On spatial analysis of wind energy potential in Malaysia

NURULKAMAL MASSERAN[1], AHMAD MAHIR RAZALI[1,2],
KAMARULZAMAN IBRAHIM[1,2], WAN ZAWIAH WAN ZIN[1]
AND AZAMI ZAHARIM[2]
[1]School of Mathematical Sciences,
Department Faculty of Sciences and Technology
[2]Solar Energy Research Institute (SERI)
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor D. E.
MALAYSIA
mall_evolution@yahoo.com, mahir@ukm.my, kamarulz@ukm.my,
w_zawiah@ukm.my, azami@vlsi.eng.ukm.my

Abstract: Statistical distribution for describing the wind speed at a particular location provides information about the wind energy potential. In this paper, nine different statistical distributions are fitted to the data of average hourly wind speed for 60 wind stations across the west and east Malaysia from the year 2000 to 2009. The distributions found to be most adequate for describing the wind speed at each particular station are determined based on several goodness of fits criteria. The spatial dependence in the data is investigated by making use of the semivariogram involving the expected speed found using the identified distribution for each individual station. Since the spatial dependence in the data of Peninsular Malaysia could not be described well by the semivariogram, the inverse distance weighting method is used for describing the spatial distribution of wind speed. For the data of East Malaysia, however, the power semivariogram is found to be fitted quite well. Accordingly, the kriging method is applied for spatial prediction. It is found that, the regions in the northeast, northwest and southeast of the peninsula have a good potential for wind energy. For East Malaysia, the northeast and southwest regions of Sarawak are found to be the most potential.

Key-Words: Wind energy, wind speed distribution, semivariogram, spatial estimation, kriging, inverse distance weighting method

1 Introduction

Various studies on wind speed have been carried out by many researchers, particularly for the purpose of generating energy [1,2,3]. The growing interest in wind as a possible source of energy for producing electricity nowadays is dated to the oil crisis that occurred in the mid-seventies [4]. Wind energy has become an important alternative renewable source of energy because it is clean and cost effective for many applications such as electric power production, water pumping, etc. Utilization of wind as an energy resource has been growing rapidly in the whole world since the consumption of other energy resources such as fuel, nuclear and coal contributes to environmental pollution and global warming. Moreover, wind energy does not impose transportation problem and does not require utilization of high technology [5].

In 2009, wind turbines built around the world are found to be generating electricity at a rate of 340TWh per year. This capacity is equivalent to 2% of worldwide electricity usage. Asia became the world’s wind locomotive in the year 2009, mainly due to the two large markets of China and India. The wind capacity installed in Asia has reached the total amount of 40.0 GW. In Malaysia, it is suggested that the potential for wind energy generation depends on the availability of the wind resource which is found to vary according to location [4]. Since wind power is seen as a potential alternative energy resource, more in depth studies have to be carried out in Malaysia in order to explore this oppurtunity [5-7].

Among the early works on wind energy research in Malaysia is the work by Sopian et al. [7]. They have analyzed 10 wind stations in Malaysia using Weibull distribution. Their results indicate that Mersing and Kuala Terengganu possess the best potential for wind energy development. It has been described by Ong et al. [4], 150 kW wind turbine which was built in Terumbu Layang-Layang in 2005 had demonstrated some success. Recently, Tenaga Nasional Berhad (TNB), which is the only electricity supplier in Malaysia, had built two units of wind turbine at Pulau Perhentian. Also, the Ministry of Rural and Regional Development had built 8 small units of wind turbine in Sabah and Sarawak for local communities [8].

Since wind speed vary according to sites, it is reasonable to consider that there exist a spatial dependence between the locations. Observations in close spatial proximity are expected to be more similar than observations that are more spatially separated [8]. In general, the greater the distance, the more is the regional independent and vice versa for the smaller distance [9]. In this study, we identify the distribution of wind speed for various stations and apply the geostatistical interpolator such as kriging and inverse distance weighting method in order to gain some insight on the wind regime in Malaysia.

Among methods that are commonly used for spatial mapping are geostatistical interpolator such as kriging and inverse distance weighting method. Cellura et al. [10], for example, use inverse distance weighting and universal kriging methods for spatial prediction of wind speed in Sicily. They found that their wind speed map are quite similar to the Italian Wind Atlas. The application of kriging method for spatial prediction can also be found in various areas of research such as rainfall analysis, health sciences, thermal sciences, etc. This paper provides a rough map of wind speed to identify which region in Malaysia has the potential of generating wind energy before a more in depth analysis is carried at the specific area.

2 Study area, regional climate and data

Malaysia is a country which lies entirely in the equatorial zone, situated in the south east part of Asia, having a geographic coordinate of 2° 30' in the north latitude and 112° 30' in the east longitude. Throughout the year, Malaysia experiences a wet and humid condition with daily temperature ranging from $25.5 \text{ to } 35°C$. The wind that blows across the peninsula as well as Sabah and Sarawak is influenced by the monsoon seasons, namely southwest monsoon, northeast monsoon and two short inter-monsoons. The two monsoons that contribute to rainy seasons are the southwest monsoon, occurring in May until September, and the northeast monsoon which occurs from November until March. The later monsoon brings about heavier rainfall in the peninsula, with the worst affected areas are in the east and south. Malaysia is a maritime country which is also influenced by the effect of sea breezes and land breezes especially when the sky is not cloudy. During the afternoon, sea breezes with the speed of between 10 to 15 knots usually occur. Meanwhile, at night, the land breezes occur. The data used in this study which consists of hourly wind speed (km/hour) from January 2000 to November 2009 for 50 wind stations across the country were obtained from the Department of Environment.

3 Wind speed probability distribution

In order to describe the behaviour of wind speed at a particular location, it is necessary to identify the distribution which best fits the data. Suitable distributions for each wind station has been determined by fitting nine types of statistical distribution to the data, namely Weibull (WE), Burr (BR), Gamma (GA), Inverse Gamma (IGA), Inverse Gaussian (IGU), Exponential (EX), Rayleigh (RY), Lognormal (LN) and Erlang (ER) to the data. Here, ER is just a special case of Gamma distribution where the shape parameter is an integer. In this study, parameter estimation for each model is done by using maximum likelihood method. Table 1 below shows the list of probability density functions with their respective mean and maximum likelihood estimator, for details see [11-14]. The maximum likelihood estimator (MLE) for the parameters of WE, GA, IGU, ER, IGA, BR distributions can be determined numerically by using methods such as Newton-Rapson, scoring, EM algorithm, quasi-Newton, Nelder-Mead method etc. In this study, Nelder-Mead method was used as an optimization technique for determining the MLE of the parameters [15]. For other distribution such as LN, RY, and EX the MLEs can be easily determined. Table 2 and Table 3 show the parameter estimates found for each distribution. Several goodness of fit tests which include Kolmogorov-Smirnov (KS), Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to determine the most suitable statistical distribution for the data of each wind station. In addition, R^2 coefficient was also used to evaluate the goodness of fit for each method. A large value of R^2 indicates a better fitted theoretical distribution to the data.

