Journal of Geospatial Surveying (2022) 2:2
DOI: http://doi.org/10.4038/jgs.v2i2.42

ORIGINAL ARTICLE

Integrating the Spatial Autocorrelation Structure of Irregularly Distributed Wind Measurement Stations to Model Wind Patterns Using R

Welikanna, D. R.*

Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Sri Lanka

ARTICLE INFO

Article history:
Received 27 August 2022
Received in revised form 21 October 2022
Accepted 22 October 2022
Available online 27 October 2022

Keywords:
Wind Speed
Wind Direction
Wind patterns
Variograms
Kriging
R 4.0.2

ABSTRACT

Wind behavior in Sri Lanka, mainly exposed to two major monsoon seasons, is complex to model spatially. The acute changes in the topography further make it difficult to map the wind patterns. Irregularly distributed wind measurements are the main source of data available to interpret and model this dynamic geographic phenomenon. Sri Lanka is in need of an accurate and timely wind map in order to support wind energy resource planning and to enable weather prediction systems to be more accurate. Strong and continuous winds in the southern part of the island encourage wind harvesting for energy. Wind energy is arguably the most affordable per megawatt-hour renewable energy source, and it is growing nearly as quickly as conventional energy generation techniques in the country. The objective of this paper is twofold. First is to use the spatial autocorrelation structure of 23 data points using diurnal data measured every 10 minutes on an irregularly distributed grid to construct continuous wind speed fields for the 4 monsoon seasons. This has been done for the year 2015 as a reference year, where a complete set of data was acquired. Secondly, it was intended to provide the reader how R opensource with advanced statistical modelling capabilities and graphics could be utilized conveniently for vague geographic phenomena analysis. The spatial autocorrelation of the wind speed and the direction was determined in the beginning by using the semi-variograms. Candidate semi-variogram models, especially spherical and exponential models, were tested to be more suitable for the observations. Ordinary Kriging, which uses the samples in the local neighborhood to avoid the mean estimation, was used to perform the spatial interpolations Kriging. For each month of the year 2015, four weekly wind maps were generated, both for the direction and speed. These 48 wind maps were further analyzed for the wind patterns. The changes in the morning and evening wind patterns, and the significant wind pattern changes along the coastal areas and the central highlands are extracted and presented in the study. The generated maps could be useful in identifying potential areas for wind farming. By feeding the data in a time series to the presented model, it would be possible to deduce wind pattern changes in Sri Lanka on a yearly basis. As a whole, the paper presents a reasonable model that could be used to model a diverse set of vague phenomena with an understanding of the uncertainties involved.

1 Introduction

Climate can be considered as a sum of atmospheric elements and their variation along time. Those main constituent elements contributing to the climate are solar radiation, temperature, humidity, clouds and precipitation (type, frequency, and amount), atmospheric pressure, and wind (speed and direction). Therefore, wind becomes one of the main components in the determination of the climatic conditions pertaining in a certain time frame. Wind is an extremely dynamic Geographic field. They are highly discontinuous. Mapping and predicting wind behavior and the directions is a particularly challenging task due to the difficulty of estimating uncertainties in the predictions as well as the discrete distribution of the measuring stations.

* Corresponding author: D.R. Welikanna, Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Sri Lanka.

E-mail address: drw@geo.sab.ac.lk (D.R. Welikanna).

Geostatistical approaches provide a feasible platform to map and predict wind patterns using ground-based measurement data. In doing so, it considers the spatial autocorrelation structure to represent the systematic spatial variation of the variable in regard to a given spatial locality. This mapping and predictions on unsampled locations are termed "wind modelling" in this paper.

A windsock is a giant sock of conical shape used in airports with a calibrated chart (Fig. 1, top) that could be used to interpret wind speed and direction arbitrarily. A cup anemometer, which typically consists of three or four conical or hemispherical cups set symmetrically around a vertical spindle, is used to measure wind speed. The spindle rotates as a result of the wind blowing into the cups. Standard instruments include cups whose construction makes it possible for the rate of rotation to be roughly proportionate to the wind speed. Anemometers are calibrated in a wind tunnel every five years at the most in order to spot any deviations from the manufacturer-specified relationship between spindle rotation and wind speed (Fig. 1, bottom left). On top of calibration adjustments, wind speed is measured.

