Deviation in PD estimate (NN MATLAB in %)
Japan	0.74	+1.5	-3.3	+0.8	-17
Saudi Arabia	0.71	No deviation	-2.3	No deviation	-17.7
Jordan	0.63	-0.7	+9.4	-1.7	-18.1
France	0.87	-2	-4	+1.3	-7.9
Russia	0.82	+1.5	-0.4	-1.1	-14.4

* PD – Power Density

The order of under prediction in Power density estimate follows the same patter as the r between the reference and target sites. The performance of the network was very poor in case of Jordan where only six months of concurrent data was available and usually this is not the case because MCP methodologies basically built on concurrent period of at-least one year. The statistical values of mean wind speeds from the dataset is provided in figure 34, in all the cases designed NN model performs equally better in predicting the mean wind speeds.

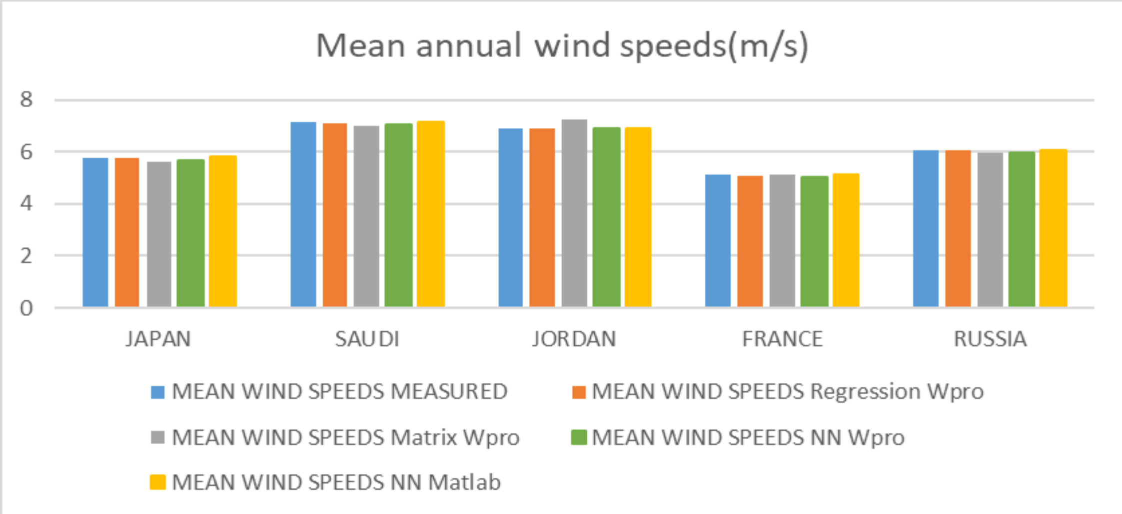


Figure 34 – Annual mean wind speeds estimations results for all sites

Weibull parameters (a, k) play an important role in estimating Power density content along with fitted wind speed (u). The inability of developed NN model to capture the minimum wind speeds resulted in deteriorated fitted mean wind speed disturbing the frequency distribution thus experiencing deviations in power density estimations. The statistical results cannot stand alone and prove themselves, the statistical results should be supported by proper distribution of wind speed in all ranges. As far as wind resource estimation is considered a model should be able to perform well on both wind speed and wind direction simultaneously, in our case the overall wind speed predictions are better but distribution is poorer in lower wind speeds and when it comes to wind directions the model is performing equally better. Thus, an improvisation should be made on the built model to capture the lower wind speeds.

As all the models available in industrial software perform similarly in case of wind distribution, wind directions, and deviation in power density, the only deciding factor could be considered is statistical error vales of wind speeds. In comparison among the methods, based on statistical error values regression method, matrix method, neural network approach tends to perform similarly with small deviations out of which regression tends to perform better in all sites.

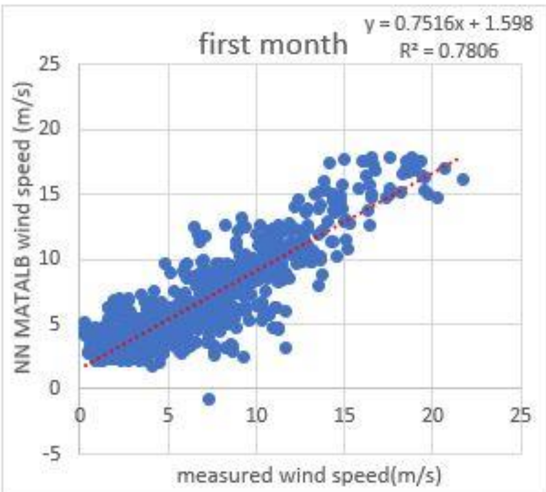
4.5 Approach 2 analysis

The analysis of approach 2 is made only for Neural network MATLAB as the results from the industrial software were the same as discussed in approach 1. The statistical error values and Power density estimate are calculated only for the month that is used in testing periods.

4.5.1 Japan site

4.5.1.1 Wind speed analysis

The scatter plot for first and last month of japan is presented in the figure 35. The wind speed data seems to be more dispersed in case of last month. The slope values of first month indicate that the data points dissipate linear relationship but slope value from last month scatter plot defines that data points drift away from ideal slope value. As an example, the time series analysis for wind speeds for first and last month compared against measured data is provided in figure 36.



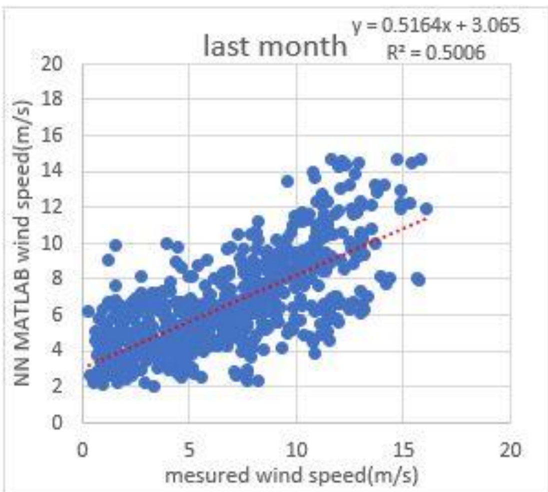


Figure 35 – Scatter plot for first and last month wind speeds (Japan site)

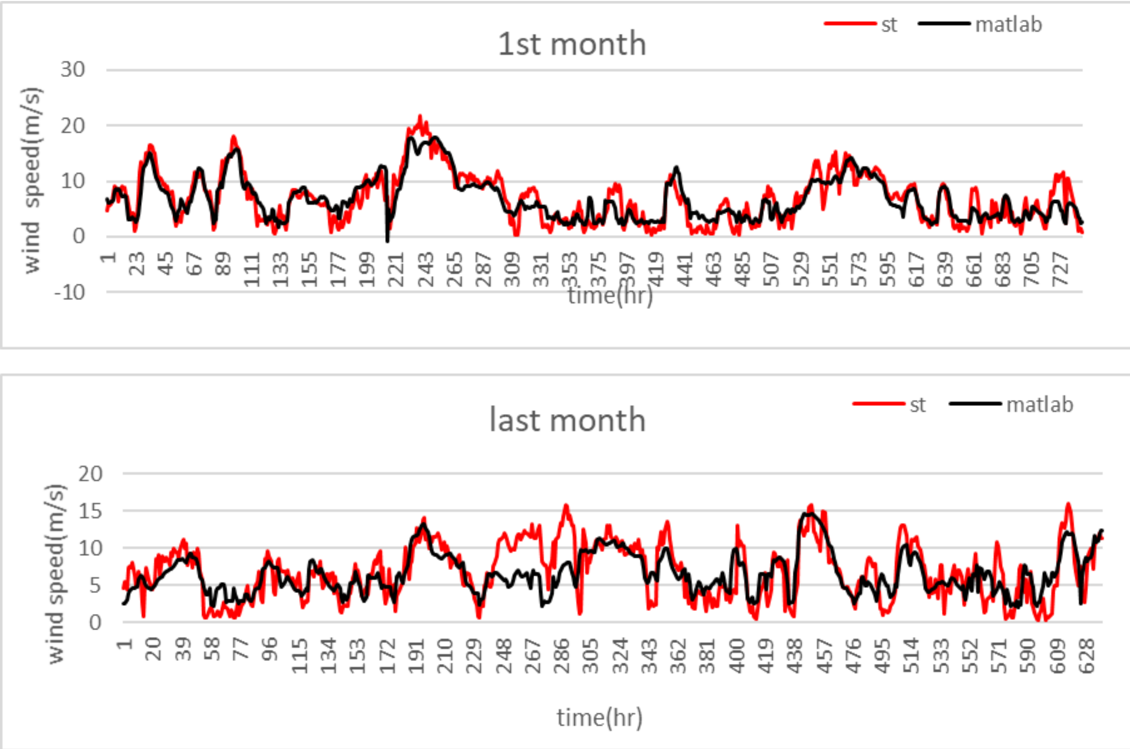
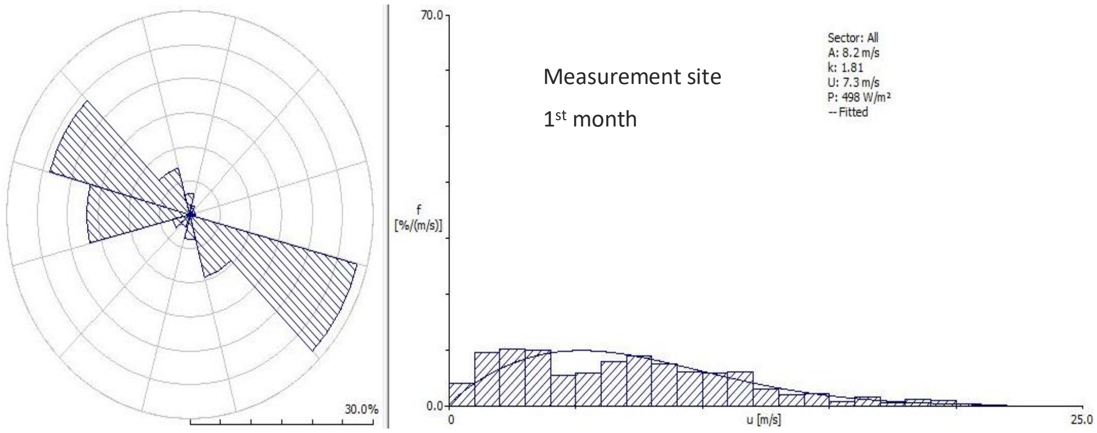
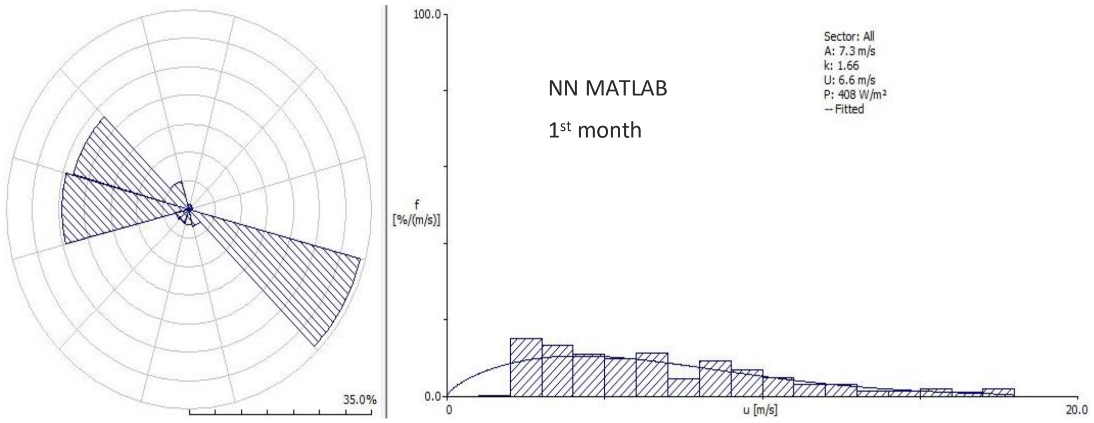


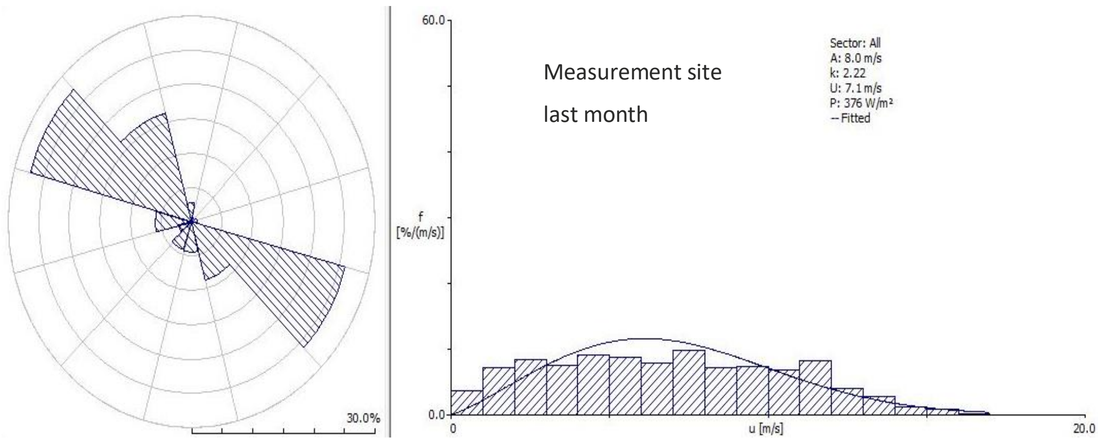
Figure 36 – Time series of first and last month with speeds comparison (Japan site)

4.5.1.2 Wind direction analysis

The study of wind rose graph in figure indicates that most of the occurrences of wind are recorded in sector 5, 10, 11 for first month and for last month most of the occurrences take place in sector 5, 11 and 12 (figure 37).