Table 1. List of probability density functions and maximum likelihood estimators

Model	Probability density function (PDF)	Maximum likelihood estimator (MLE)	Mean
Lognormal (LN)	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right]$	$\hat{\mu}=\frac{\sum_{i=1}^{n}\ln x_i}{n}$ *and* $\sigma=\frac{\sum_{i=1}^{n}(\ln x_i-\hat{\mu}\quad)^2}{n}$	$e^{\mu+\frac{\sigma^2}{2}}$
Weibull (WE)	$f(x)=\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$	$\hat{\beta}=\left[\left(\sum_{i=1}^{n}x_i^{\hat{\beta}}\ln x_i\right)\left(\sum_{i=1}^{n}x_i^{\beta}\right)^{-1}-n^{-1}\sum_{i=1}^{n}\ln x_i\right]^{-1}$ *and* $\hat{\alpha}=\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}x_i^{\hat{\beta}}\right]^{\frac{1}{\hat{\beta}}}$	$\alpha\Gamma\left[\frac{(\beta+1)}{\beta}\right]$
Rayleigh (RY)	$f(x)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	$\hat{\sigma}=\sqrt{\sum_{i=1}^{n}x_i \Big/ 2n}$	$\sigma\sqrt{\frac{\pi}{2}}$
Exponential (EX)	$f(x)=\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$	$\hat{\theta}=\overline{x}$	θ
Gamma (GA)	$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$	$\hat{\beta}=\frac{\overline{x}}{\alpha}$ *and* $\ln(\hat{\alpha})-\psi(\alpha)=\ln\left(\frac{1}{n}\sum_{i-1}^{n}x_i\right)-\frac{1}{n}\sum_{i=1}^{n}\ln x_i$	$\alpha\beta$
Inverse Gaussian (IGU)	$f(x)=\left[\frac{\lambda}{2\pi x^3}\right]^{\frac{1}{2}}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}$	$\hat{\mu}=\overline{x}$ *and* $\hat{\lambda}=n\left[\sum_{i=1}^{n}x_i^{-1}-(\overline{x})^{-1}\right]^{-1}$	μ
Burr (BR)	$f(x)=\frac{aqx^{a-1}}{b^a\left[1+(x/b)^a\right]^{1+q}}$	$\frac{n}{a}+\sum_{i=1}^{n}\ln(x_i/b)=(1+q)\sum_{i=1}^{n}\ln(x_i/b)\left[\left(\frac{b}{x_i}\right)^{a}+1\right]^{-1}$ *and* $n=(1+q)\sum_{i=1}^{n}\left[\left(\frac{b}{x_i}\right)^{a}+1\right]^{-1}$ *and* $\frac{n}{q}=\sum_{i=1}^{n}\ln\left[\left(\frac{x_i}{b}\right)^{a}+1\right]$	$\frac{b\Gamma\left(\frac{a+1}{a}\right)\Gamma\left(\frac{q-1}{a}\right)}{\Gamma(q)}$
Inverse Gamma (IGA)	$f(x)=\frac{\beta^{p}}{\Gamma(p)}x^{-p-1}\exp\left(-\frac{\beta}{x}\right)$	$\frac{np}{\beta}=\sum_{i=1}^{n}\frac{1}{x_i}$ *and* $n\ln\beta-n\psi(p)=\ln\sum_{i=1}^{n}x_i$	$\hat{\gamma}(\mathbf{h})=\frac{1}{2N(\mathbf{h})}\times\sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(\mathbf{h})}\left[Z(\mathbf{s}_i)-Z(\mathbf{s}_j)\right]^2$

Table 4 provide the result on the goodness of fit statistics, the associated value of R^2 and the selected distribution to describe the data for each station. Based on the results for all the distributions and with respect to all goodness of fit methods, all R^2 value are found to be grater than 0.97, indicating that these distributions fit the data well. However, for the purpose of selection of the 'best' distribution, we use the criteria of largest value of R^2. Weibull, Gamma, Erlang, Burr and Inverse Gamma are distributions that are found to be the most suitable for explaining the hourly mean speed in Malaysia. The most frequent distribution selected will be based on the highest number of stations that have been successfully fitted using the particular distribution. Based on results shown in Table 4, GA provides the best fit to wind speed observed at 22 stations, indicating that it is the most frequent selected distribution. The second most frequent selected distribution is BR. Thirteen station was successfully fitted with BR distribution. IGA are the third most frequent selected distribution, as shown by 5 stations. This followed by WE and ER which are found to fit adequately the data observed at 4 and 3 stations respectively. However, LN, EX, RY and IGU fail to fit well the distribution of wind speed at all stations.

Table 2. The parameter estimates for Lognormal, Weibull Rayleigh, Exponent, Gamma and Inverse Gaussian based on maximum likelihood method.