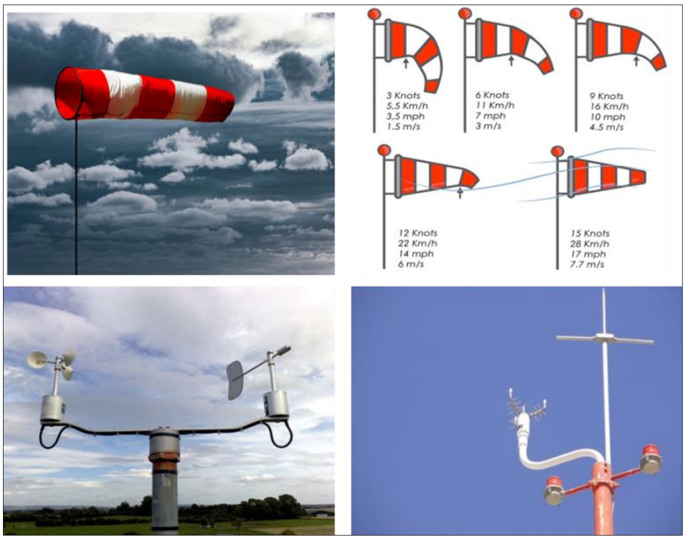

Fig. 1: Wind measuring, top windsocks and the readings, bottom left cup anemometer, bottom right sonic anemometer (*Source*: https://www.metoffice.gov.uk/).

A heated sonic anemometer, which has no moving parts, is used to measure wind in places where it is extremely windy, as on top of mountains. Two transducers at the ends of thin arms are used in the device to measure the speed of acoustic impulses being conveyed between them. An estimation of wind speed and direction can be obtained by combining measurements from two pairs of transducers. The average speed and direction over the ten minutes preceding the reporting time serve as a better indicator of the total wind severity. Calculating mean wind over different averaging intervals is also possible. A vane that consists of a slender horizontal arm holding a vertical flat plate with its edge towards the wind at one end and a balance weight that also acts as a pointer at the other end is used to measure the direction of the wind. The arm is supported by a vertical spindle that is positioned on bearings and may move freely in the wind. Each wind vane and anemometer are fastened to a horizontal supporting arm. To get the wind intensity measured with precision and accuracy, it is possible to average speed and direction at an interval of 10 minutes. Mean wind values averaged over different periods could also be used for the purpose. Surface mean wind speeds in the range 34–40 knots averaged in a period of 10 minutes is termed as a “gale”. Different forms as “severe gale”, “storm”, and do on, results with winds reaching over 41 knots (Department of Meteorology, United Kingdom). Sri Lanka is an island which is positioned within the tropics between 5° 55' and 9° 51' North latitude and between 79° 42' and 81° 53' East longitude, and holds a tropical climate. Due to the tropical position of the island, it is not affected by seasonal variations due to the ecliptic of the sun. Though this is the case if we consider temperature variations on the island, the climate is largely controlled by the seasonal winds. If closely observed, the main element that determines the climate of Sri Lanka is its native wind. Of these seasonal wind regimes, two stand out as the Southwest and Northeast monsoons, which have regional scale winds. The climate experienced during a 12-month period in Sri Lanka can be characterised by up to 4 distinct climate seasons depending mainly on the seasonal winds experienced. The first inter-monsoon season (March and April) brings in thunderstorm-type rain, particularly during the afternoons or evenings. The distribution of rainfall during this period shows that the entire Southwestern sector of the hill country is receiving heavy rainfall, with a localised area on the south-western slopes experiencing rainfall in excess of 700 mm. Windy weather during this monsoon cools off the heat of the first inter-monsoon season. Southwest monsoon rains are experienced at any time of the day and night, mainly focused on the southern part of the country. The second inter-monsoon season (October to November) brings in thunderstorm-types of rain, particularly during the afternoon or evening. Under such conditions, the whole country experiences strong winds with widespread rain, sometimes leading to floods and landslides. The Northeast monsoon season (December and February) is characterised by dry and cold winds blowing from the Indian landmass, creating a cooling effect all over the island. Cloud-free skies provide days full of sunshine and pleasant, cool nights. During this period, the highest rainfall figures are recorded in the northern and eastern slopes of the hill country (Met-SL, 2016). Wind as a hazy spatial phenomenon, particularly in its intensity.

Wind, as mentioned earlier, is a vague and highly dynamic phenomenon occurring on top of a spatial grid along time. Starting from the most probable value (MPV) for the wind intensity estimation from the measured data, the handling of a large set of wind data is challenging. The amount of measurement density with respect to the time is very high for the wind speed and direction. This makes it very difficult to decide approaches to estimate and predict wind at unsampled locations. Sri Lanka, with its dynamic topography and the strong wind regime system by way of monsoons, provides a difficult case to model the wind patterns. Yet the behavior of the winds during these time periods are generally know from years of observations and experiences. Thus, in this study, we attempt to use a combination of semivariogram, kriging, and stochastic simulation with a comprehensive all-year wind data set to predict and map the wind patterns belonging to the major monsoon seasons. In general, this study provides an insight into vague spatial data (spatially autocorrelated) modelling (eg. permeability, sediments, air pressure,) in an irregular lattice structure. One main limitation of the study is the difficulty of accessing the wind data in Sri Lanka and the incompleteness of the datasets available for different years. Furthermore, it should also be noted that more than 50,000 records belonging to the year 2015 were manually extracted from the data portals of the Meteorological Department of Sri Lanka.