Measurement site
1st month
Sector: All
A: 8.2 m/s
k: 1.81
U: 7.3 m/s
P: 498 W/m²
-- Fitted
70.0
0.0
f [%/(m/s)]
0
u [m/s]
25.0
30.0%


NN MATLAB
1st month
Sector: All
A: 7.3 m/s
k: 1.66
U: 6.6 m/s
P: 408 W/m²
-- Fitted
100.0
0.0
f [%/(m/s)]
0
u [m/s]
20.0
35.0%


Measurement site
last month
Sector: All
A: 8.0 m/s
k: 2.22
U: 7.1 m/s
P: 376 W/m²
-- Fitted
60.0
0.0
f [%/(m/s)]
0
u [m/s]
20.0
30.0%

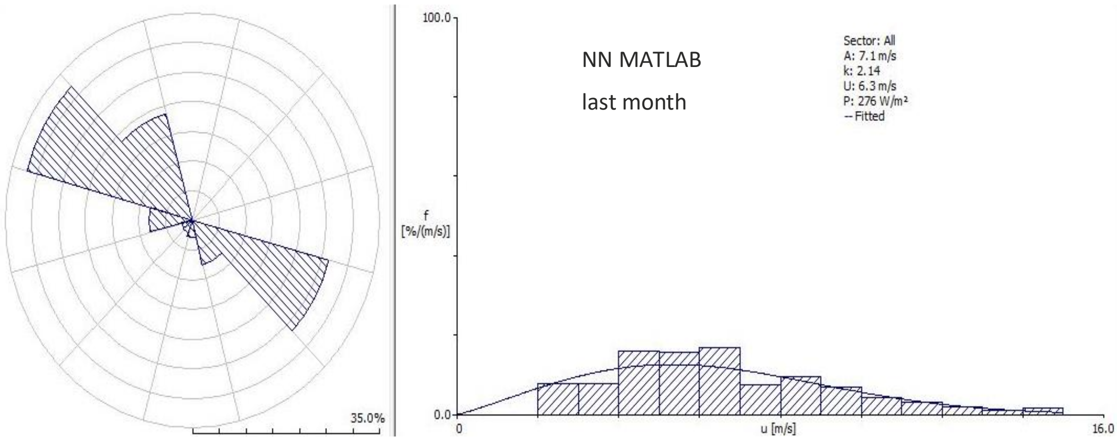


Figure 37 – Wind speeds frequency rose and histogram comparison, for the first and last month (Japan site)

As far as last month wind rose graph is considered the directional sectors are predicted well but the same problem of discarding the minimum wind speeds continues resulting in distorted fit and deteriorated power density estimate. The A, k parameters and fitted mean velocity are presented in the table10, in the case of first and last month the k parameter is less than its measured counterpart and mean velocity also faces a depreciation of 9.6% for first month and 11.3% for last month. The comparison of wind speed frequency for measured and predicted dataseries for first and last month are presented in figure 38 and 39 respectively.

Table 10 – Results from WAsP wind analysis software approach 2 (japan site)

Method	Major sector(s) of wind	A(m/s)	K	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m^2) In all sectors
Measurement data 1st month	5th,10th,11th	8.2	1.81	7.3	498
NN MATLAB 1st month	5th,10th,11th	7.3	1.66	6.6	408
Measurement data last month	5th,11th,12th	8.0	2.22	7.1	376
NN MATLAB last month	5th,11th,12th	7.1	2.14	6.3	276

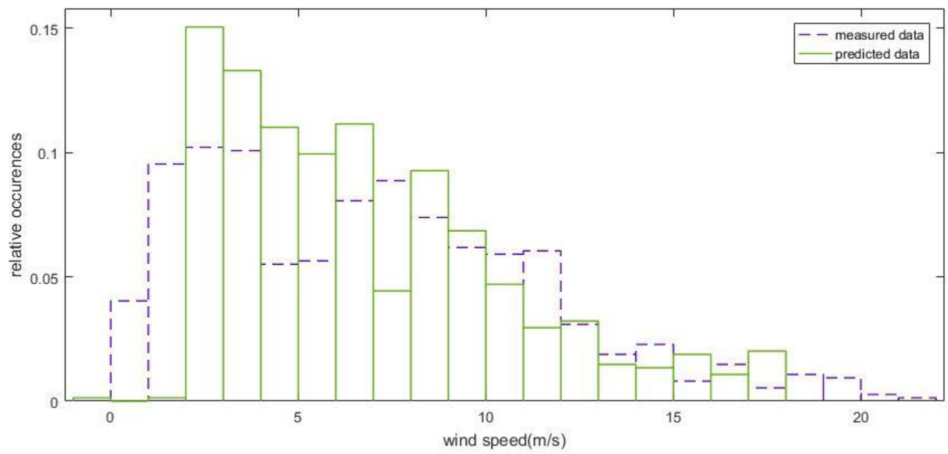


Figure 38 – Wind speed histogram comparison, for first month (Japan site)

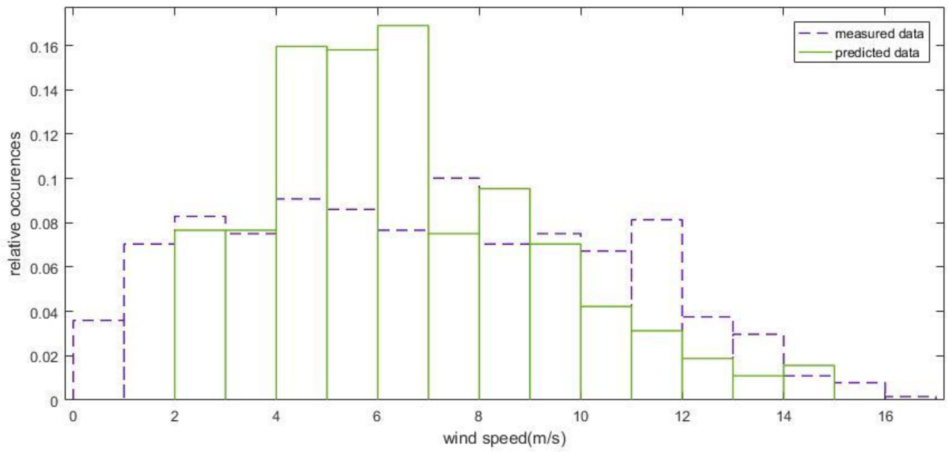


Figure 39 – Wind speed histogram comparison, for last month (Japan site)

4.5.2 Saudi Arabia site

The concurrent data period for Saudi is about one year 10 months and 28 days. The data period of last 1 year and 28 days was used to train and build the model. The first unused 10-month data was used to test the model. This is one of the interesting sites where one-year data is used to build the model so that the functional relationship is established all around the year covering all the seasons and variations in wind characteristics.

4.5.2.1 Wind speed analysis

The scatter plot determines the nature of dispersion of data around the trend line (fig 40). The inferences from the figure state that predicted datapoints converge linearly well with the measured data but from measure of slope readings, there is an indication that datasets deviate from the ideal slope line (m=1).

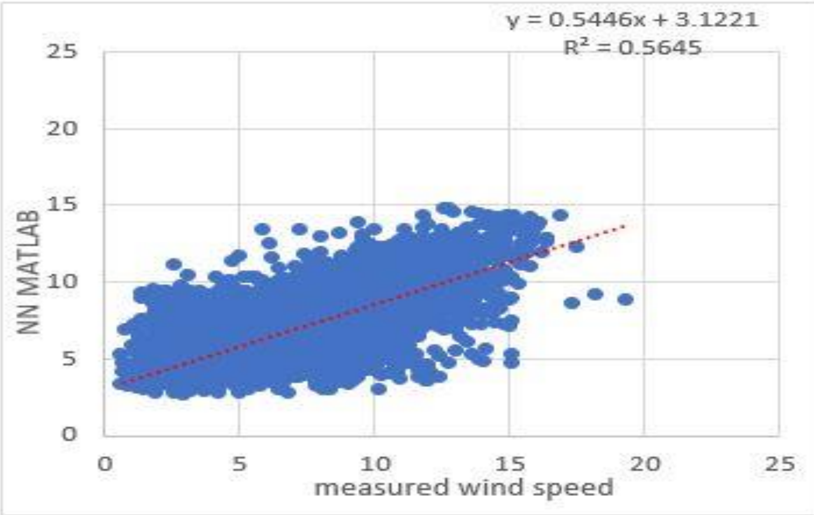


Figure 40 – Scatter plot for 10 months wind speeds (Saudi site)

To support the scatter plot results the time series analysis between measured and predicted wind speeds is provided in figure 41.

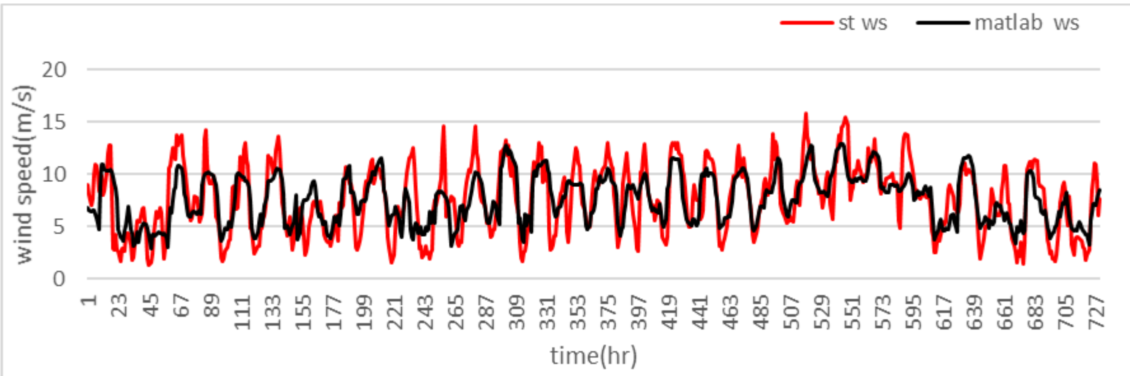


Figure 41 – Time series of 10 months wind speeds comparison (Saudi site)

4.5.2.2 Wind direction analysis

The wind rose representations for measured and approach 2 predicted is presented in the figure 42.