Station	Lognormal		Weibull		Rayleigh	Exponential	Gamma		Inverse Gaussian	
	μ	σ	α	β	Θ	Θ	α	β	μ	λ
1	1.565	0.460	6.611	1.723	4.863	5.871	2.589	2.268	5.871	3.720
2	1.360	0.419	5.395	1.557	4.066	4.798	2.555	1.880	4.798	3.138
3	1.771	0.302	7.647	2.027	5.391	6.755	3.756	1.799	6.755	4.966
4	1.415	0.567	5.902	1.582	4.463	5.282	2.162	2.445	5.281	3.060
5	0.865	0.477	3.377	1.490	2.650	3.205	2.232	1.355	3.025	1.909
6	1.330	0.565	5.612	1.526	4.174	4.879	2.117	2.304	4.789	2.833
7	1.049	0.457	3.991	1.656	2.982	3.545	2.465	1.439	3.545	2.280
8	1.779	0.442	7.933	2.036	5.546	6.930	3.357	2.066	6.931	3.481
9	1.487	0.432	6.124	1.634	4.666	5.444	2.565	2.119	5.444	3.560
10	1.695	0.370	7.308	1.847	5.336	6.462	3.086	2.092	6.463	4.469
11	1.491	0.329	5.843	2.002	4.090	5.162	3.495	1.477	5.161	3.713
12	1.664	0.374	7.126	1.802	5.200	6.302	2.984	2.112	6.302	4.351
13	1.895	0.338	8.848	1.929	6.316	7.807	3.279	2.381	7.807	5.606
14	1.170	0.371	4.316	1.882	3.087	3.816	3.111	1.227	3.816	2.638
15	1.448	0.439	5.874	1.676	4.371	5.217	2.601	2.004	3.508	2.154
16	1.016	0.508	3.905	1.546	2.994	3.508	2.241	1.565	3.491	2.142
17	1.383	0.346	5.311	1.886	3.811	4.691	3.235	1.451	4.691	3.330
18	1.506	0.340	5.960	1.980	4.225	5.265	3.373	1.560	5.264	3.756
19	1.472	0.504	6.124	1.619	4.609	5.461	2.368	2.307	5.461	3.315
20	1.046	0.414	3.938	1.628	2.967	3.497	2.577	1.357	3.497	2.342
21	1.328	0.773	5.774	1.350	4.652	5.283	1.632	3.236	5.283	2.574
22	1.347	0.646	5.711	1.362	4.673	5.197	1.808	2.875	5.197	2.797
23	1.374	0.281	5.089	2.106	3.559	4.497	4.021	1.119	4.497	3.389
24	1.288	0.708	5.455	1.378	4.379	4.967	1.737	2.857	4.966	2.544
25	1.602	0.524	6.926	1.750	5.060	6.165	2.460	2.506	6.165	3.656
26	1.351	0.441	5.236	1.895	3.748	4.644	2.869	1.618	1.351	0.441
27	1.572	0.336	6.366	1.954	4.526	5.622	3.388	1.658	5.622	4.022
28	1.150	0.646	4.660	1.466	3.628	4.206	1.899	2.215	4.206	2.285
29	1.449	0.553	6.094	1.553	4.655	5.457	2.169	2.519	5.457	3.195
30	1.698	0.651	8.135	1.513	6.187	7.185	2	3.636	7.186	3.803
31	1.482	0.381	5.916	1.874	4.250	5.232	3.055	1.712	5.232	3.583
32	1.152	0.374	4.260	1.837	3.078	3.767	3.027	1.245	3.767	2.611
33	1.078	0.276	3.798	2.053	2.670	3.353	3.946	1.066	3.353	2.551
34	1.464	0.353	5.776	1.883	4.145	5.102	3	1.605	5.102	3.599
35	1.522	0.431	6.193	1.895	4.434	5.490	3	1.883	5.490	3.549
36	1.586	0.468	6.714	1.781	4.887	5.962	2.663	2.237	3.962	3.727
37	1.336	0.394	5.218	1.637	3.931	4.631	2.694	1.099	4.631	3.144
38	1.360	0.780	5.944	1.386	4.714	5.421	1.662	3.262	5.421	2.625
39	1.687	0.283	6.998	2.088	4.902	6.174	3.906	1.580	6.174	4.661
40	1.581	0.460	6.675	1.795	4.848	5.924	2.682	2.207	5.924	3.745
41	1.321	0.328	4.976	1.866	3.582	4.392	3.299	1.079	4.392	3.174
42	1.598	0.316	6.399	2.210	4.440	5.663	3.835	1.477	5.663	4.114
43	1.440	0.384	5.746	1.738	4.226	5.086	2.843	1.789	5.086	3.492
44	1.649	0.451	7.124	1.755	5.215	6.326	2.708	2.336	6.326	4.008
45	1.354	0.350	5.199	1.784	3.795	4.596	3.082	1.490	4.596	3.252
46	1.787	0.270	7.667	2.213	5.315	6.767	4.161	1.063	6.767	5.171
47	1.385	0.460	5.537	1.714	4.079	4.917	2.559	1.923	4.917	3.139
48	1.372	0.452	5.549	1.512	4.332	4.960	2.336	2.122	4.960	3.188
49	1.148	0.534	4.524	1.514	3.498	4.055	2.137	1.897	4.055	2.433
50	1.099	0.375	4.054	1.776	2.962	3.589	2.961	1.212	3.589	2.483
51	2.208	0.255	11.59	2.193	8.005	10.143	4.376	2.339	10.237	7.885
52	2.005	0.362	9.872	1.946	6.660	7.846	3.260	2.677	8.729	6.057
53	1.796	0.423	8.226	1.746	5.843	6.838	2.764	2.640	7.296	4.788
54	1.603	0.505	7.024	1.593	5.021	5.591	2	2.717	6.267	3.835
55	1.802	0.334	7.993	1.933	5.629	6.891	3.145	0.483	7.064	5.051
56	1.486	0.365	5.844	2.101	3.492	3.763	3.350	1.543	5.170	3.603
57	1.648	0.499	7.361	1.554	5.466	6.177	2	2.890	6.577	4.042
58	1.702	0.340	7.246	1.929	4.957	5.892	3.379	1.897	6.408	4.552
59	1.818	0.547	8.763	1.613	6.189	6.926	2.239	3.496	7.827	4.596
60	1.813	0.473	8.522	1.690	6.121	7.129	3	3.021	7.577	4.760

Table 3. The parameter estimates for Burr and Inverse Gamma distribution based on maximum likelihood method