2 R Open Source

R is a free software environment for statistical computing and graphics (Crawley, 2005; Crawley, 2007; Dalgaard, 2002; Pebesma and Bivand, 2005; Ripley, 2001). Furthermore, it is an integrated suite of software facilities for data manipulation, calculation, and graphical display. It includes an effective data handling and storage facility; a suite of operators for calculations on arrays, in particular matrices; a large, coherent, integrated collection of intermediate tools for data analysis; graphical facilities for data analysis and display either on-screen or on hardcopy; and a well-developed, simple, and effective programming language which includes conditionals, loops, user-defined recursive functions, and input and output facilities. While it is a GNU project, R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, remote sensing data processing, GIS, variogram modelling, spatial data handling, and graphical techniques (www.r-project.org). R Packages provide more power in data handling; they are collections of functions and data sets developed by the community. They increase the power of R by either improving existing base R functionalities or by adding new ones. A set of advanced R packages specifically used in this study is shown below in Table 1.

Table 1: R Packages additional to basic packages employed in the study.

R Package	Usage
rgeos	Interface to Geometry Engine
maptools	Manipulating and reading geographic data, in particular 'ESRI Shapefiles'
rgdal	Access to projection/transformation operations
sp	Classes and Methods for Spatial Data
rgoogleMaps	Use the google map as a background image
ggmap	Spatial Visualization with ggplot2
ggplot2	System for 'declaratively' creating graphics
latticeExtra	Powerful and elegant high-level data visualization system
ggrepel	Automatically Position Non-Overlapping Text Labels
gstat	Variogram modelling; simple, ordinary and universal point or block (co)kriging; spatio-temporal kriging; sequential Gaussian or indicator (co)simulation; variogram and variogram map plotting utility functions
dplyr	Consistent tool for working with data frame like object
scales	Graphical scales map data to aesthetics
magritr	Provides a mechanism for chaining commands

3 Study Area and Data Preparation

However, though wind is primarily an important climatic parameter; Sri Lanka does not have timely wind maps to realize the changes in patterns. But in practice, the study area, especially in the monsoon seasons, provides a reasonable understanding of the temporal patterns identified in the past years. Wind measuring stations are located as shown in Fig. 2.

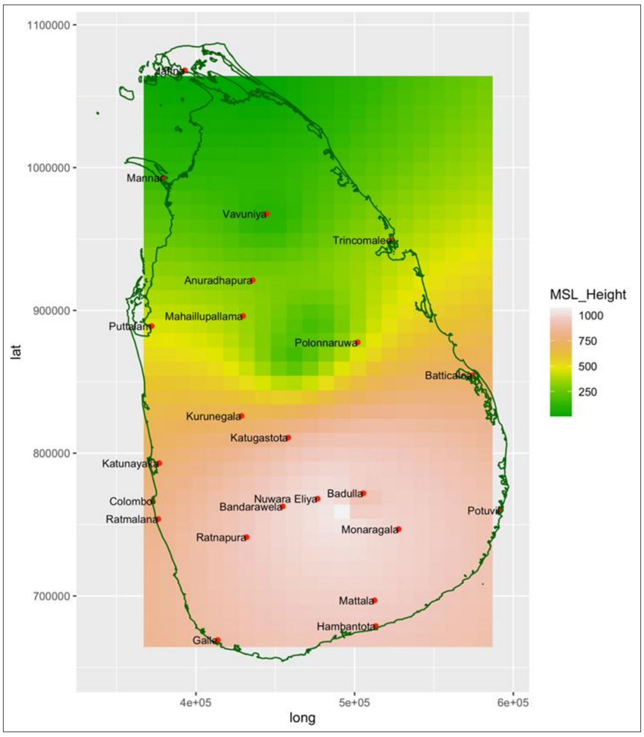

Fig. 2: Wind measuring stations (23- Stations) of Sri Lanka, represented on the topographic variation modelled using reduced levels of the fundamental BM with IDW interpolation.