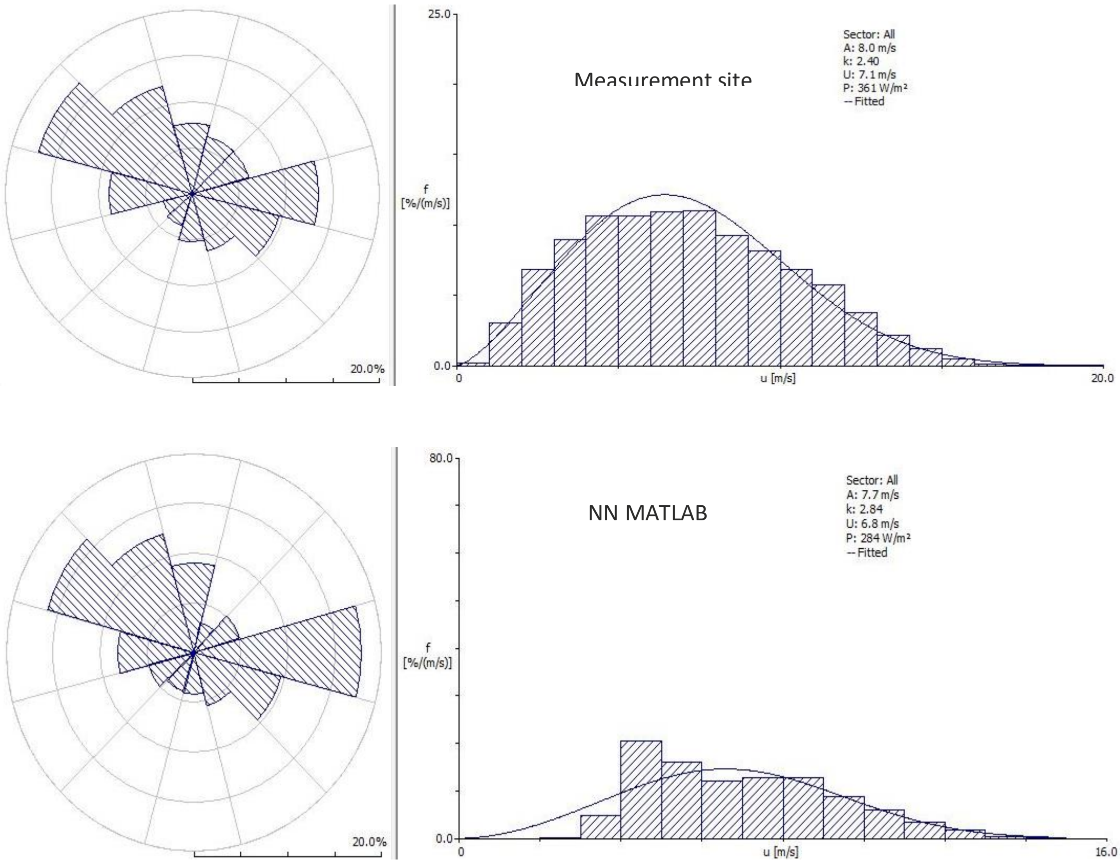


Figure 42 – Wind speeds frequency rose and histogram comparison approach 2 (Saudi site)

The results of prediction of major wind flow sectors is satisfactory. The value of k the shape parameter plays a vital role in shape of wind distribution. The lower values of k indicate that the distribution will be wide spread whereas the higher values of k indicates that more wind speeds are binned in narrow range. The k value for measured site is 2.4 and for predicted data is 2.84. The fitted mean wind speed suffers a depreciation of 4.2% from the measured data (table 11).

Table 11 – Results from WAsP wind analysis software approach 2 (Saudi site)

Method	Major sector(s) of wind	A(m/s)	k	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m^2) In all sectors
Measurement data	4th,11th,12th	8.0	2.4	7.1	361
Designed NN	4th,11th,12th	7.7	2.84	6.8	284

The comparison of wind speed frequency for measured and predicted dataseries is presented in figure 43.

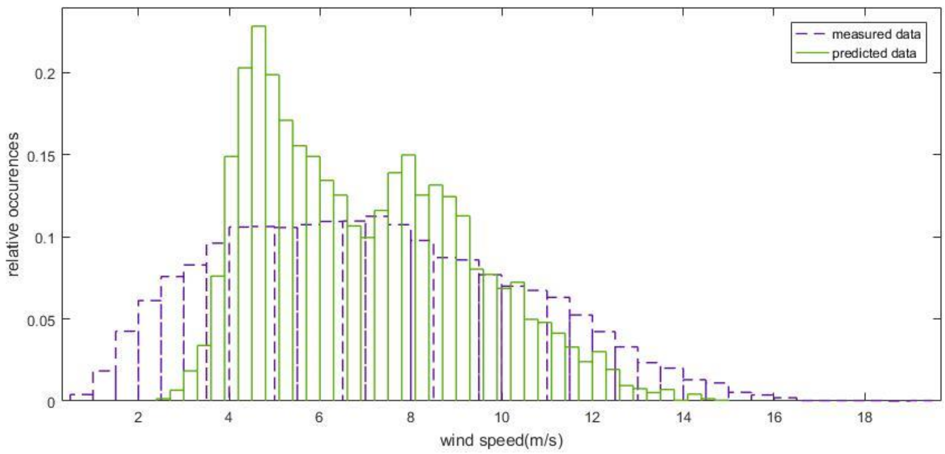


Figure 43 – Wind speed histogram comparison approach 2 (Saudi site)

4.5.3 France

One-year concurrent data was available for France. The last consecutive ten-month data was used to train the network and first two months data was used to test the model performance.

4.5.3.1 Wind speed analysis

The scatter plot for measured wind speed and predicted wind speed in presented in figure 44.

The slope value and scatter of data points along the trendline appears to be promising. The comparison of measured and predicted mean hourly wind speeds for the first month is presented in figure 45.

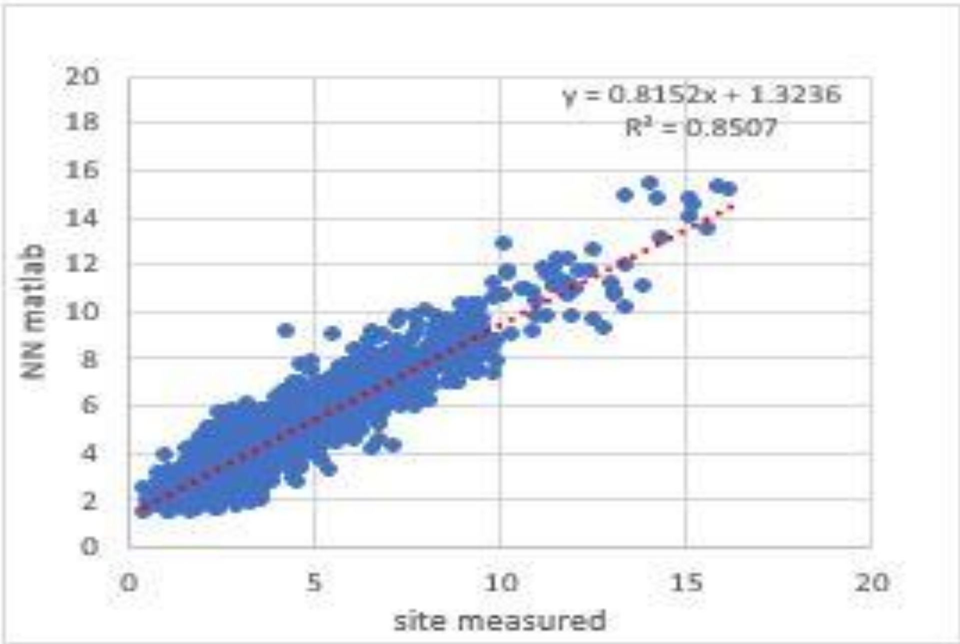


Figure 44 – Scatter plot for wind speeds approach 2 (France site)

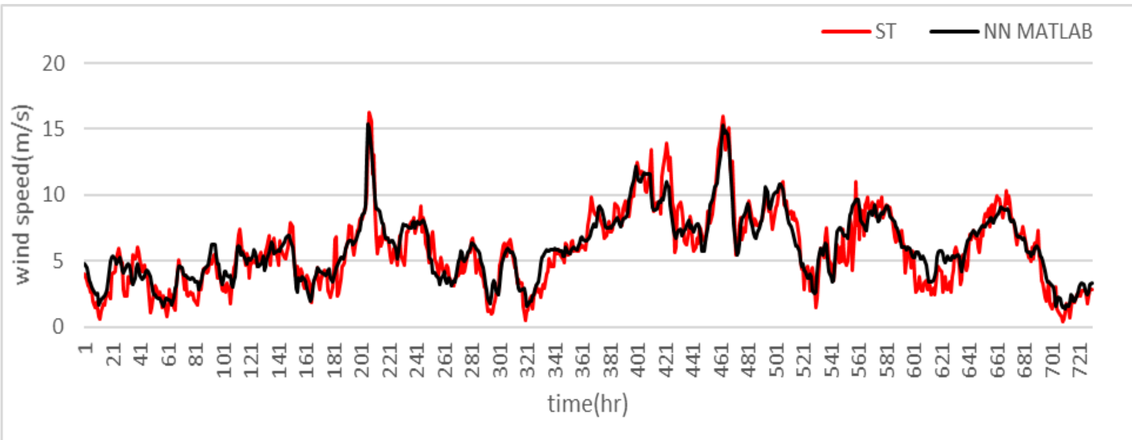


Figure 45 – Time series of wind speeds comparison approach 2 (France site)

4.5.3.2 Wind direction analysis

The approach 2 seems to perform well in prediction of major frequencies of wind sector. The distribution can be analysed with A and k parameters. The k parameter for predicted data series is high compared to the measured data series (table 12).

The fitted mean wind speed (u) of predicted data series experiences a positive deviation of 9.5% from the measured data.

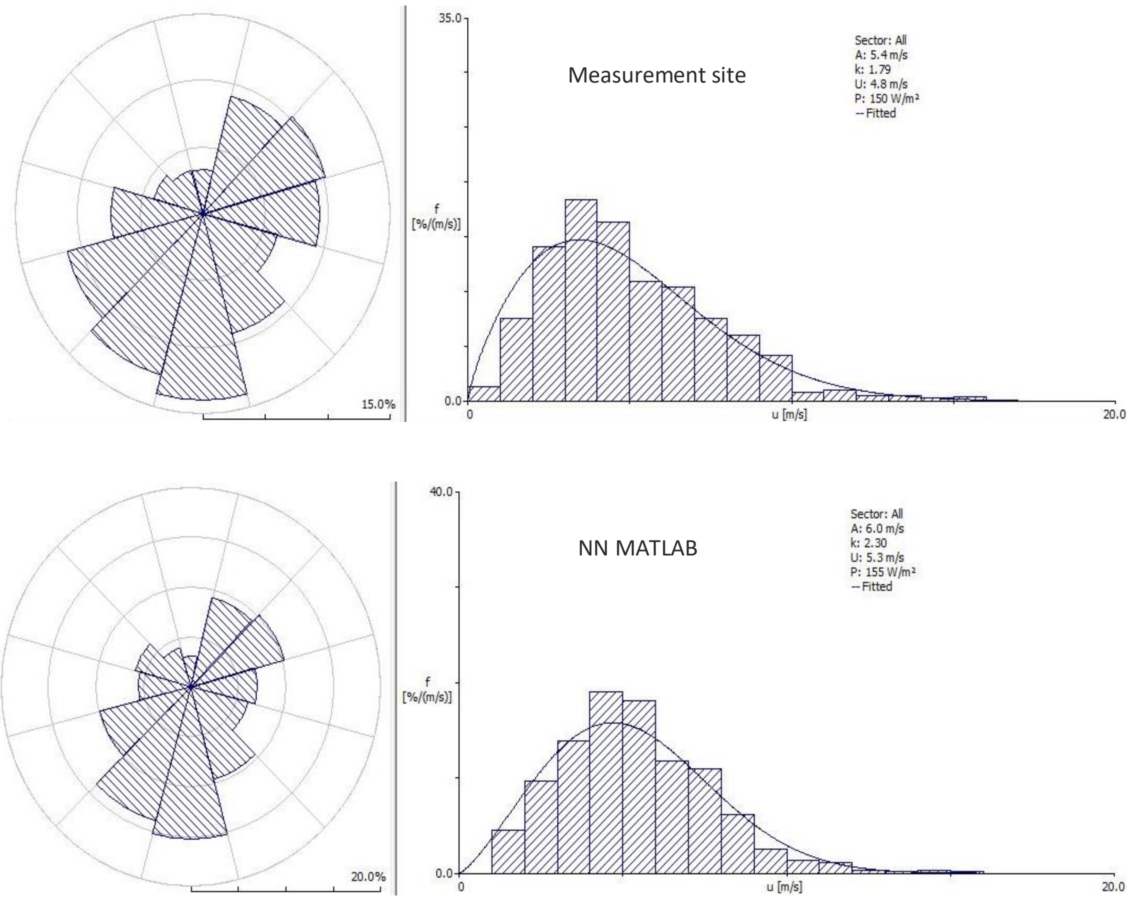


Figure 46 – Wind speeds frequency rose and histogram comparison approach 2 (France site)

Table 12 – Results from WAsP wind analysis software approach 2 (France site)

Method	Major sector(s) of wind	A(m/s)	K	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m²) In all sectors
Measurement data	2nd,3rd,6th,7th,8th,9th	5.4	1.79	4.8	150
NN MATLAB	2nd,3rd,6th,7th,8th,9th	6.0	2.3	5.3	155

The distributional characteristic of wind speed frequencies can be clearly understood from the figure 47.