St.	Parameter Estimates					St.	Parameter Estimates				
	Burr			Inverse Gamma			Burr			Inverse Gamma	
	q	*a*	*b*	*p*	β		*q*	*a*	*B*	*P*	β
1	11.357	1.809	24.397	2.145	7.976	**31**	5.892	2.059	13.12	2.581	9.249
2	1.608	2.265	5.283	2.464	7.732	**32**	3.739	2.133	7.154	2.756	7.196
3	2.136	2.667	8.798	3.120	15.491	**33**	2.495	2.630	4.751	3.692	9.419
4	152.89	1.589	139.60	1.833	5.608	**34**	3.356	2.211	8.965	2.880	10.36
5	0.797	2.624	2.019	2.427	4.633	**35**	99.73	1.906	69.03	2.115	7.504
6	20.321	1.568	37.40	1.879	5.323	**36**	16.86	1.844	30.31	1.998	7.444
7	3.878	1.913	7.282	2.443	5.648	**37**	0.917	2.813	3.585	2.780	8.738
8	0.797	2.624	2.019	1.075	3.742	**38**	0.797	2.624	2.019	1.409	3.697
9	1.660	2.249	6.130	2.455	8.738	**39**	2.823	2.545	9.403	3.548	16.54
10	2.912	2.246	10.39	2.677	11.96	**40**	74.32	1.808	71.86	2.070	7.752
11	4.041	2.308	9.831	2.942	10.92	**41**	1.349	2.732	4.406	3.175	10.07
12	2.283	2.272	8.752	2.735	11.89	**42**	16.73	2.294	21.42	2.880	11.84
13	7.026	2.081	21.40	3.078	17.25	**43**	1.533	2.400	5.461	2.793	9.753
14	3.878	1.913	7.282	2.653	6.999	**44**	4.397	2.016	13.59	2.070	8.295
15	2.365	1.764	7.434	2.335	7.895	**45**	1.365	2.631	4.613	3.013	9.800
16	3.890	1.764	7.434	2.214	4.717	**46**	6.355	2.415	15.671	3.629	18.76
17	1.660	2.249	6.130	2.930	9.759	**47**	46.86	1.734	50.439	2.227	6.988
18	4.041	2.308	9.831	2.903	10.90	**48**	0.762	2.813	3.300	2.500	7.969
19	4.505	1.837	12.63	1.980	6.564	**49**	5.409	1.671	11.386	2.084	5.073
20	0.873	2.720	2.581	2.727	6.387	**50**	2.887	2.186	5.795	2.784	6.912
21	218.43	1.354	307.9	1.451	3.735	**51**	2.823	2.545	9.403	3.669	28.931
22	223.31	1.366	298.5	1.719	4.810	**52**	7.026	2.081	21.400	2.604	15.778
23	2.283	2.272	8.752	3.413	11.56	**53**	20.32	1.568	37.404	2.331	11.161
24	135.11	1.384	188.2	1.557	3.963	**54**	4.397	2.016	13.594	2.081	7.978
25	213.32	1.763	1.462	1.832	6.839	**55**	6.356	2.423	15.669	2.904	14.668
26	286.75	1.899	102.8	2.106	6.299	**56**	5.892	2.059	13.125	2.602	9.374
27	2.283	2.272	8.752	2.931	11.78	**57**	6.355	2.415	15.671	2.140	8.650
28	161.08	1.471	146.81	1.692	3.868	**58**	4.397	2.016	13.594	2.833	12.895
29	164.33	1.559	160.05	1.900	6.038	**59**	7.026	2.081	21.400	1.857	8.536
30	20.321	1.568	37.404	1.526	5.804	**60**	7.132	2.111	20.899	2.130	10.140

Table 4. The result of goodness of fit tests found based on Kolmogorov Smirnov test, Akaike's Information Criterion, Bayesian Information Criterion and the selected distribution (in **bold**) for each station.

St.	Goodness-of-fit method						St.	Goodness-of-fit method					
	KS	R^2 (%)	AIC	R^2 (%)	BIC	R^2 (%)		KS	R^2 (%)	AIC	R^2 (%)	BIC	R^2 (%)
1	**GA**	99.70	GA	99.70	GA	99.70	**31**	**GA**	99.65	GA	99.65	GA	99.65
2	**BR**	99.52	GA	98.98	GA	98.98	**32**	**GA**	99.62	GA	99.62	GA	99.62
3	**BR**	99.60	GA	99.57	GA	99.57	**33**	**BR**	99.43	BR	99.43	BR	99.43
4	**BR**	99.34	GA	99.20	GA	99.20	**34**	**ER**	99.09	GA	99.01	GA	99.01
5	**IGA**	98.00	IGA	98.00	IGA	98.00	**35**	**ER**	99.38	WE	99.24	WE	99.24
6	**WE**	99.34	GA	99.30	GA	99.30	**36**	**BR**	99.54	BR	99.54	BR	99.54
7	**GA**	98.75	GA	98.75	GA	98.75	**37**	**IGA**	99.72	IGA	99.72	IGA	99.72
8	**WE**	97.87	RY	98.07	WE	97.87	**38**	**GA**	96.97	WE	97.32	WE	97.32
9	**BR**	99.34	GA	98.91	GA	98.91	**39**	**GA**	99.59	GA	99.59	GA	99.59
10	**GA**	99.79	GA	99.79	GA	99.79	**40**	**GA**	99.84	WE	99.72	WE	99.72
11	**GA**	99.76	GA	99.76	GA	99.76	**41**	**IGA**	99.61	IGA	99.61	IGA	99.61
12	**BR**	99.21	GA	99.09	GA	99.09	**42**	WE	99.92	**BR**	99.94	BR	99.94
13	**GA**	98.74	GA	98.74	GA	98.74	**43**	**IGA**	99.42	GA	99.38	GA	99.38
14	**GA**	99.73	GA	99.73	GA	99.73	**44**	BR	99.67	**GA**	99.71	GA	99.71
15	**GA**	99.18	GA	99.18	GA	99.18	**45**	BR	99.46	**GA**	98.85	GA	98.85
16	**BR**	98.69	GA	98.31	GA	98.31	**46**	**BR**	99.25	BR	99.25	BR	99.25
17	**GA**	99.63	GA	99.63	GA	99.63	**47**	**GA**	99.53	GA	99.53	GA	99.53
18	**GA**	99.70	GA	99.70	GA	99.70	**48**	**IGA**	99.61	IGA	99.61	IGA	99.61
19	**GA**	99.86	GA	99.86	GA	99.86	**49**	**BR**	98.44	GA	98.33	GA	98.33
20	BR	98.12	**IGA**	98.52	IGA	98.52	**50**	**GA**	99.30	GA	99.30	GA	99.30
21	**WE**	97.79	GA	97.70	GA	97.70	**51**	**GA**	98.66	GA	98.66	GA	98.66
22	**GA**	99.16	GA	99.16	GA	99.16	**52**	WE	93.89	**GA**	94.93	GA	94.93
23	WE	99.09	**GA**	99.90	GA	99.90	**53**	WE	96.58	**GA**	97.16	GA	97.16
24	**BR**	98.77	GA	98.66	GA	98.66	**54**	**ER**	97.24	GA	92.46	GA	92.46
25	**WE**	99.86	WE	99.86	WE	99.86	**55**	**BR**	98.33	WE	97.82	WE	97.82
26	WE	99.53	**BR**	99.54	WE	99.53	**56**	**RY**	81.55	WE	73.11	WE	73.11
27	**GA**	99.29	GA	99.29	GA	99.29	**57**	**ER**	98.99	GA	96.37	GA	96.37
28	**BR**	98.36	GA	98.10	GA	98.10	**58**	**RY**	96.67	GA	96.04	GA	96.04
29	**GA**	98.94	GA	98.94	GA	98.94	**59**	**ER**	96.26	GA	92.90	GA	92.90
30	**ER**	99.66	WE	99.65	WE	99.65	**60**	**GA**	96.97	GA	96.97	GA	96.97