Fig. 2 shows the wind station location with respect to the Mean Sea Level (MSL) variation of Sri Lanka using 22 Fundamental Bench Marks from the Survey Department of Sri Lanka. Inverse Distance Weighting (IDW) interpolation has been performed to construct the topography map. The irregular distribution of the wind measuring stations not covering the island along a regular grid is a concern in wind predictions and the mapping process. In this study weekly wind measurement values for the twenty-three stations recorded by the Sri Lanka Meteorological Department for the year 2015 have been used. The original data sets contain daily wind measurements using GPS-positioned cup anemometers every 10 seconds. The time period for the morning is from midnight (00:00:00 AM) to noon (12:00:00 PM) for a period of 12 hours, and for the evening, from 12:00:00 PM to midnight (12 hours). Thus, a 12-hour period contains roughly 72 readings for wind speed and direction. In a day, for the morning and evening, an average of these 72 readings is taken to represent the average wind intensity. Then a weekly average is performed, and 4 weeks per month are considered in order to estimate the representative wind intensity. For the 1st inter-monsoon, March and April were considered. For the Southwest monsoon, the months of May, June, July, August, and September were considered. The 2nd inter-monsoon runs through the months of October and November. The Northeast monsoon season was represented by taking the data from December, January, and February respectively.

Weekly wind measurement values were averaged for the four main wind seasons based on the wind values in the morning and the evening hours. Summary of the wind measurement statics are shown in Table 2. Slightly higher wind speeds with respect to the mornings could be seen for all the four seasons (Fig. 3). Maximum wind speeds could be observed during the Southwest monsoon season, while almost steady or windless conditions could be seen during the 1st and 2nd Inter monsoon seasons. Large variation in the wind speeds as well as incoming rain could be expected during the Southwest monsoon season. According to Fig. 3, the evening wind speeds are generally slightly higher than the morning wind speeds, across the island. These changes in the average wind speed measure and the variance depending on the morning and the evening makes the regionalized variability of wind modelling more acute.

The linear models fitted to the morning and evening wind speeds (Fig. 3) indicate the directional behavior of the winds as well as their possible orientation. It could be observed the linear models indicate that the direction and the magnitude of each wind regime are different, but they are quite the same for the morning and the evening wind behaviors. The average distances between the 23 regionalized variables are in the range 50 km to 100 km (Fig. 4). The average distance determined by the matrix is important in the case of defining lag distance for the modelling of spatial autocorrelation. The Euclidian distance between the geodetic locations of the measurement stations is tabulated and represented by matrix notation with upper diagonal elements showing the average distance variation within the occupied spatial range of the stations. According to the measurements, the coastal zones in the Northeast and the Northwest are closer than expected, with an average separation exceeding 200 km compared to the rest of the stations clustered in the centre of the island. When modelling the wind for islands such as Sri Lanka, the spatial correlation structure within the country is heavily determined by the coastal wind speeds and the directions.

Table 2: Summary statistics of the four wind regimes for the morning and evening, year 2015.

1st Inter-monsoon March-April						**Southwest Monsoon May- September**					
Morning 00:00:00 AM-12:00:00 PM/ Every 10minutes						Morning 00:00:00 AM-12:00:00 PM/Every 10minutes					
Min	1st Qu.	Median	Mean	3rd Qu.	Max.	Min	1st Qu.	Median	Mean	3rd Qu.	Max.
0.853	1.608	2.585	3.004	3.328	7.907	1.207	3.101	5.404	6.062	8.030	15.653
Evening 12:00:00 PM-00:00:00 AM/ Every 10minutes						Evening 12:00:00 PM-00:00:00 AM/ Every 10minutes					
Min	1st Qu.	Median	Mean	3rd Qu.	Max.	Min	1st Qu.	Median	Mean	3rd Qu.	Max.
2.196	3.566	4.859	5.461	6.575	12.675	2.952	5.328	6.019	7.282	8.119	17.872
2nd Inter-monsoon October-November						**Northeast Monsoon December-February**					
Morning 00:00:00 AM-12:00:00 PM/ Every 10minutes						Morning 00:00:00 AM-12:00:00 PM/ Every 10minutes					
Min	1st Qu.	Median	Mean	3rd Qu.	Max.	Min	1st Qu.	Median	Mean	3rd Qu.	Max.
0.654	1.375	2.449	2.822	3.1726	7.7060	0.493	1.618	3.242	3.814	4.056	11.587
Evening 12:00:00 PM-00:00:00 AM/ Every 10minutes						Evening 12:00:00 PM-00:00:00 AM/ Every 10minutes					
Min	1st Qu.	Median	Mean	3rd Qu.	Max.	Min	1st Qu.	Median	Mean	3rd Qu.	Max.
0.915	2.181	3.463	3.832	4.734	10.339	2.017	3.238	4.863	5..861	7.211	14.513