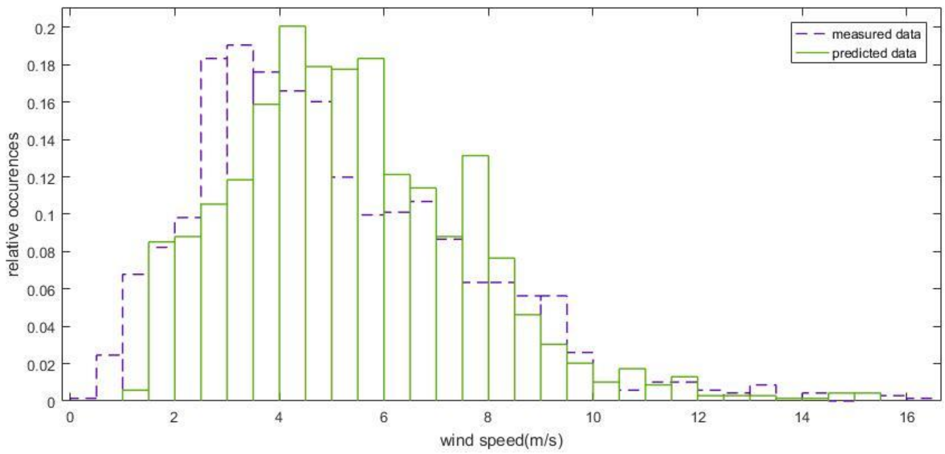


Figure 47 – Wind speed histogram comparison approach 2 (France site)

4.5.4 Russia

One-year concurrent wind data was available for Russia. The last nine months data were used to train the network and the established functional relationship was applied to first three months of reference data resulting in new predicted three-month data which was compared against the actual three-month data to evaluate the performance of the built model.

4.5.4.1 *Wind speed analysis*

The scatter plot for measured wind speed and predicted wind speed are provided in figure 48. The predicted data series seems to converge linearly well with the measured data series, but it could be viewed that minimal range of wind speeds are left over.

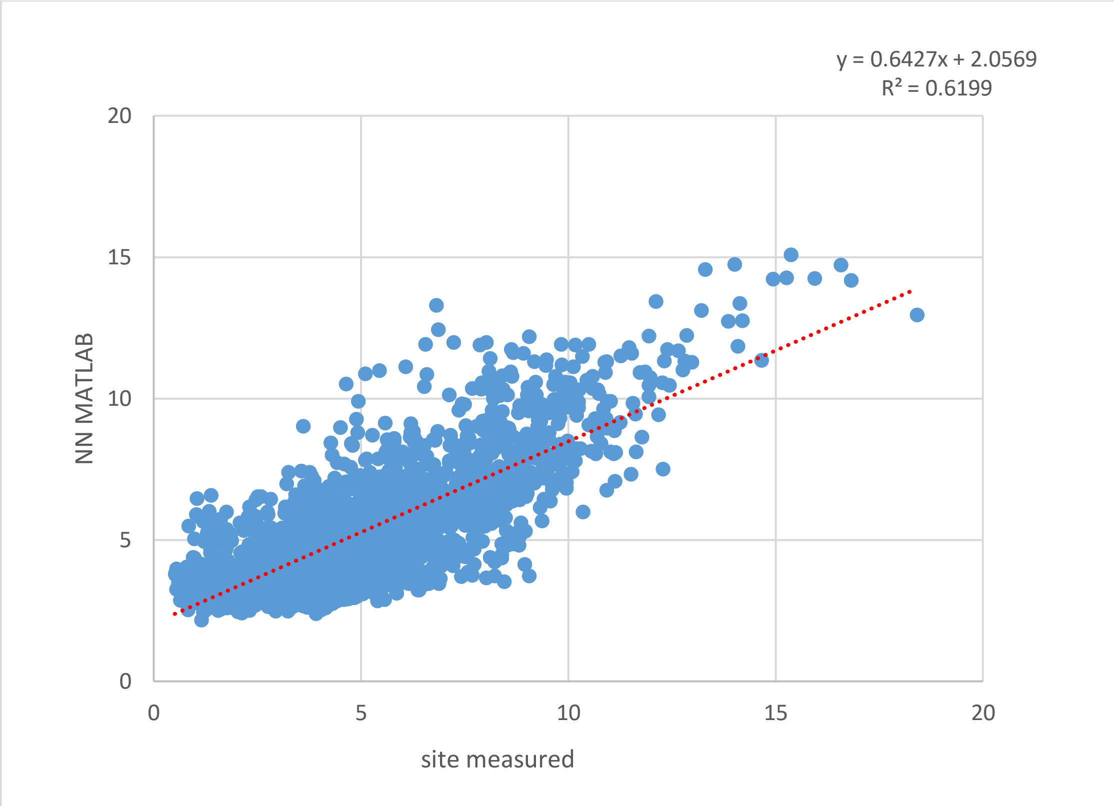


Figure 48 – Scatter plot for wind speeds approach 2 (Russia site)

The time series comparison of wind speed predicted and measured dataseries can be viewed figure 49.

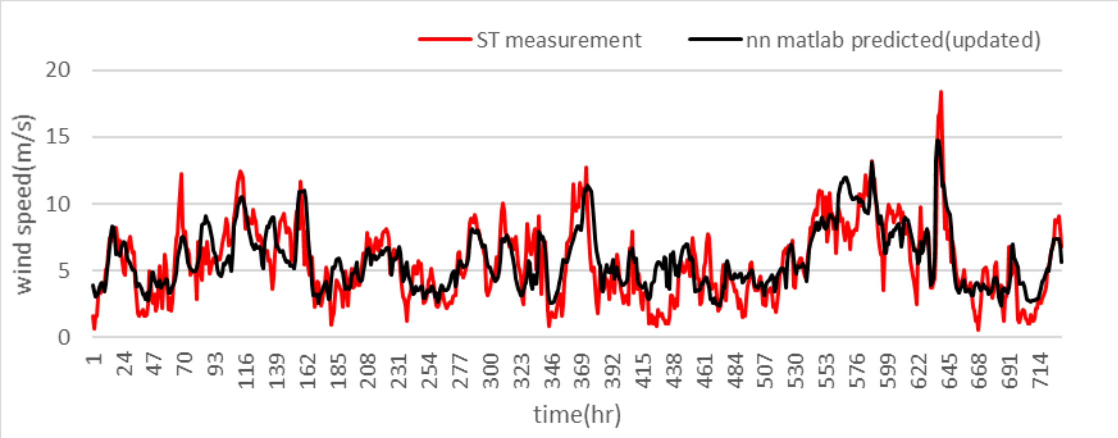


Figure 49 – Time series of wind speeds comparison approach 2 (Russia site)

4.5.4.2 Wind direction analysis

Wind rose graph for measured and predicted data are presented in figure 50.

The major frequency wind direction sectors are predicted but sectors with sectors 4 and 10 being predicted with higher frequencies than the measured site.

Table 13 – Results from WAsP wind analysis software approach 2 (Russia site)

Method	Major sector(s) of wind	A(m/s)	k	Fitted mean velocity(u) (m/s) In all sectors	Available Power density (W/m^2) In all sectors
Measurement data	3rd,4th,5th,8th,9th,10th	5.7	1.95	5.1	155
NN MATLAB	3rd,4th,5th,9th,10th	5.6	2.0	5.0	147

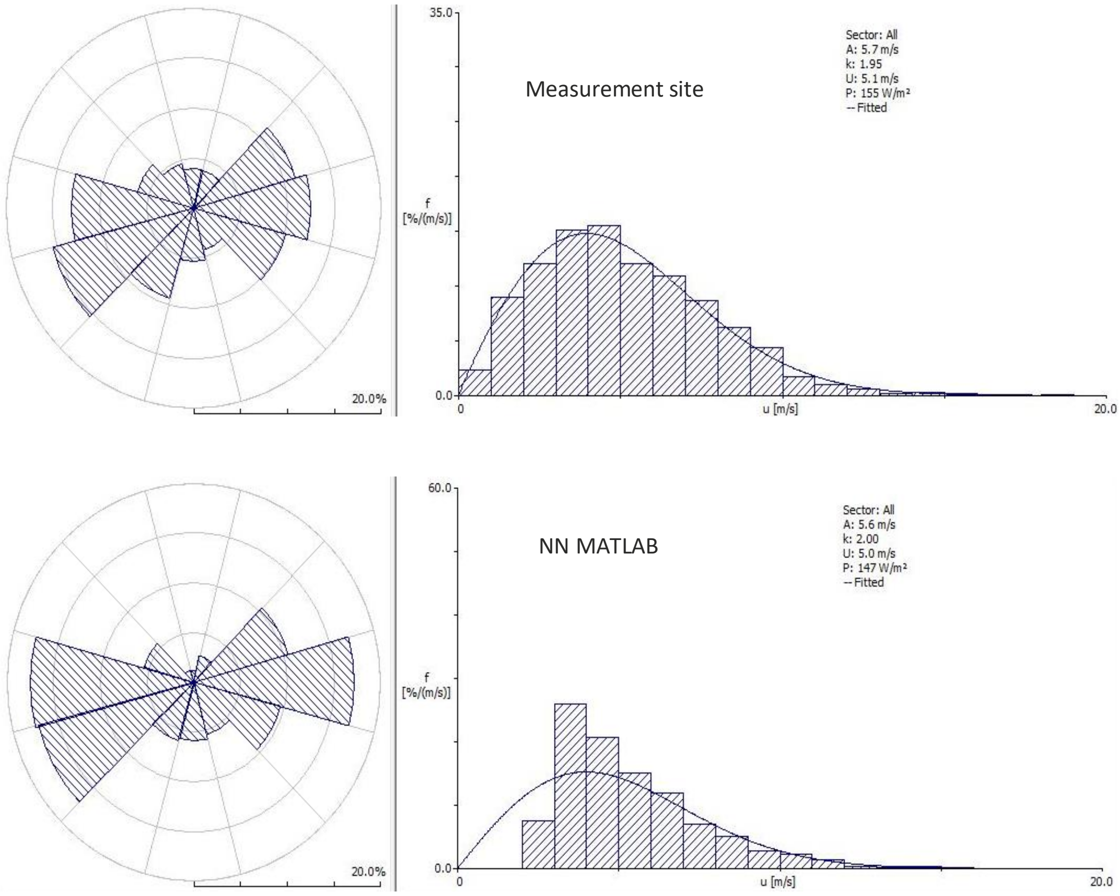


Figure 50 – Wind speeds frequency rose and histogram comparison approach 2 (Russia site)

The distortion in distribution between predicted and measured wind speed frequencies can be clearly viewed in figure 51.

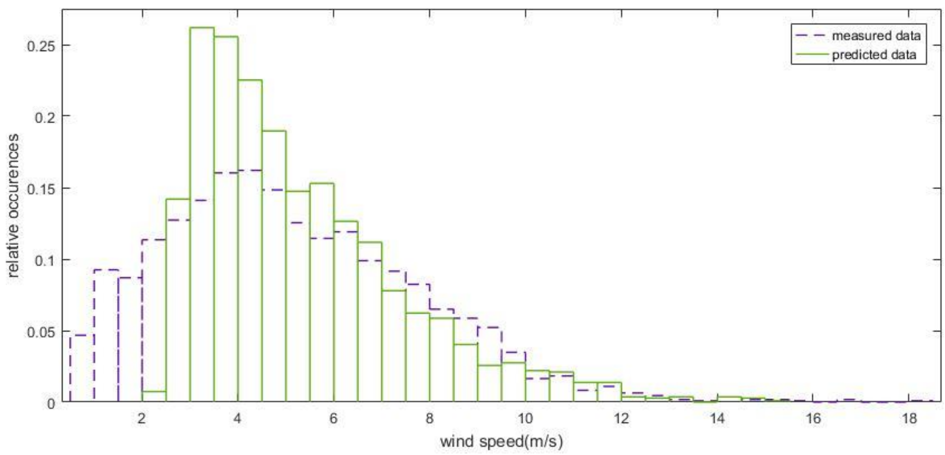


Figure 51 – Wind speed histogram comparison approach 2 (Russia site)

4.5.5 Statistical error analysis for approach 2

The statistical results for all the sites in approach are presented in figure 52, 53, 54 in the order of root mean square error (RMSE), mean absolute percentage error (MAPE), index of agreement (IoA) and mean wind speeds of the data series is also presented in figure 55.