4 Semivariogram

Semivariogram is a tool which is often used to investigate spatial dependence of the data before spatial prediction is done. Let $Z(s_i)$ denote the mean speed for the *i*-th station given a particular choice of the wind speed distribution. Semivariogram reconstruct the properties of auto-covariance for the spatial process in *d* dimension denoted as $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, where **s** is the location at which attribute Z is observed. Semivariogram is defined as

$$\begin{aligned}\gamma(\mathbf{s}_i - \mathbf{s}_j) &= \frac{1}{2} Var\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right] \\ &= \frac{1}{2}\left\{Var\left[Z(\mathbf{s}_i)\right] + Var\left[Z(\mathbf{s}_i)\right] - 2Cov\left[Z(\mathbf{s}_i), Z(\mathbf{s}_i)\right]\right\}\end{aligned} \quad (1)$$

Semivariogram function that depends upon separation vector only through its length $\left\|\mathbf{s}_i - \mathbf{s}_j\right\|$ is called isotropic; if not, it is anisotropic. A valid semivariogram function can also be constructed from a valid covariance function, where a valid covariance function is a positive-definite function, such that $\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(\mathbf{s}_i - \mathbf{s}_i) \geq 0$ for any set of real number $a_1, \cdots, a_k$, where $C(\mathbf{s}_i - \mathbf{s}_i)$ is a covariance function. Parametric forms that are available as candidates for semivariogram are linear, spherical, exponential, wave, rational quadratic, etc (see, Appendix). The candidate model is chosen based on the "closeness" between the theoretical semivariogram and the empirical semivariogram which is calculated by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} \left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right]^2, \quad (2)$$

where $\mathbf{h} = \left\|\mathbf{s}_i - \mathbf{s}_j\right\|$ is a distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$ and $N(\mathbf{h})$ denotes the set of pairs of locations at distance $\mathbf{h}$.This measure of "closeness" may be based on mean square error.

5 Kriging method

Kriging is a geostatistical technique used in spatial interpolation of geostatistical data. It has an ability of incorporating information about regional and local trends [7]. The general mathematical model for kriging is given by

$$\mathbf{Z}(\mathbf{s}) = \mu\mathbf{1} + \mathbf{e}(\mathbf{s}) \quad (3)$$

where $\mathbf{Z}(\mathbf{s}) = \left[Z(s_1), Z(s_2) \cdots, Z(s_n)\right]'$ denote the mean of wind speed for each *i-th* station, where *i=1,2,...n*. μ is unknown and assumed to be constant, $\mathbf{e}(\mathbf{s}) \sim (0, \Sigma)$, in which Σ is the information about the spatial dependence of the data that needs to be specified from the semivariogram function. The predicted value of mean speed at the pivot point $\mathbf{s}_0$, denoted as $p_k(\mathbf{Z}; s_0)$ can be expressed in term of a linear combination of $Z(\mathbf{s}_i)$ which is given by

$$p_k(Z; s_0) = \lambda_0 + \lambda' Z(s) \quad (4)$$

where λ_0 and the element of vector $\boldsymbol{\lambda} = [\lambda_1, \dots \lambda_n]'$ are unknown coefficients. In order to get some reasonable estimates of $\hat{Z}(\mathbf{s}_0)$, the unbiasedness constraint such that $E\left[p(\mathbf{Z}; \mathbf{s}_0)\right] = E\left[Z(s_0)\right]$, or equivalently, $E\left[\lambda_0 + \boldsymbol{\lambda}'\mathbf{Z}(\mathbf{s})\right] = E\left[Z(s_0)\right]$, which implies that $\lambda_0 + \mu(\boldsymbol{\lambda}'\mathbf{1} - 1) = 0$ is required. Since this must hold for every μ, it also holds for $\mu = 0$. Thus, the unbiasedness constraint require that $\lambda_0 = 0$ and $\boldsymbol{\lambda}'\mathbf{1} = 1$. From that point, we know that $\boldsymbol{\lambda} = [\lambda_1, \dots \lambda_n]'$ will be chosen based on minimizing

$$E\left[\left(\lambda' Z(s) - Z(s_0)\right)^2\right] \qquad \textit{subject to} \quad \lambda'\mathbf{1} = 1$$

It can be accomplished by using Lagrange multiplier *m* where

$$\arg\min_{\lambda,\lambda} Q = \arg\min \; E\left[\left(\lambda' Z(s) - Z(s_0)\right)^2\right] - 2m\left(\lambda'\mathbf{1} - 1\right)$$

Expanding the above function by putting

$$Var\left[Z(s_0)\right] = \gamma(s_0) \text{ and } \Sigma = \Gamma = \left[\gamma(s_i - s_j)\right]$$

yield a solution given by

$$Q = -\boldsymbol{\lambda}'\boldsymbol{\Gamma}\boldsymbol{\lambda} + 2\boldsymbol{\lambda}'\gamma(s_0) - 2m(\boldsymbol{\lambda}'\mathbf{1} - 1)$$

where $\boldsymbol{\Gamma} = \left[\gamma\left(s_i - s_j\right)\right]$ is the information about semivariogram of the data. Differentiation with respect to $\boldsymbol{\lambda}$ and m, setting to zero and solving the simultaneous equation yields

$$\lambda' = \left(\gamma\left(s_0\right) + \mathbf{1}\frac{1-\mathbf{1}'\Gamma^{-1}\gamma\left(s_0\right)}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}\right)'\Gamma^{-1} \tag{5}$$

and

$$m = -\frac{1-\mathbf{1}'\Gamma^{-1}\gamma\left(s_0\right)}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} \tag{6}$$

Thus, $p_k\left(Z;s_0\right) = \lambda' Z\left(s\right)$ is the predictor of Kriging method with the variance of prediction is $\sigma_k^2 = \boldsymbol{\lambda}\gamma\left(s_0\right) + m$. However, if none of the semivariogram model is found suitable for modelling the empirical semivariogram, inverse distance weighting method is found to be more reasonable in estimation of random field instead (for detail see [8,17,18])

6 Inverse Distance Weighting Method (IDW)

IDW is an alternative method for spatial estimation of random field when the spatial dependence of the data could not be described well by the semivariogram. IDW is a weighted average interpolator which can either be an exact or a smoothing interpolator. In IDW, data is weighted during interpolation such that the influence of one point relative to another decline as the distance increased. The value of $p\left(\mathbf{Z};s_0\right)$ at the pivot point s_o can be predicted by using a weighted mean of the available measurements through the expression

$$p\left(\mathbf{Z};s_0\right) = \frac{\sum_{i=1}^{n} W\left(s_i, s_0\right) Z\left(s_i\right)}{\sum_{i=1}^{n} W\left(s_i, s_0\right)} \tag{7}$$

where $Z\left(s_i\right)$ is the observed data for station i, $\left(s_i, s_0\right)$ is the distance between the station i to the pivot point s_o and $W\left(s_i, s_0\right)$ is the weighting factor which is decreasing as the distance increase. The value of $p\left(\mathbf{Z};s_0\right)$ decreases with the distance following a quadratic or exponential law [10].