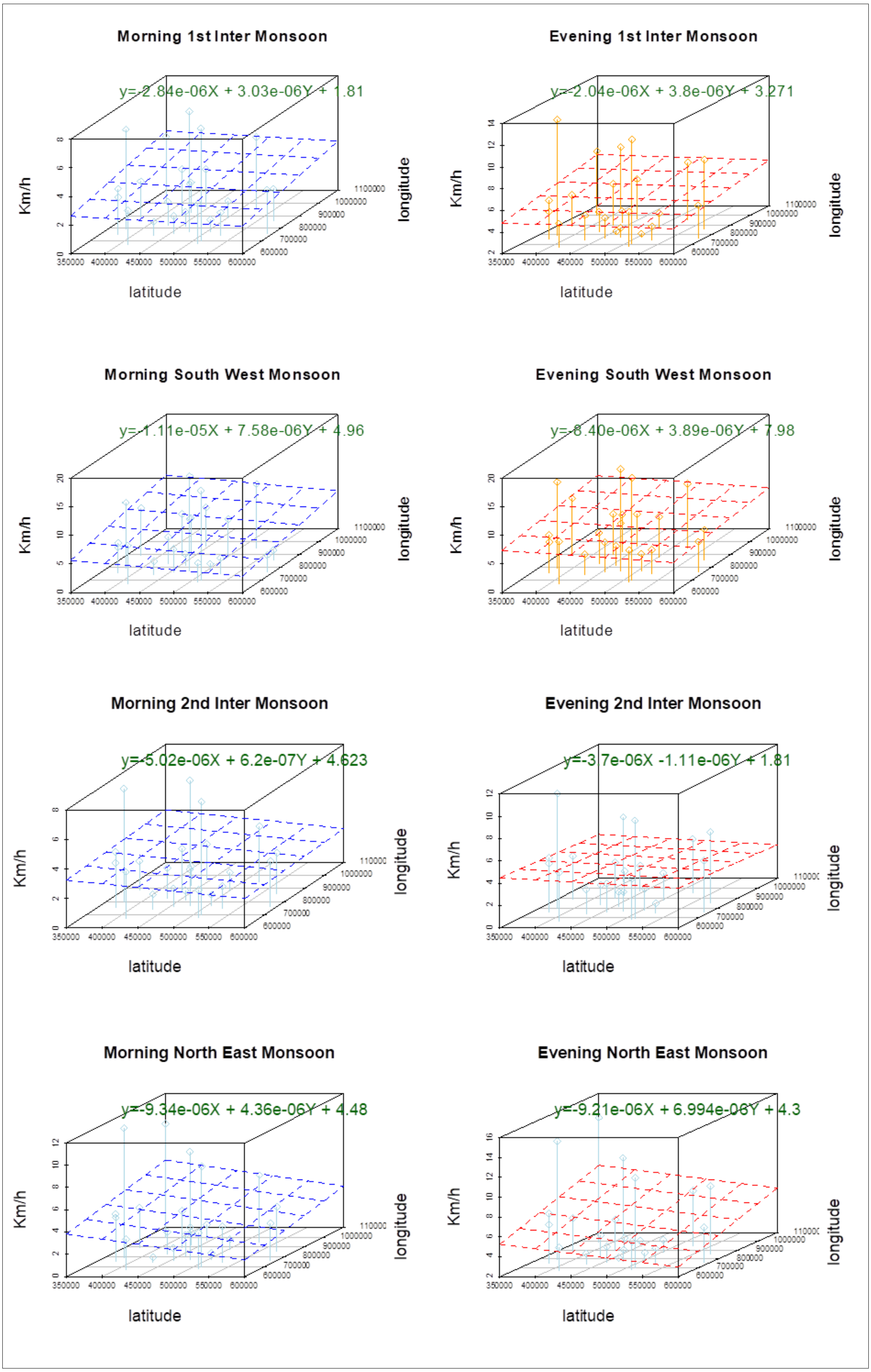

Fig. 3: Changes in the wind speeds during the morning and the evening times of the four wind seasons for the 23 wind stations, year 2015. Linear planes were fitted to represent the wind field as random variable and the orientation.

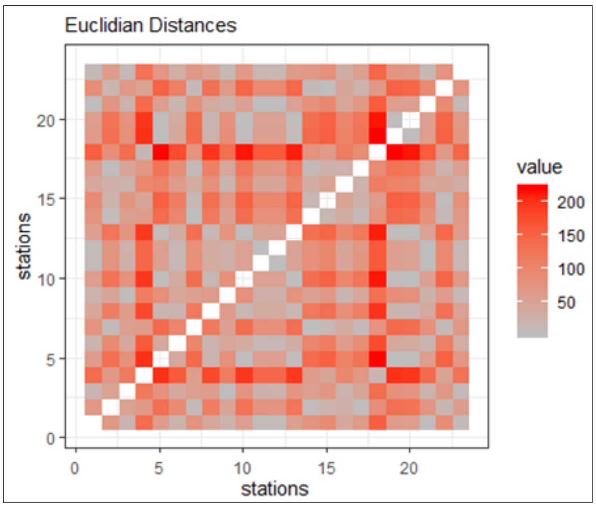

Fig. 4: Euclidian distances matrix between regionalized observation stations (refer to the upper diagonal or the lower diagonal).