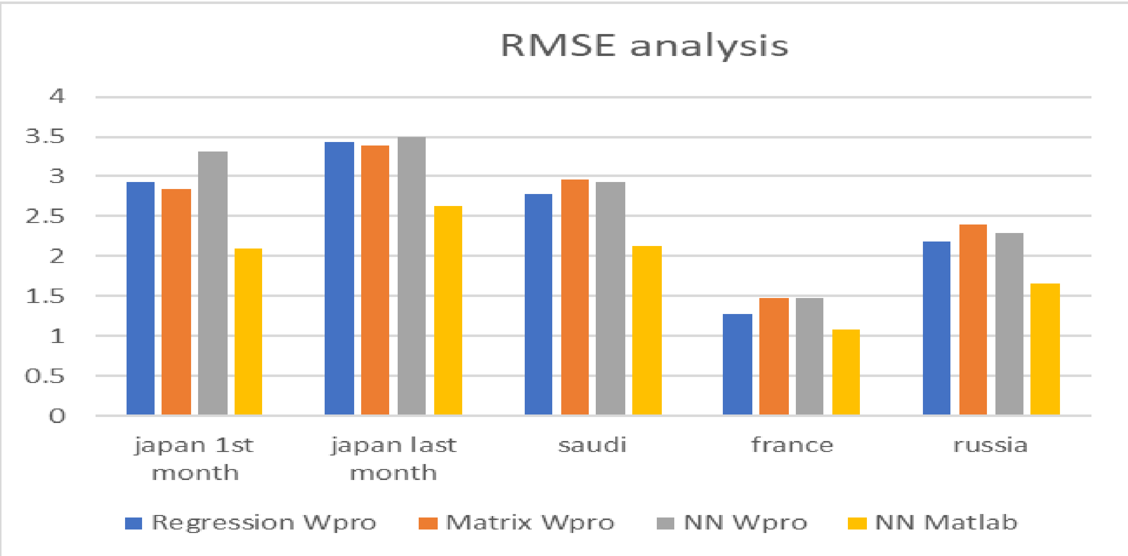


Figure 52 – Different methods RMSE results for all sites approach 2

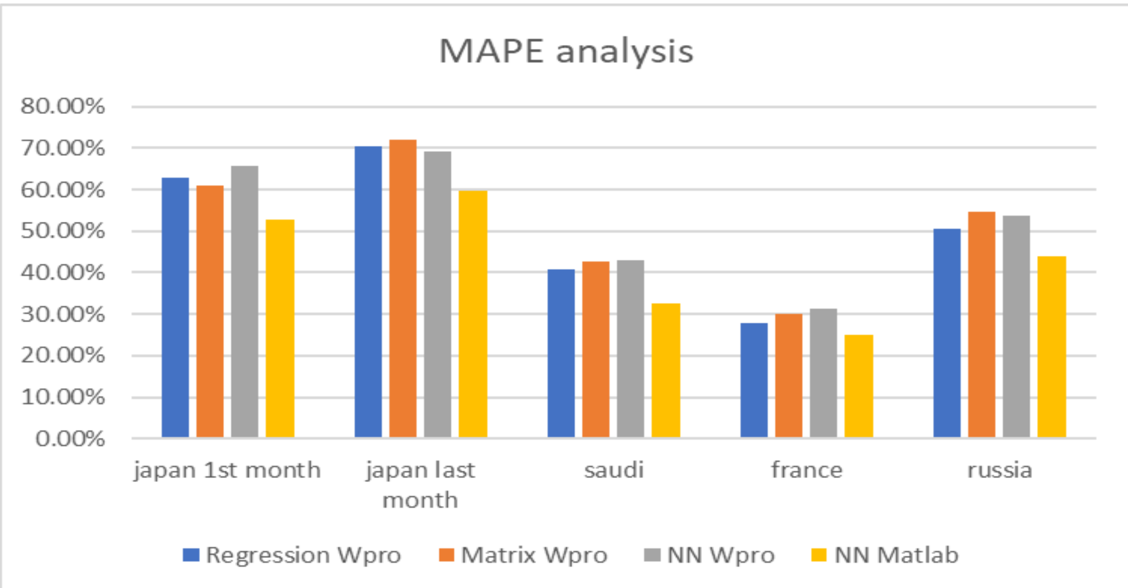


Figure 53 – Different methods MAPE results for all sites approach 2

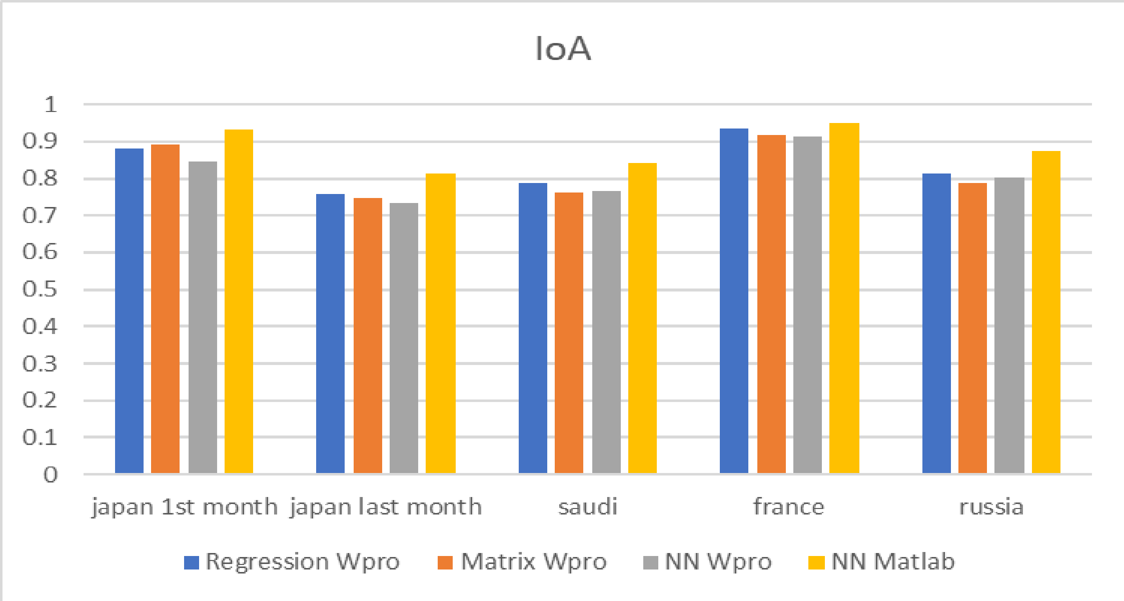


Figure 54 – Different methods IoA results for all sites approach 2

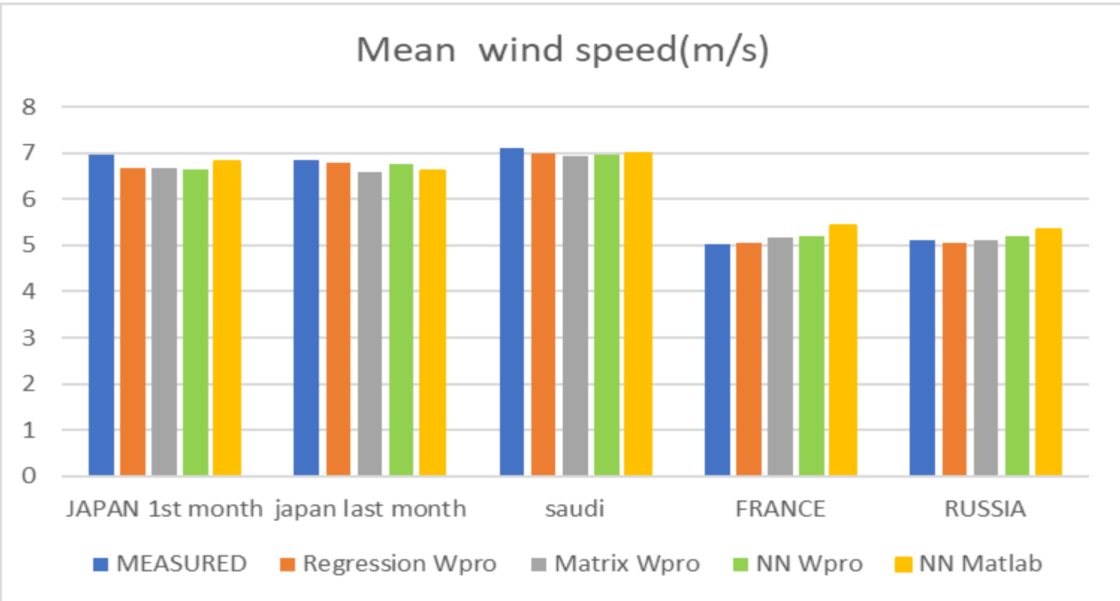


Figure 55 – Mean wind speeds estimations results for all sites in approach 2

The results from statistical analysis proves to be in favour of developed NN. As discussed previously the statistical results cannot alone be enough to prove a model to be a good one. The prediction of wind sector is the second factor, the model built on different sets of data and tested on other sets data seems to catch the prominent wind sectors promisingly. The third and important factor is the wind distribution, again the model suffers with same problem of minimum speeds. One interesting conclusion from approach 2 is that the model built on training data set having correlation coefficient (r) value say suppose x, then performance of the model is good when correlation coefficient for testing datasets is more than x and model's performance depreciates when testing dataset r value is less than x in case of NN MATLAB for example, the correlation coefficient between reference and target site datasets for the first month in Japan is 0.832 and for last month is 0.623. The functional relationship was built on the datasets with r value between reference and target site of about 0.737. Thus, a conclusion could be drawn from this experiment that correlation coefficient of datasets is a prime factor that decides the performance of the neural network, as the function built on common datasets perform better in datasets with higher r and poorer in lower values of r. The Power density deviations of different methods are presented in table 14. The correlation coefficient r is the relation between datasets measured and reference that were used to build the model.

Table 14 – Correlation coefficient between reference and measured site data and Power density deviations in approach 2

Sites	correlation coefficient (r) training dataset	correlation coefficient(r) testing dataset	Deviation in PD estimate (regression method) in %	Deviation in PD estimate (matrix method) in %	Deviation in PD estimate (NN method) in %	Deviation in PD estimate (NN MATLAB) in %
Japan first month	0.737	0.832	-9.5	-5.9	-11.5	-18.1
Japan last month	0.737	0.623	-2.4	-11	-7	-26.6
saudi	0.704	0.725	-4.2	-4.2	-2.8	-21.4
France	0.872	0.898	0.6	4.45	6.5	3.2
Russia	0.824	0.782	No deviation	8.3	9.4	-5.2

* PD – Power Density

As far as proposed model is considered, it performs much better in case of France and very poorer in case of Japan. The conclusion from the approach 2 is that neural network is a robust model that when built on datasets with higher values of correlation coefficient they perform well and on lower values of r their performance deteriorates but in both the cases they generalise the data with mean value of the dataset thus reducing the statistical errors but failing to capture the minimum wind speeds. The neglection of minimum wind speeds should be studied in the future and proper investigation to be carried out to sort out the issue and make this model a real competitor to the methods in industrial software.

CONCLUSIONS

5 CONCLUSIONS AND FUTURE WORK PROPOSAL

5.1 Conclusions

Number of experiments have been carried out in using neural networks for short term prediction of wind speed and wind direction. Five different sites with different complexity of terrains were tested. As far as neural network approach is considered the complexity of terrain does not have any effect in performance of the network. The correlation between the reference and target sites is the major factor deciding the performance of the network along with the length of the concurrent period available for the model. The methods in industrial software also tends to perform well in higher correlation datasets but the length of concurrent period seems to have no effect in their performance, but the uncertainties will be very high in case of lower concurrent periods.

The conclusion from approach 1 is that neural network model is performing statistically better the prediction of wind speed values are close to the measured values and prediction of wind direction is also better but the inability to capture the low wind speeds make the model to perform poorer in terms of power density. The maximum deviation of Power density ranges from -18.0% in Jordan to minimum of -7.9% in France. In case of industrial software, the almost all the model performs similarly, but to be precise based on statistical error values tends to perform similarly with small deviations out of which regression tends to perform better in all sites. As all the models perform similarly in case of frequency distribution, wind direction and deviation in Power density the only deciding factor could be considered is statistical error vales of wind speeds, regression method outperforms other methods statistically. In case of France the even the best performing regression method has deviation of -2% in Power density.

The approach 2 is constructed only for the purpose of understanding the neural network and gain a knowledge about how well the model performs when new sets of data are introduced to it. The second approach was constructed for sites in Japan, Saudi, France and Russia. Where each site is tested with different sets of data. As far japan is considered the performance of the model in first month was better than last month. The reason for this behaviour the reason could be rooted down to the correlation coefficient(r), the r value for first month data was higher than the last months data. But this kind of conclusions cannot be drawn for methods in industrial software because they were tested on entire dataset. The results for the respective months were snipped out from those datasets. The performance of the model in concern with statistical parameters/errors were good and predicting the sector with higher frequency wind was also accomplished. The model faces problem in predicting lower wind speed as in previous case. The deviation in Power density for approach 2 ranges from maximum of -26.6% in case of last month for Japan and minimum of +3.2% in case of France also site in Russia faces a deviation of -5.2%.

After all these studies it is evident that the performance of the model is directly related to the degree of correlation between the reference and target site datasets. But practically, one cannot expect the datasets to be highly correlated all the time. Thus, the neural network model constructed using wind speed and sine and cosine vectors of wind direction can perform statistically better but further investigation to be carried out to analyse the reason behind inefficiency of model to predict lower wind speeds.

5.2 Future work

The shortcoming of the built NN model is its inefficiency in predicting lower windspeeds. The first step in the future work will be to investigate this behaviour of the neural network. One interesting future work will be analysing the key performance indicators of different MCP methodologies in industrial software, involving study on performance of different methods on complexity of terrains, different lengths of concurrent periods, correlation coefficient between datasets etc. To Study the possibility of applying neural networks in mapping of solar radiation and estimation of solar potential.