7 Result and Discussion

After determining the distribution of wind speed for each station, the spatial interpolation is carried out on the mean value of each selected distribution. However, before determining the spatial distribution of wind speed, an assessment is made on the semivariogram of the data. Mean square error (MSE) is used to make an assessment on the suitability of the particular semivariogram model. Although MSE for the fitted model may indicate a minimum value, it does not necessarily imply that the fitted model will have a similar form as the theoretical model. Thus, in study we use our subjective assessment to select the best semivariogram model. The fitted semivariogram with relatively small MSE and found to satisfy the particular form of semivariogram will be chosen as the best fitted model. Based on Table 5 which shows the value of MSE for fitted semivariogram model, it is found that Linear, Power and Wave semivariogram are chosen for modelling the spatial dependence in the peninsula data. While for the data of East Malaysia, semivariogram model such as Exponential, Gaussian, Linear, Power, Quadratic, Rational Quadratic and Wave are found to a have relatively small value of MSE. Based on the subjective assessment after making a comparison for each fitted model with the theoretical semivariogram shape, linear model has been chosen as the best semivariogram model for the theoretical mean of wind speed in Peninsular Malaysia. Figure 1 shows a fitted linear semivariogram with estimated nugget and scale effect of 0.752 and 0.688 respectively, implying that the peninsula data is random. Figure 1 also shows that the linear semivariogram cannot adequately fit the data. This implies that there is a lack of spatial dependence of wind speed in Peninsular Malaysia region. The presence of nugget effect for fitted linear semivariogram indicates the roughness of a data. Based on the properties of semivariogram discussed here, we can conclude that there is no clear pattern on how the wind speed at a particular location is influenced by the wind speed at a neighbouring location. However, we suggest that a more comprehensive analysis involving more stations need to be conducted in the future in order to get a better understanding about semivariogram of the data as well as the spatial dependence of wind speed in Peninsular Malaysia. Thus, instead of using kriging interpolator, inverse distance weighting method is used to avoid the problem of over or underestimation. The result found based on the analysis for data of Sabah and Sarawak is slightly different from those found for the peninsular. As shown in Figure 2, it is found that the fitted power semivariogram with estimated scale and power of 0.662 and 0.272 respectively has been chosen as the best model for describing the spatial dependence in the data. The absence of nugget effect indicates that the wind speed data is smooth.

Table 4. List of geographical coordinate and theoretical mean for each station

St	Latitude	Longitude	$Z(s_i)$	St	Latitude	Longitude	$Z(s_i)$
1	N01° 28.225	E103° 53.637	5.872	**31**	N05° 18.455	E103° 07.213	5.230
2	N04° 16.260	E103° 25.826	4.854	**32**	N01° 27.308	E110° 29.498	3.769
3	N05° 23.470	E100° 23.213	6.765	**33**	N01° 14.425	E111° 27.629	3.359
4	N01° 33.734	E110° 23.329	5.301	**34**	N05° 21.528	E100° 17.864	4.815
5	N03° 15.702	E101° 39.103	3.247	**35**	N04° 15.016	E117° 56.166	5.649
6	N02° 15.510	E102° 10.364	5.056	**36**	N06° 08.218	E100° 20.880	5.972
7	N03° 58.238	E102° 20.863	3.547	**37**	N04° 12.038	E100° 39.841	4.909
8	N04° 37.781	E101° 06.964	7.028	**38**	N05° 19.980	E115° 14.315	5.421
9	N05° 23.890	E100° 24.194	5.505	**39**	N02° 12.789	E102° 14.055	6.171
10	N02° 49.246	E101° 48.877	6.456	**40**	N02° 03.715	E102° 35.587	5.919
11	N03° 00.620	E101° 24.484	5.162	**41**	N03° 41.267	E101° 31.466	4.630
12	N03° 49.138	E103° 17.817	6.333	**42**	N04° 33.155	E101° 04.856	5.667
13	N03° 57.726	E103° 22.955	7.807	**43**	N02° 43.418	E101° 58 .105	5.439
14	N03° 06.612	E101° 42.274	3.817	**44**	N03° 19.592	E101° 15.532	6.331
15	N05° 37.886	E100° 28.189	5.212	**45**	N05° 20.313	E116° 09.769	4.648
16	N01° 29.815	E103° 43.617	3.475	**46**	N05° 51.865	E118° 05.479	6.781
17	N04° 53.940	E100° 40.782	4.694	**47**	N01° 29.068	E103° 41.064	4.936
18	N06° 09.520	E102° 17.262	5.262	**48**	N02° 55.915	E101° 40.909	4.921
19	N06° 09.520	E102° 15.059	5.463	**49**	N03° 06.376	E 101° 43.072	4.073
20	N02° 59.645	E101° 44.417	3.690	**50**	N02° 00.875	E112° 55.640	3.590
21	N04° 35.880	E103° 26.096	5.295	**51**	N02° 27.000	E103° 103.00	10.235
22	N03° 06.287	E101° 33.368	5.461	**52**	N06° 10.000	E102° 17.000	8.727
23	N02° 18.856	E111° 49.906	4.507	**53**	N04° 28.000	E101° 22.000	6.297
24	N03° 10.587	E113° 02.433	4.985	**54**	N03° 47.000	E103° 13.000	5.434
25	N04° 25.456	E114° 00.731	6.168	**55**	N05° 23.000	E103° 06.000	6.796
26	N02° 07.992	E111° 31.351	4.647	**56**	N06° 29.000	E100° 16.000	4.377
27	N05° 53.623	E116° 02.596	5.617	**57**	N06° 12.000	E100° 24.000	5.780
28	N04° 45.529	E115° 00.813	4.212	**58**	N04° 34.000	E101° 06.000	6.212
29	N06° 19.903	E099° 51.517	5.464	**59**	N05° 18.000	E100° 16.000	6.992
30	N06° 25.424	E100° 11.046	7.272	**60**	N02° 16.000	E102° 15.000	7.589