Though it is not considered in the study, regional coastal wind mapping should be carried out in tandem with the central regions to better understand the overall behavior. As the topography of the island suggests, the coastal belt representing Jaffana, Katunayaka, Puttalam, and Trincolmalee receives very high wind effects, while Colombo, Galle, Hambanthota, and Baticaloa eventually reach moderately high yet steady wind effects (Fig. 5). Most of the central part of the country, especially Badulla, Anuradhapura, Kurunagala, Nuwara Eliya, and Ratnapura, receive low wind effects.

4 Materials and Methods

According to the Regionalized Variable Theory (Journel and Huijbregts, 1979; Sterk and Stein, 1997), modelling data deviating from normal distribution and with non-stationary mean and variance is difficult. Spatial structure which describes the spatial dependence with the distance and direction for the observations is an important parameter in the process of estimation. Semivariance is one of the main parameters determining the auto-covariance or spatial correlation. Semivariance measures the dissimilarity between observations, while the covariance and the correlation both measure the similarity. In the case of dynamic variables such as wind speed and direction, dissimilarity provides more meaningful quantification than similarity. This is because the use of similarity will not attribute the continuous changes that the random variable shows within a small-time interval. As a whole, semivariance and covariance have an inverse relationship. The semivariance function (Eq. 1) shown below quantifies the spatial correlation between measurement points based on distance and direction. Further, it can also account for the direction dependent variability or anisotropic spatial patterns (Luo et al., 2008).

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \left[Z(x_i - x_j)\right]^2 \qquad \text{Eq. (1)}$$

Here $m(h)$ is the number of point pairs separated by a vector. The direction dependence and variability of the regionalized variables are estimated using a semivariance function. These patterns in autocorrelation structure with respect to the direction could be reflected by using an anisotropic variogram. From the empirical variogram computed by Eq. 1, we derive the acceptable variogram model, which expresses the semivariance as a function of the separation vector. The model allows us to infer the characteristics of the underlying process from its functional form and its parameters. This computed semivariance between any two-point pairs is used in optimal interpolation kriging to make predictions at unsampled locations. Due to theoretical and mathematical limitations, several functional forms can be used to model the variogram. These models are also called the authorized models.

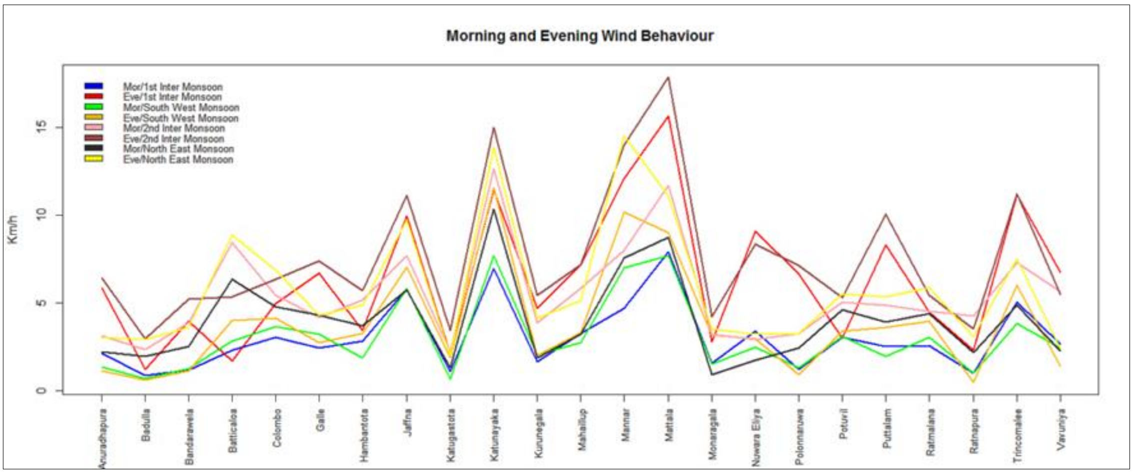

Fig. 5: Time series seasonal Wind speeds four each of the 23 measuring stations, the year 2015. Data represented by the daily and weekly averages of 10 minutes measuring intervals.

In this study, three of the authorized models, as shown below, were used to determine the weights for the nearby supporting data to compute the interpolated values (Ord, 1975; Ripley, 1988; Stein, 1999).

The exponential model:

$$\gamma(h) = c\left\{1 - e^{-\frac{h}{a}}\right\} \qquad \text{Eq. (2)}$$

Where h is the lag distance, sill is c and effective range is $3a$.