BIBLIOGRAPHY

6 BIBLIOGRAPHY AND OTHER SOURCES OF INFORMATION

1. http://ec.europa.eu/environment/action-programme/ accessed 31/01/2019.
2. https://www.apren.pt/pt/energias-renovaveis/o-que-sao/ accessed 31/01/2019.
3. N. Vogiatzis, K. Kotti, S. Spanomitsios, and M. Stoukides, "Analysis of wind potential and characteristics in North Aegean, Greece," *Renewable Energy ELSEVIER Journal*, vol. 29, pp 1193-1208, 2004.
4. Lars Landberg, Lisbeth Myllerup, Ole Rathmann. Wind resource assessment an overview. *Wind Energ.* 2003; **6**:261–271 (DOI: 10.1002/we.94).
5. E.W. Hewson, J.E. Wade, and R.W. Baker, Department of Atmospheric Sciences, Oregon state University, Oregon 97331, "A Handbook on the use of trees as an indicator of wind power potential,", *Department of Energy, Division of Distributed Solar Technology, Federal Wind Energy Program, DOE contract No. EY-76-506-227, Task Agreement-24*, June 1979.
6. Wind resource hand book fundamentals of conducting a successful monitoring program by AWS Truepower. April 1997.
7. Wind resource assessment hand book by AWS Truepower. Final report October 2010.
8. Lalit Anjum. Wind resource estimation technique -An overview International Journal of wind and Renewable Energy Volume 3 Issue 2 (pp, 26-38), ISSN: 2277-3975.
9. Addison JF, Hunter A, Bass J, Rebbeck M.A neural network version of the measure–correlate–predict algorithm for estimating wind energy yield. In: Proceedings of the 13th international congress and exhibition on condition monitoring and diagnostic engineering management, Houston, Texas;3–8 December 2000.p.917–922.

10.`Derrick A, RaveyI, MartiI, Glinou G, Pahlke T, Schwenk B, Brand A, Antoniou I, FrandsenS .A unified approach to the evaluation of site specific wind characteristics for use in both energy and load modelling of a potential wind turbine development.

11.Ali Dinler. A new low -correlation MCP (measure-correlate-predict) for wind energy forecasting.

12.S.M. Weekes, A.S. Tomlin. Comparison between the bivariate Weibull probability approach and linear regression for assessment of the long-term wind energy resource using MCP.

13.Jose A. Carta, Sergio Velazquez, Pedro Cabrera. A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site.

14.Anthony L. Rogers, John W. Rogers, James F. Manwell. Comparison of the performance of four measure–correlate–predict algorithms.

15.Jie Zhang, Souma Chowdhury, Achille Messac, Bri-Mathias Hodge. Assessing long-term wind conditions by combining different measure – correlate - predict algorithms.

16.Alejandro Romo Perea, Javier Amezcua and Oliver Probst. Validation of three new measure - correlate-predict models for the long-term prospection of the wind resource.

17.J.C. Woods, S.J Watson A new matrix method of predicting long term wind roses with MCP.

18.Morten Lybech Thøgersen. An introduction to MCP facilities in WindPRO.

19.Basheer, I.A. & Hajmeer, M. 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, Vol. 43, pp. 3-31.

20.Hagan, M.T, Demuth, B.H & Beale M.H. 1996. *Neural Network Design*. Boston: PWS Publishing Company.

21.Fadlalla, A. & Lin, C. 2001. An analysis of the applications of neural networks in finance. *Interfaces*, Vol. 31 (4), pp. 112–122.

22. More, A. & Deo, M.C. 2003. Forecasting wind with neural networks. *Marine Structures*, Vol. 16, pp. 35-49.

23. Tsai, T.-H. & Lee, C.-K. 2005. Design of Dynamic Neural Networks to Forecast Short-term Railway Passenger Demand. *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, pp. 1651-1666.

24. Nagendra, S. M & Khare, M. 2006. Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecological Modelling*, Vol. 190, pp. 99-115.

25. http://radio.feld.cvut.cz/matlab/toolbox/nnet/adv1212.html.

26. Fahlman, S. E. 1988. An empirical study of learning speed in back-propagation networks. Technical report, CMU-CS-88- 162. Pittsburgh: Carnegie Mellon University.

27. Vijendra Singh. Application of Artificial neural network for predicting generated wind power. International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3, 2016.

28. Nurhashinmah Mahamad, Muhamad Kamal b. Mohammed Amin. Power Prediction analysis using Artificial Neural Network in MS Excel platform.

29. Colin John Ritter Sheppard. A master thesis on “Analysis of the measure correlate predict methodology for wind resource assessment.”

30. Mohamed A. Mohandes, Shafiqur Rehman, and Talal O. Halawani.A neural network approach for wind speed prediction. Renewable energy, Vol. 13, No. 3, pp. 345-354, 1998.

31. José A. Carta, Pedro Cabrera, José M. Matías, Fernando Castellano. Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study. Applied energy 158 (2015) 490-507.

32. Jose Carlos Quadrado and Pedro Fonte. Wind speed prediction using artificial neural networks. https://www.researchgate.net/publication/262315667.

33. P. Lopez, R. Velo, F. Maseda. Effect of direction on wind speed estimation in complex terrain using neural networks. Renewable energy 33 (2008) 2266-2272.

34. Sergio Valezquez, Jose A. Carta, J.M. Matias. Comparison between ANNS and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: A case study in the Canary Islands.

35. D.A. Fadare. Modelling of solar energy potential in Nigeria using an artificial neuralnetwork model. Applied Energy 86 (2009) 1410–1422.

36. Mehmet Bilgili, Besir Sahin, Abdulkadir Yasar. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. Renewable Energy 32 (2007) 2350–2360.

37. A. Sagbas, T. Karamanlioglu. The application of Artificial Neural Networks in the estimation of wind speed: A case study. 6th International Advanced Technologies Symposium (IATS'11), 16-18 May 2011, Elazığ, Turkey.

38. D.A. Fadare. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. Applied Energy 87 (2010) 934-942.

39. Anthony L. Rogers, John W. Rogers, James F. Manwell. Uncertainties in Results of Measure-Correlate-Predict Analyses. American Wind Energy Association, Denver, May 2005

40. Mark Taylor, Pawel Mackiewicz, Michael C. Brower, Michael Markus. An Analysis of Wind Resource Uncertainty in Energy Production Estimates. AWS Truewind

41. Matthew Lackner, Anthony L. Rogers, James F. Manwell. Uncertainty Analysis in MCP-Based Wind Resource Assessment and Energy Production estimation. Renewable Energy Research Laboratory, University of Massachusetts, Amherst, MA, 01003 http://works.bepress.com/matthew_lackner/3/

APPENDIXES

7 APPENDICES

7.1 Appendix A – Uncertainties in MCP

The wind resource at a site is elementarily quantified by mean wind speed also in consideration with other factors like probability wind speed, predominant wind direction and turbine intensity. Rogers et al. [39] confronts that wind resource assessment is an uncertain process influenced by large number of factors from wind speed measurements to intrinsic physical variations in wind characteristics. As far as uncertainties in MCP are considered, prediction uncertainties are the primary source and are defined as function of concurrent data length between reference and target sites, statistical model poses some uncertainties as well. The MCP tool is considered most useful in wind power development only when the uncertainties related to it are understood properly. The basic idea behind MCP approach is to obtain a relationship between data from reference site and period of concurrent data from measurement site, thus the obtained relationship is applied to long term reference data to estimate long term wind speed at target site by binning data into different directional sector. Moreover, uncertainties of MCP predictions can be defined by using data from sites for which concurrent long-term data exist in such cases multiple lengths of non-overlapping shorter concurrent data sets can be used with long term reference data to different predict long term characteristics of target site. The resulting standard deviation of different concurrent data sets can be to estimate precision of predictions, author expects that standard deviation should decrease as the square root of concurrent data length. The uncertainty in linear regression of MCP model is calculated by calculating uncertainty of slope, also to understand uncertainty the reason for strong seasonal variability between sites should be understood better. The important conclusion is that uncertainties of MCP estimates in long term wind speed gradually decreases as length of concurrent data increases. For example, the standard deviation of MCP are about 11% of mean wind speed when 1000 hours of concurrent data are available and standard deviation drops to 5% when 9000 hours of data are used.

Taylor et al. [40] conducts a study on two important uncertainty sources namely, uncertainty in MCP methods and uncertainty in wind shear extrapolation to hub height.

The usual height of measuring mast is about 40m to 60m whereas the actual turbine hub height will be around 80m to 100m and still high for some modern turbines. Extrapolation of hub height measurements from such towers intrinsically contains uncertainty α.

Wind shear profile can be defined by power law.

$$\frac{V_2}{V_1} = \left(\frac{h_2}{h_1}\right)^{\alpha} \qquad \text{(a-1)}$$

Where V_1, V_2 are the measured wind speeds at heights h_1 and h_2 respectively.

$$\alpha = \frac{\ln\left(\frac{V_2}{V_1}\right)}{\ln\left(\frac{h_2}{h_1}\right)} \qquad \text{(a-2)}$$

thus ,

$$V_3 = V_2\left(\frac{h_3}{h_2}\right)^{\alpha} \qquad \text{(a-3)}$$

V_3 is the extrapolated wind speed at height h_3.

The random error of wind speed V_3 and other normally distributed uncertainties is given by equation

$$\Delta V_3 = \sqrt{\sum_{i=1}^{j}(G_i \Delta x_i)^2} \qquad \text{(a-4)}$$

Where, G_i is sensitivity coefficient,

ΔX_i is measurement error of input parameters.

Also, author says that sources of uncertainty of wind speed(V_3) extrapolated to hub height are mostly due to uncertainty in wind speeds V_1 and V_2 and thus uncertainties due to h_1 and h_2 are neglected. The most important conclusion from this research paper conveys that anemometer measurement uncertainty, vertical spacing on tower, monitoring period and length of concurrent data period used in MCP are the vital factors should take care of, also author confronts that there are uncertainty at almost all stages of wind resource assessment and should be accounted for in final results considering each source of uncertainty is statistically independent from each other σ.

$$\sigma_{total} = \sqrt{\sum_{i=1}^{j} \sigma_i^2} \tag{a-5}$$

Where, σ_{total} is total uncertainty,

σ_i is uncertainty associated with each source (measurement, climatological adjustment, extrapolation to hub height).

Lackner et al. [41] Provides a frame work for an accurate and straight forward way of accounting uncertainty three important aspects of uncertainty are defined and are classified accordingly, combining uncertainties that arise in assessing wind resource, handling of uncertainties in wind turbine power output and energy losses and modelling Weibull distribution to estimate overall annual energy production (AEP) uncertainties. The author considers that Weibull distribution as an approximation of wind speed distribution.

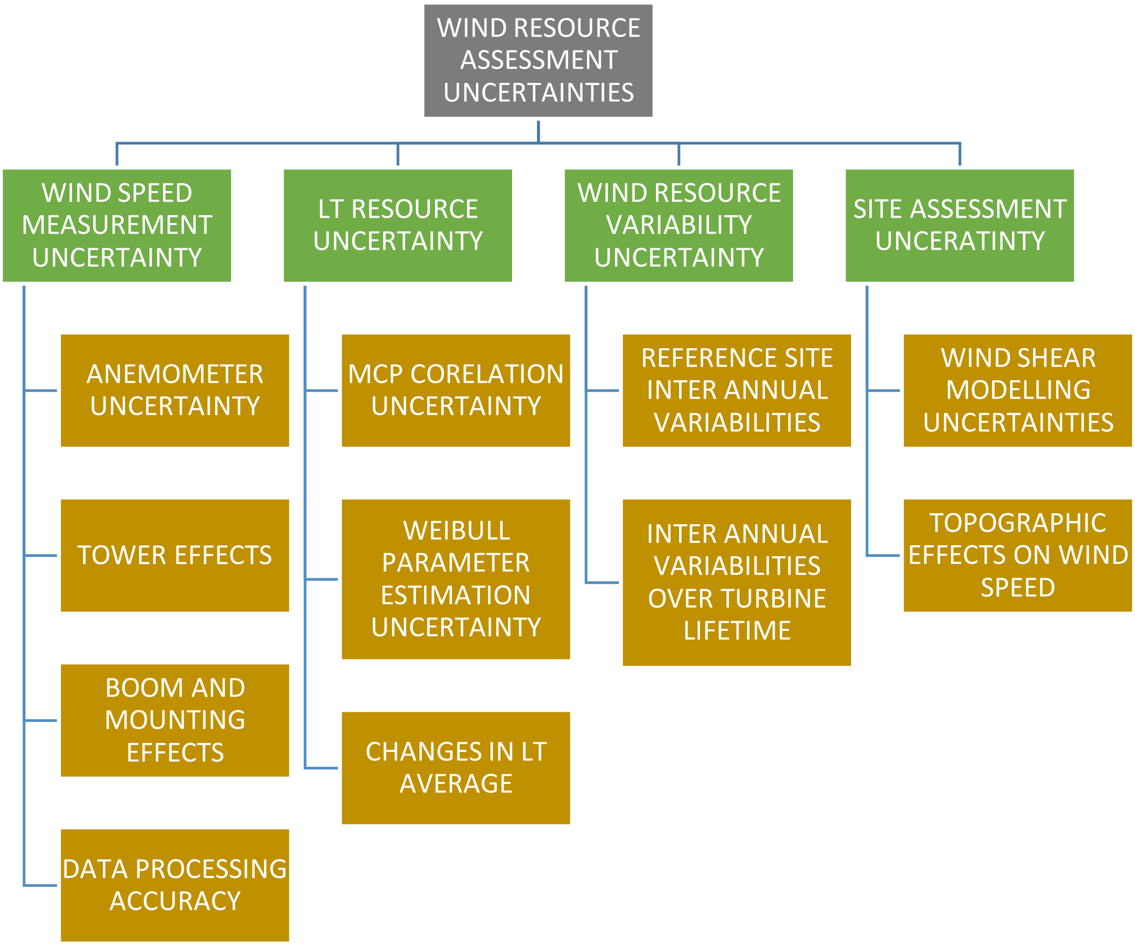


Figure 56 – Uncertainties in various stages of wind resource assessment [41]