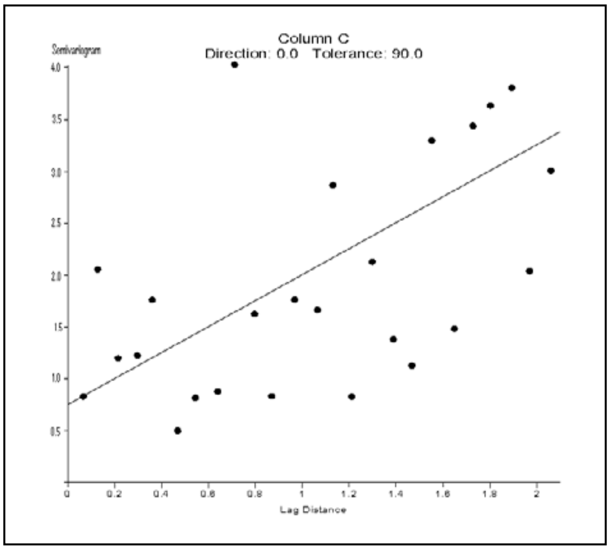

Fig 1. The fitted linear semivariogram model for mean wind speed in Peninsular Malaysia

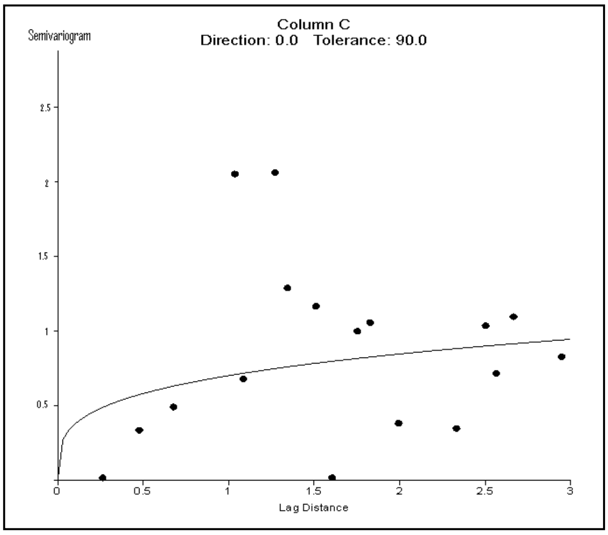

Fig 2. The fitted power semivariogram model for mean wind speed in East Malaysia.

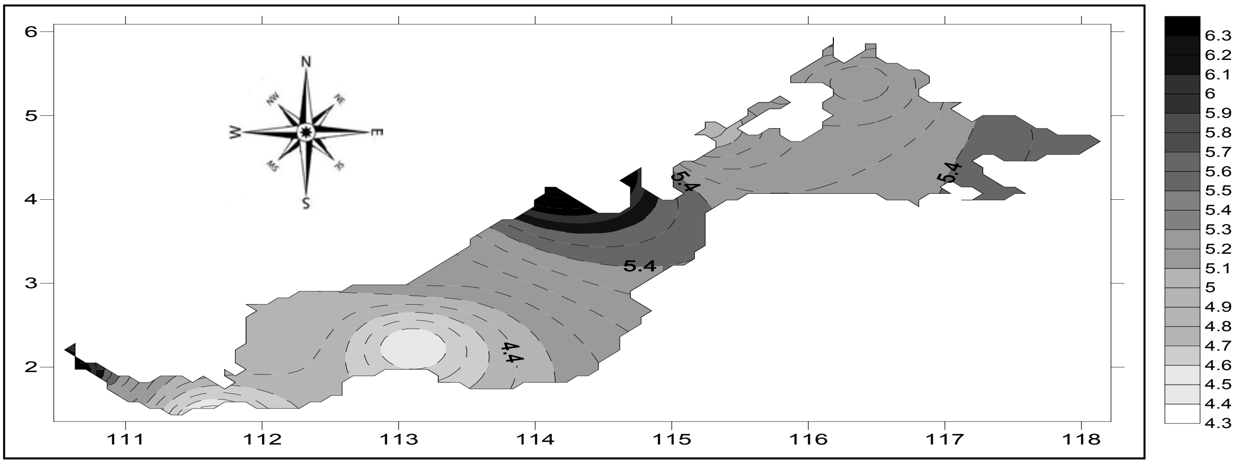

Fig 3. Map of wind speed in East Malaysia

The information about the semivariogram model discussed above can be used for kriging interpolation. Once the properties of spatial dependence of the data have been determined, it is possible to perform the spatial prediction in the random field. Table 4 shows a list of geographical coordinate and theoretical mean of the wind speed for each station, while Figure 3 shows the result of spatial prediction for the theoretical mean wind speed, where the solid line represents the boundary for region with wind speed greater than 6 km/hour while dotted line indicate the boundary of regions with wind speed less than 6 km/hour. Most regions in East Malaysia have the wind speed in the range of 4.3 to 6.3 km/hour. However, northeast and southwest region of Sarawak state was found to have the highest wind speed as compared to the other regions. The average wind speed for the southern region is about 6 km/hour. This indicates that the northeast and southwest region of Sarawak has the highest potential of producing energy. Assessment on wind potential for Peninsular Malaysia also follows the same rule. Figure 4 shows a rough map of spatial distribution for the theoretical mean in Peninsular Malaysia. The map shows that the mean wind speed across the peninsula is found to be in range between 3 to 10 km/hour. As shown in Figure 4, we found that the wind speed for southeast, northeast and northwest regions in the peninsula is greater when compared to all the other regions. Based on the results discussed above, northeast, northwest and southeast region in Peninsular Malaysia will be chosen as the region that could be explored in details in order to find a specific area that has a good wind regime and a large tendency to develop wind energy.