The Spherical model (Fig. 6)

$$\gamma(h) = \begin{cases} c\left\{\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} & h < a \\ c & h \geq a \end{cases} \qquad \text{Eq. (3)}$$

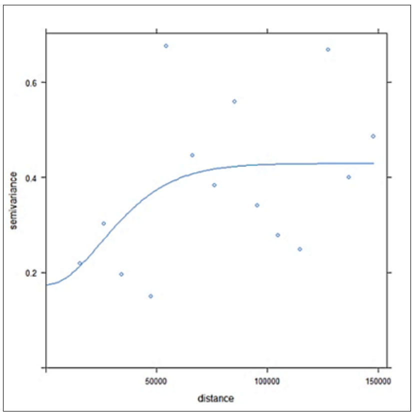

Fig. 6: A variogram model with a Spherical model fitting.

And the Gaussian model:

$$\gamma(h) = c\left\{1 - e^{-3\frac{h^2}{a^2}}\right\} \qquad \text{Eq. (4)}$$

In the equations 2, 3, and 4, h represent the lag distance as mentioned earlier, a represents the range of the practical range and c the sill, where the correlation structure could be expected. Each of these parameters needs to be estimated separately for the authorized models using the empirical variogram cloud. Within various variogram models, the spherical model is the most widely used and often preferred when the nugget variance (nonspatial variations) is essential and there is a clear range and sill effect (Cressie, 1993; Luo et al., 2008). In this study, a combination of exponential, spherical, and Gaussian variogram models were incorporated to determine the weights for the nearby supporting data to compute the interpolated values. It has been considered in this study that the sill, the limiting values of the variogram model, and the range that provides the distance at which the spatial dependencies exist are important parameters for the actual estimation. The anisotropic variograms were first analyzed for the most reasonable authorized model before the kriging interpolation execution.

A stochastic method called kriging (Goovaerts, 1997; Krige, 1951) is comparable to IDW in that it estimates a value at an unknown point using a linear combination of weights at known sites. Kriging addresses the issue of surface estimation differently than deterministic approaches by accounting for the spatial correlation, which is measured by the semivariance function. As a result, kriging now estimates a stochastic variable rather than a parameter (expectation of an observation). There are several varieties of kriging that have been created, including ordinary, universal, simple, and indicator kriging (Sterk and Stein, 1997). From the possible kriging methods, we have employed ordinary and universal kriging methods. Further ordinary kriging (OK) experimental results are presented. Our aim is to use the local neighboring measuring stations as much as possible to provide estimations on unsampled locations. Thus, it is reasonable to leave the mean value being estimated, as it changes across the spatial grid. Ordinary kriging provides this possibility on top of the other methods. Ordinary kriging, which assumes the observation mean is unknown, focuses on the spatial component and uses only the samples in the local neighbourhood for the estimation. Cressie (1993), Isaaks and Srivastava (1989), and Webster and Oliver (2007) provide more information on kriging. In OK, we model the value of a variable at a location as the sum of two components, where m is the regional mean and the spatially correlated random component. The regional mean m is estimated from the sample but not as the simple average, because there is spatial dependence. It is implicit in the OK system. This mean value is constant across the field; that is, the expected value is the same but unknown; this is the "ordinary" situation. The spatially-correlated random component is estimated from the spatial covariance structure determined by the variogram model. The estimated value at a point is predicted as the weighted average of the values at all sample points:

$$\hat{Z}\,(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \qquad \text{Eq. (5)}$$

Where $\sum_{i=1}^{N} \lambda_i = 1$ so that the estimation is unbiased, because $E[\hat{Z}\,(x_0) - Z(x_0)] = 0$.

5 Experimental results

The variogram parameters for the fitted anisotropic variograms for the main four seasons were calculated and modelled. These models for the morning and evening hours of the Southwest monsoon and the Northeast monsoon are shown below in figures 7, 8, 9, and 10 respectively. The exponential variogram model has been the model of choice in fitting the empirical variograms produced for each wind season. Estimating these functions in the case of irregularly distributed data could be extremely challenging, mainly due to the need to pool the data pairs into larger bins (Bargaoui and Chebbi, 2009).