7.2 Appendix B – ANN code

```
function [Y,Xf,Af] = myNeuralNetworkFunction(X,~,~)
%MYNEURALNETWORKFUNCTION neural network simulation function.
%
% Generated by Neural Network Toolbox function genFunction,
%
% [Y] = myNeuralNetworkFunction(X,~,~) takes these arguments:
%
%   X = 1xTS cell, 1 inputs over TS timesteps
%   Each X{1,ts} = 3xQ matrix, input #1 at timestep ts.
%
% and returns:
%   Y = 1xTS cell of 1 outputs over TS timesteps.
%   Each Y{1,ts} = 3xQ matrix, output #1 at timestep ts.
%
% where Q is number of samples (or series) and TS is the
number of timesteps.

%#ok<*RPMT0>

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
x1_step1.xoffset = [0.1; -1;-1];
x1_step1.gain = [0.124223602484472;1;1];
x1_step1.ymin = -1;

% Layer 1
b1 = [-
0.22371202743185886;4.8301867117221624;1.6959306527346047;0.3226
3628146778978;1.2085227837462851;0.33960590272965252;0.973088989
86252346;0.63049838614195719; -
0.28111400891672728;3.4467808412422682];
IW1_1 = [-1.9169826663587533 -0.52458321215751991 -
0.29989380617310807;-2.374316516870107 3.8214722906052736 -
1.7786883059851473;-0.79845811447342618 0.44617872474110454 -
2.6539533071691852;0.94504835403579401 -0.99657596978105234 -
1.2420460718630792;-0.62777886430836471 1.3809823131271752
0.045390879958295292;0.79537853082171128 1.1689236966123659 -
1.1746403790961644;5.441964516152769 0.20296415628217754
1.1339806212318284;1.6668704305862208 0.16331813749084081 -
1.7727870255287541;-0.79243375574777819 -0.047894147014499698
0.54476794890371838;1.3578102961632432 0.11565171644104594
1.8321287791391609];

% Layer 2
b2 = [0.46011582904656628; -
0.40436612566310443;1.8362639380854131];
LW2_1 = [0.44143131140714142 -0.49618958918919531
0.41563114386998384 -0.88972870388085212 0.48718406447050586 -
1.0189532523428178 0.10621201194013601 -0.79125102067812503 -
3.6666613524999248 -1.1543450014822307;-0.82687583294368783
```

```
0.015527702281444754 0.32758923231171971 0.14810888058571511
0.80597051156064181 0.24420175500517038 -0.092249747929822698
0.39296422753156862 1.5650302207199962 -0.088043942772152661;-
0.10534211348518398 0.085144276919809192 -0.041385620276784O1 -
0.63238347589407651 -0.49219823454897488 -0.43384473723824862 -
0.025538901632557127 -0.34606326392650255 -0.90976025747251887 -
1.4868633757756595];

% Output 1
y1_step1.ymin = -1;
y1_step1.gain = [0.0953288846520496;1;1];
y1_step1.xoffset = [0.54; -1;-1];

% ===== SIMULATION ========

% Format Input Arguments
isCellX = iscell(X);
if ~isCellX, X = {X}; end;

% Dimensions
TS = size(X,2); % timesteps
if ~isempty(X)
Q = size(X{1},2); % samples/series
else
Q = 0;
end

% Allocate Outputs
Y = cell(1,TS);

% Time loop
for ts=1:TS

% Input 1
Xp1 = mapminmax_apply(X{1,ts},x1_step1);

% Layer 1
a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*Xp1);

% Layer 2
a2 = repmat(b2,1,Q) + LW2_1*a1;

% Output 1
Y{1,ts} = mapminmax_reverse(a2,y1_step1);
end

% Final Delay States
Xf = cell(1,0);
Af = cell(2,0);

% Format Output Arguments
if ~isCellX, Y = cell2mat(Y); end
end
```

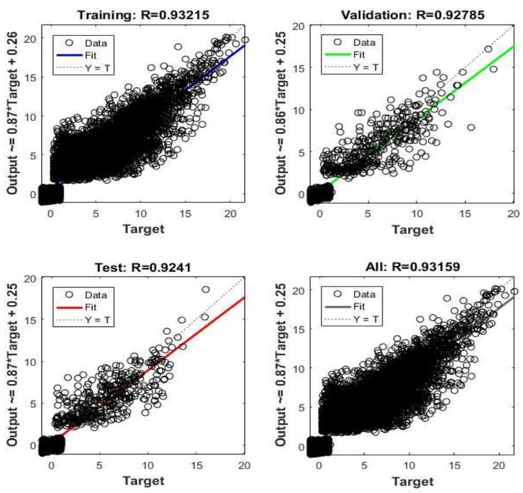
```
% ===== MODULE FUNCTIONS ========

% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end

% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end
```

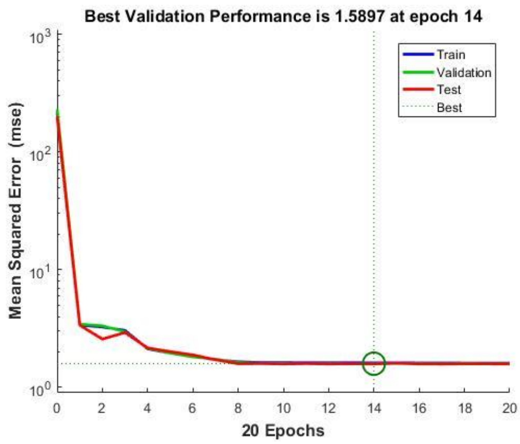
7.3 Appendix C – Linear regression and performance graphs for ANN model

The linear regression graph and performance curve for final optimized neural network model is presented for approach 1 and approach 2.