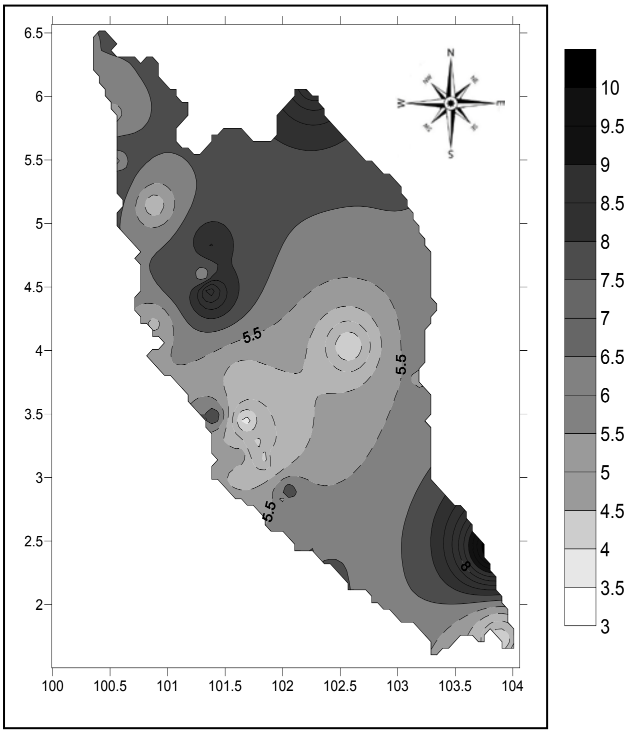

Fig 4. Map of mean wind speed in Peninsular Malaysia

8 Conclusion

Gamma distribution is the distribution that most frequently found adequate to describe the distribution of wind speed at 50 stations considered in this study. In order to gain some insight about the potential of wind power in Malaysia, the theoretical mean is considered as an important index that have been derived from each selected statistical distribution. The mean wind speed is used to evaluate the potential of wind energy based on kriging and inverse distance weighting method in order to get the rough information about the areas with a good wind regime and have a capability to generated sustainable wind power. The mappings of the theoretical mean of wind speed over the Peninsular Malaysia indicates that region in the northeast, northwest and southeast are found to be the most potential to be explored in detail for the purpose of generating wind energy. While for East Malaysia, northeast and southwest region of Sarawak is found as the best region to be investigated in the future for developing wind energy. A more comprehensive analysis need to be conducted in the future, involving more stations to get a better map of wind speed in Malaysia.

Appendix:

Table A1. List of Semivariogram models

Model	Variogram
Exponential	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left(1-\exp(-h)\right) & if\ \ h>0\\ 0 & otherwise\end{cases}$
Gaussian	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left(1-\exp(-h^2)\right) & if\ \ h>0\\ 0 & otherwise\end{cases}$
Linear	$\gamma(h)=\begin{cases}\tau^2+\sigma^2 h & if\ \ h>0\\ 0 & otherwise\end{cases}$
Logarithmic	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left[\log_e(h)\right] & if\ \ h>0\\ 0 & otherwise\end{cases}$
Pentaspherical	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left(1.87h-1.25h^3+0.375h^5\right) & if\ \ h>0\\ 0 & otherwise\end{cases}$
Power	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left\lfloor h^n\right\rfloor & if\ \ h>0\\ 0 & otherwise\end{cases}$ where 0< n< 2
Quadratic	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left(2h-h^2\right) & if\ h<1\\ \sigma^2 & if\ h\geq 1\end{cases}$
Rational Quadratic	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left[\dfrac{h^2}{1+h^2}\right] & if\ \ h>0\\ 0 & otherwise\end{cases}$
Spherical	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left[1.5h-0.5h^3\right] & if\ \ h<1\\ \sigma^2 & if\ \ h\geq 1\end{cases}$
Wave	$\gamma(h)=\begin{cases}\tau^2+\sigma^2\left(1-\dfrac{\sin(h)}{h}\right) & if\ \ h>0\\ 0 & otherwise\end{cases}$

Acknowledgement

The authors are indebted to the staff of the Department of Environment for providing data to make this paper possible. This research would not have been possible without the sponsorship from Universiti Kebangsaan Malaysia and Ministry of Higher Education, Malaysia (grant number UKM-GUP-TK-08-16-061 and UKM-GUP-2011-213).

References:

[1] A.M. Razali, R. Zainal Abidin, A. Zaharim, K.Sopian, Fitting Statistical Distributions to Wind Speed Data. *Proceedings of the 4th IASME/WSEAS International Conference on Energy, Environment, Ecosytems, and Sustainable Development (EEESD'o8)*, Algarve, Portugal, 2008.

[2] A. Zaharim, S. K. Najid, A. M. Razali, K. Sopian, Analyzing Malaysian Wind Speed Data Using Statistical Distribution. *Proceedings of the 4th IASME/ WSEAS International Conference on Energy and Environment (EE '09)*, Cambridge, United Kingdom, 2009.

[3] A. M. Razali, M. S. Sapuan, K. Ibrahim, A. R. Ismail, A. Zaharim, K. Sopian,. Fitting of Weibull Distribution to Study Wind Energy Potential in Kuala Terengganu, Malaysia. *Proceedings of the 4th IASME/ WSEAS International Conference on Applied Mathematics (MATH '09)*, Canary Island, 2009.

[4] H. C. Ong, T. M. I. Mahlia, H. H. Masjuki, A review on energy scenario and sustainable energy in Malaysia,*Renewable and Sustainable Energy Reviews*, 2011, 15, pp 639-647.

[5] M. R. Islam, R. Saidur, N. A Rahim, Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function. *Renewable Energy*, 2011, 36, pp 985-992.

[6] B. Ali, K. Sopian, C. H. Yen, S. Mat, A. Zaharim, Key Success Factors in Implementing Renewable Energy Programme. . *Proceedings of the 4th IASME/WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development (EEESD'o8)*, Algarve, Portugal, 2008.

[7] K. Sopian, M.Y. Hj. Othman, A.Wirsat, The wind energy potential in Malaysia. *Renewable Energy* , 1995, Vol 6, No 8, pp1005-1016.

[8] O. Schabenberger, C.A. Gotway, *Statistical Methods for Spatial Data Analysis*, Chapman & Hall, Florida, 2005.
[9] A. Oztopal, Artificial neural network approach to spatial estimation of wind speed data. *Energy Conversion and Management,* 2006, 47, pp 395-406.
[10] M. Cellura, G. Cirrincione, A. Marvuglia, A., Miraoui, Wind speed spatial estimation for energy planning in Sicily: Introduction and statistical analysis. *Renewable Energy* 2008, 33, pp 1237-1250.
[11] J. A. Carta, P. Ramirez, S. Velazquez, A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews, 2009*. 13, pp 933-955.
[12] E. C Morgan,. M. Lackner, R. M.Vogel, L. G. Baise, Probability distributions for offshore wind speeds. *Energy Conversion and Management , 2011*. 52, pp 15-26.
[13] L. Zhou, E. Erdem, G. Li, J. Shi, Comprehensive evaluation of wind speed distribution model: A case study for North Dakota sites. *Energy Conversion and Management, 2010*. 51, pp1449-1458.
[14] M. Evans, N. Hastings, B. Peacock, *Statistical Distribution*s: Second Edition. John Wiley. & Son, New York, 1993.
[15] Development Core Team. R, 2008: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http: //www.R-project.org.
[16] R. H. Holdaway, Spatial modeling and interpolation of monthly temperature using kriging. *Climate Research* 1996; Vol 3, pp 215-225
[17] B. Banerjee, B. P. Carlin, A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data,* Chapman & Hall, Florida, 2004.
[18] N. Cressie, *Statistics for Spatial Data*, Revised Edition. John Wiley & Sons, New York,1993.