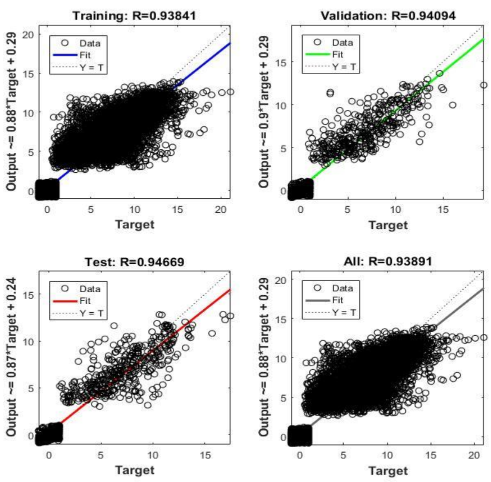
Approach 1

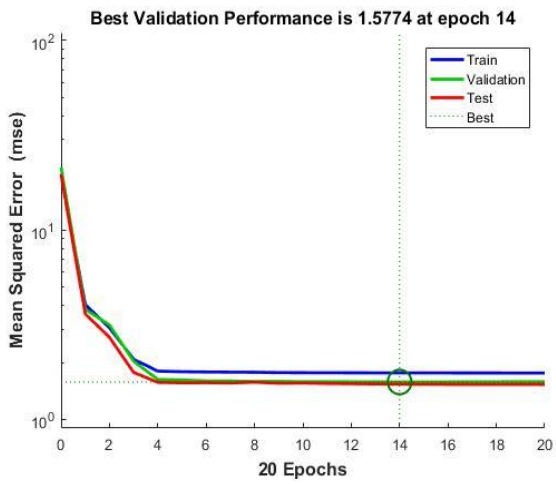
Japan



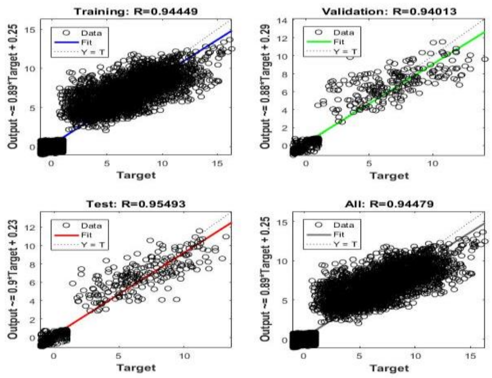


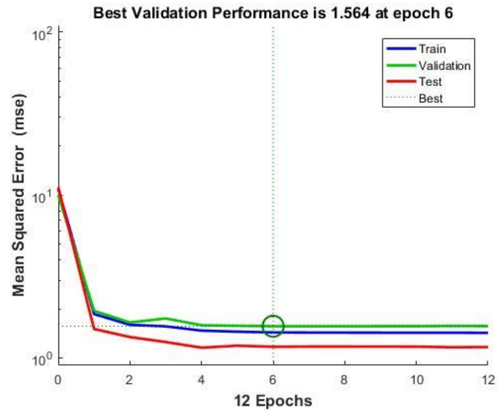
Saudi Arabia



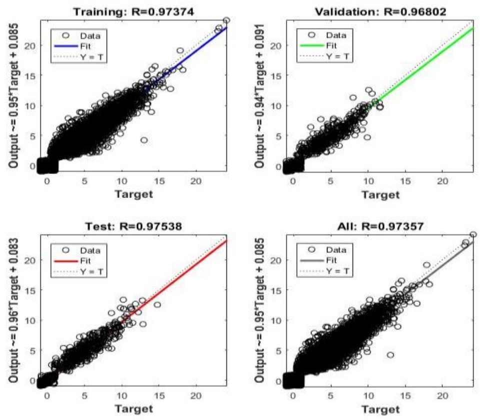


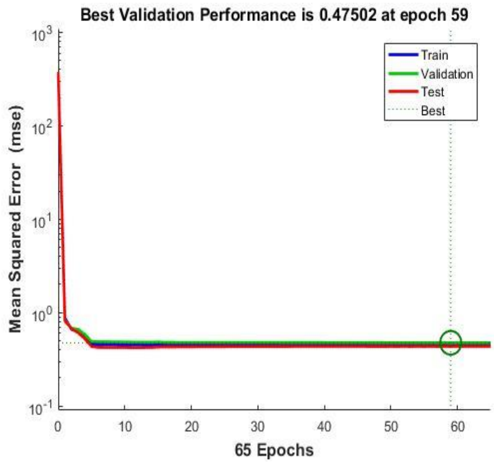
Jordan





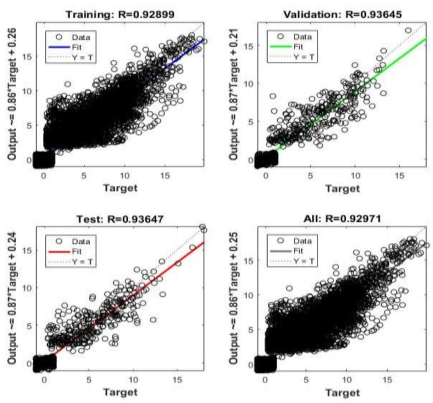
France

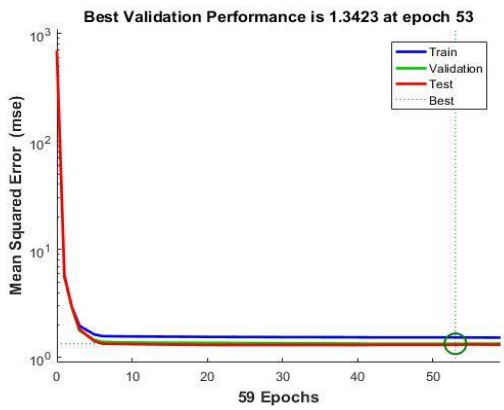




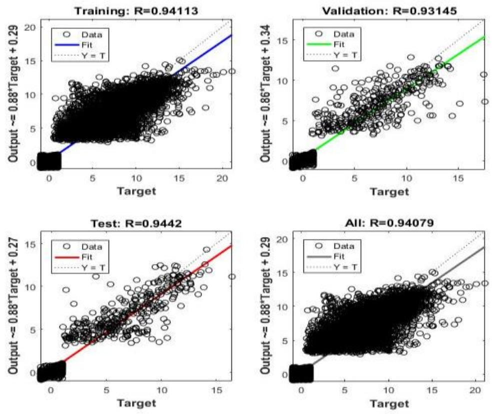
Approach 2

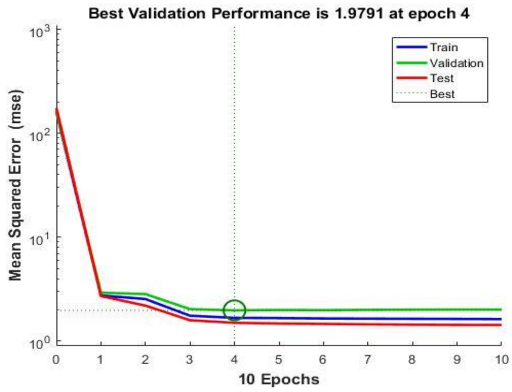
Japan



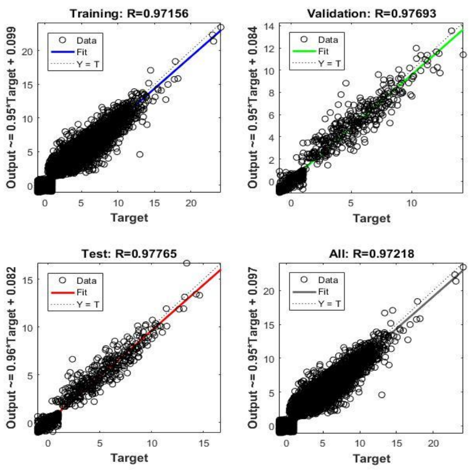


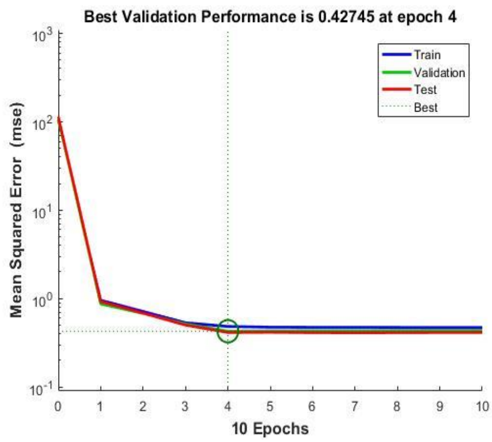
Saudi Arabia





France





7.4 Appendix D – Determination of number of neurons in hidden layer

There were number experiments conducted to fix the number of hidden neurons. The datasets were divided into random subsets of data 60% for training, 20% for validation and 20% for testing and experimented 10 times. The mean of MSE value from the testing dataset is used to see the convergence.

Approch 1

Japan

	5 NEURONS		10 NEURONS		15 NEURONS		20 NEURONS		25 NEURONS		30 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R
	1.7212	0.927	1.62	0.9304	1.667	0.931	1.619	0.929	1.712	0.927	1.625	0.934
	1.795	0.924	1.74	0.923	1.678	0.927	1.558	0.934	1.611	0.931	1.648	0.93
	1.67	0.929	1.6	0.931	1.59	0.93	1.603	0.931	1.591	0.934	1.66	0.93
	1.71	0.928	1.68	0.929	1.647	0.929	1.647	0.933	1.552	0.931	1.725	0.924
	1.71	0.929	1.64	0.929	1.462	0.939	1.623	0.929	1.683	0.928	1.683	0.926
	1.701	0.928	1.582	0.93	1.629	0.932	1.7	0.928	1.589	0.93	1.511	0.934
	1.69	0.93	1.627	0.929	1.659	0.927	1.617	0.933	1.6	0.929	1.583	0.931
	1.739	0.926	1.694	0.93	1.577	0.932	1.674	0.928	1.634	0.928	1.68	0.926
	1.76	0.926	1.796	0.922	1.624	0.933	1.488	0.934	1.663	0.926	1.73	0.927
	1.69	0.93	1.691	0.926	1.671	0.928	1.653	0.93	1.557	0.936	1.74	0.92
mean of MSE	1.71862		1.667		1.6204		1.6182		1.6192		1.6585	

Saudi Arabia

3 NEURONS		5 NEURONS		7 NEURONS		10 NEURONS	
MSE	R	MSE	R	MSE	R	MSE	R
1.78	0.936	1.61	0.941	1.716	0.923	1.672	0.941
1.79	0.938	1.683	0.942	1.615	0.939	1.606	0.943
1.82	0.938	1.702	0.939	1.6	0.943	1.735	0.938
1.765	0.937	1.64	0.942	1.702	0.94	1.817	0.938
1.759	0.937	1.753	0.938	1.9	0.934	1.773	0.937
1.742	0.939	1.66	0.939	1.757	0.938	1.694	0.939
1.76	0.94	1.701	0.94	1.754	0.939	1.783	0.937
1.756	0.939	1.766	0.938	1.77	0.938	1.804	0.936
1.855	0.935	1.739	0.938	1.668	0.941	1.735	0.939
1.817	0.935	1.79	0.936	1.8	0.939	1.593	0.943
1.7844		1.7044		1.7282		1.7212	

Jordan

	5 NEURONS		10 NEURONS		15 NEURONS	
	MSE	R	MSE	R	MSE	R
	1.397	0.944	1.541	.94.	1.387	0.945
	1.549	0.939	1.412	0.944	1.411	0.945
	1.478	0.936	1.278	0.949	1.427	0.944
	1.444	0.9414	1.422	0.943	1.338	0.947
	1.337	0.943	1.333	0.948	1.401	0.946
	1.51	0.94	1.381	0.945	1.5	0.942
	1.5	0.939	1.504	0.942	1.484	0.941
	1.4	0.944	1.415	0.943	1.466	0.943
	1.456	0.942	1.312	0.949	1.4411	0.943
	1.443	0.943	1.46	0.941	1.373	0.946
mean of MSE	1.4514		1.4058		1.42281	

France

	5 NEURONS		10 NEURONS		15 NEURONS		20 NEURONS		25 NEURONS		30 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R
	0.499	0.97	0.4871	0.97	0.473	0.972	0.469	0.972	0.452	0.972	0.505	0.97
	0.488	0.97	0.461	0.973	0.473	0.972	0.471	0.973	0.493	0.971	0.481	0.972
	0.484	0.971	0.483	0.973	0.463	0.973	0.482	0.971	0.448	0.973	0.446	0.974
	0.498	0.97	0.455	0.974	0.486	0.971	0.468	0.972	0.476	0.971	0.476	0.971
	0.524	0.968	0.483	0.972	0.478	0.971	0.475	0.972	0.4515	0.973	0.4504	0.973
	0.477	0.97	0.49	0.97	0.464	0.972	0.449	0.973	0.457	0.9704	0.476	0.971
	0.479	0.972	0.493	0.97	0.481	0.972	0.486	0.971	0.481	0.9707	0.467	0.971
	0.484	0.971	0.471	0.972	0.452	0.974	0.465	0.972	0.475	0.972	0.4402	0.973
	0.456	0.972	0.478	0.972	0.48	0.971	0.468	0.972	0.457	0.973	0.483	0.971
	0.462	0.972	0.445	0.974	0.465	0.971	0.457	0.973	0.4801	0.972	0.464	0.973
mean of MSE	0.4851		0.47461		0.4715		0.469		4.6706		0.46886	

Russia

	5 NEURONS		10 NEURONS		15 NEURONS		20 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R
	1.258	0.948	1.17	0.946	1.26	0.945	1.31	0.944
	1.211	0.948	1.23	0.47	1.22	0.946	1.27	0.9469
	1.17	0.949	1.22	0.946	1.233	0.947	1.23	0.946
	1.2	0.946	1.25	0.946	1.19	0.949	1.24	0.948
	1.27	0.942	1.19	0.949	1.18	0.949	1.12	0.951
	1.211	0.945	1.227	0.948	1.256	0.944	1.25	0.944
	1.18	0.949	1.19	0.945	1.3	0.943	1.29	0.946
	1.23	0.948	1.22	0.948	1.25	0.947	1.21	0.946
	1.25	0.947	1.25	0.947	1.288	0.943	1.25	0.9469
	1.32	0.944	1.19	0.948	1.22	0.947	1.25	0.944
mean of MSE	1.23		1.2137		1.2397		1.242	

Approach 2

Japan

	5 NEURONS		10 NEURONS		15 NEURONS		20 NEURONS		25 NEURONS		30 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R	MSE	R
	1.67	0.921	1.621	0.916	1.613	0.922	1.453	0.933	1.602	0.921	1.589	0.926
	1.68	0.921	1.674	0.922	1.53	0.925	1.557	0.927	1.554	0.929	1.626	0.924
	1.61	0.924	1.604	0.924	1.608	0.924	1.551	0.927	1.5871	0.925	1.721	0.913
	1.773	0.916	1.588	0.925	1.556	0.927	1.63	0.922	1.583	0.928	1.492	0.929
	1.676	0.922	1.689	0.924	1.589	0.929	1.557	0.926	1.527	0.929	1.627	0.927
	1.72	0.921	1.607	0.924	1.596	0.924	1.72	0.917	1.565	0.927	1.484	0.929
	1.73	0.925	1.642	0.924	1.572	0.928	1.572	0.925	1.551	0.929	1.52	0.929
	1.63	0.923	1.631	0.9221	1.601	0.926	1.541	0.926	1.498	0.93	1.655	0.924
	1.58	0.926	1.687	0.929	1.689	0.921	1.668	0.928	1.502	0.926	1.535	0.93
	1.685	0.921	1.596	0.923	1.623	0.924	1.532	0.933	1.613	0.925	1.548	0.925
mean of MSE	1.6754		1.6339		1.594889		1.5781		1.55821		1.583222	

Saudi Arabia

	5 NEURONS		10 NEURONS		15 NEURONS	
	MSE	R	MSE	R	MSE	R
	1.716	0.9401	1.832	0.935	1.738	0.939
	1.839	0.934	1.791	0.937	1.806	0.938
	1.903	0.933	1.758	0.936	1.738	0.938
	1.728	0.94	1.731	0.939	1.754	0.939
	1.72	0.939	1.72	0.939	1.812	0.937
	1.78	0.932	1.862	0.935	1.853	0.934
	1.777	0.937	1.779	0.937	1.726	0.939
	1.827	0.935	1.736	0.938	1.736	0.937
	1.817	0.935	1.699	0.939	1.71	0.938
	1.82	0.935	1.713	0.939	1.839	0.935
mean of MSE	1.7927		1.7621		1.7712	

France

	15 NEURONS		20 NEURONS		25 NEURONS		30 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R
	0.5	0.97	0.495	0.9706	0.488	0.97	0.493	0.973
	0.482	0.972	0.493	0.9706	0.451	0.973	0.493	0.971
	0.482	0.971	0.53	0.9685	0.468	0.972	0.513	0.972
	0.478	0.972	0.506	0.971	0.478	0.971	0.48	0.9726
	0.487	0.971	0.485	0.9716	0.4444	0.973	0.486	0.973
	0.483	0.9706	0.476	0.972	0.469	0.972	0.489	0.97
	0.508	0.972	0.491	0.97	0.464	0.9726	0.467	0.971
	0.489	0.9706	0.471	0.971	0.5	0.9711	0.4908	0.9716
	0.516	0.97	0.506	0.969	0.505	0.97	0.458	0.972
	0.471	0.973	0.464	0.973	0.503	0.97	0.535	0.969
mean of MSE	0.4896		0.4917		0.47704		0.49048	

Russia

	5 NEURONS		10 NEURONS		15 NEURONS		20 NEURONS	
	MSE	R	MSE	R	MSE	R	MSE	R
	1.28	0.951	1.27	1.952	1.349	0.948	1.3	0.949
	1.3	0.95	1.36	0.945	1.298	0.95	1.276	0.951
	1.134	0.944	1.28	0.95	1.304	0.947	1.28	0.951
	1.286	0.951	1.27	0.949	1.298	0.952	1.278	0.954
	1.402	0.941	1.3	0.951	1.281	0.95	1.339	0.945
	1.371	0.949	1.32	0.951	1.211	0.952	1.296	0.95
	1.345	0.944	1.24	0.95	1.29	0.949	1.216	0.952
	1.27	0.949	1.26	0.948	1.353	0.947	1.15	0.956
	1.436	0.943	1.32	0.948	1.21	0.953	1.38	0.943
	1.33	0.945	1.28	0.951	1.2	0.955	1.323	0.946
MEAN OF MSE	1.3154		1.29		1.2794		1.2838	