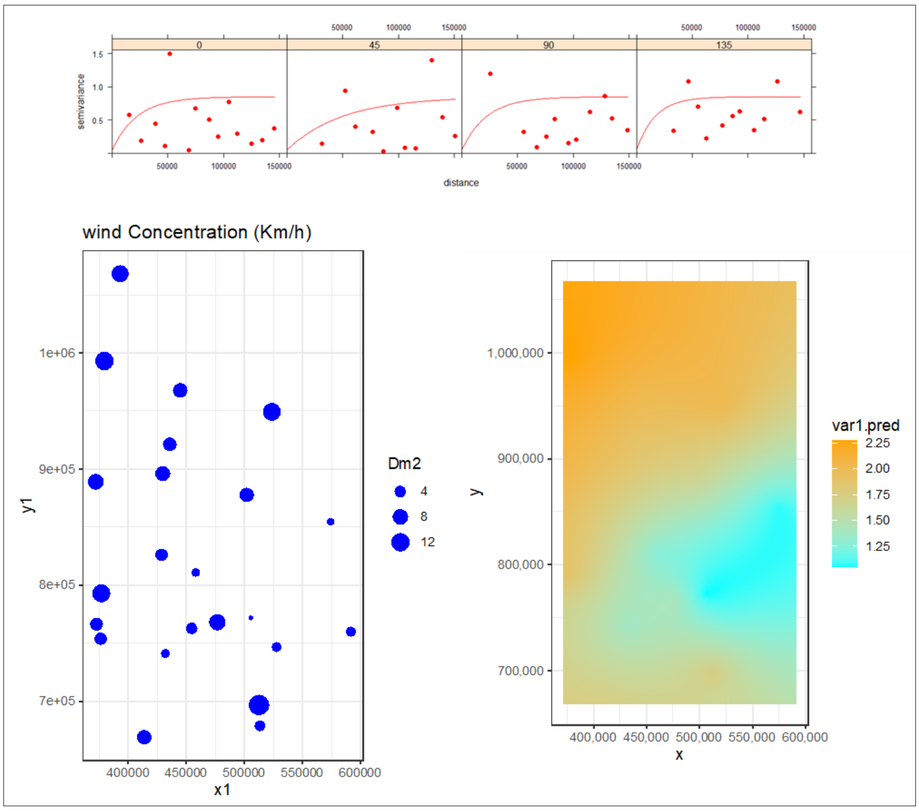

Fig. 7: Southwest monsoon, morning hours, Anisotropic Variogram models and the initial wind predictions, with a spherical fit, N 90° E, Partial Sill-0.75, Range-50 km, Nugget-0.05.

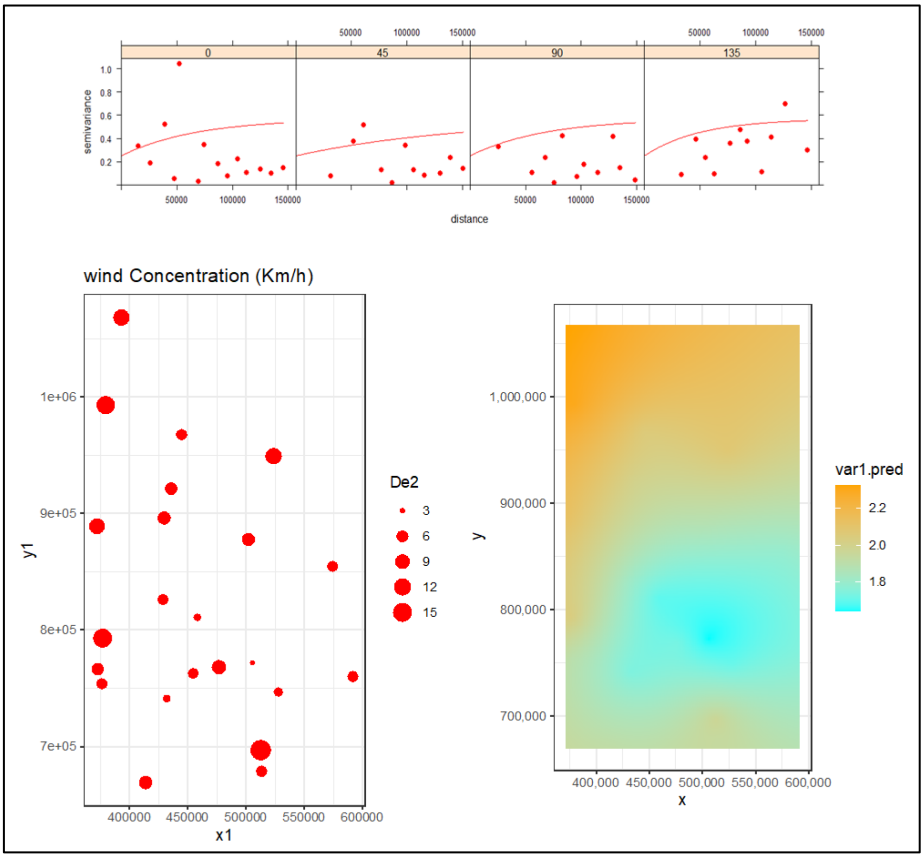

Fig. 8: Southwest monsoon, evening hours, Anisotropic Variogram models and the initial wind predictions, with an exponential fit, N 90° E, Partial Sill-0.7, Range-15 km, Nugget-0.25.

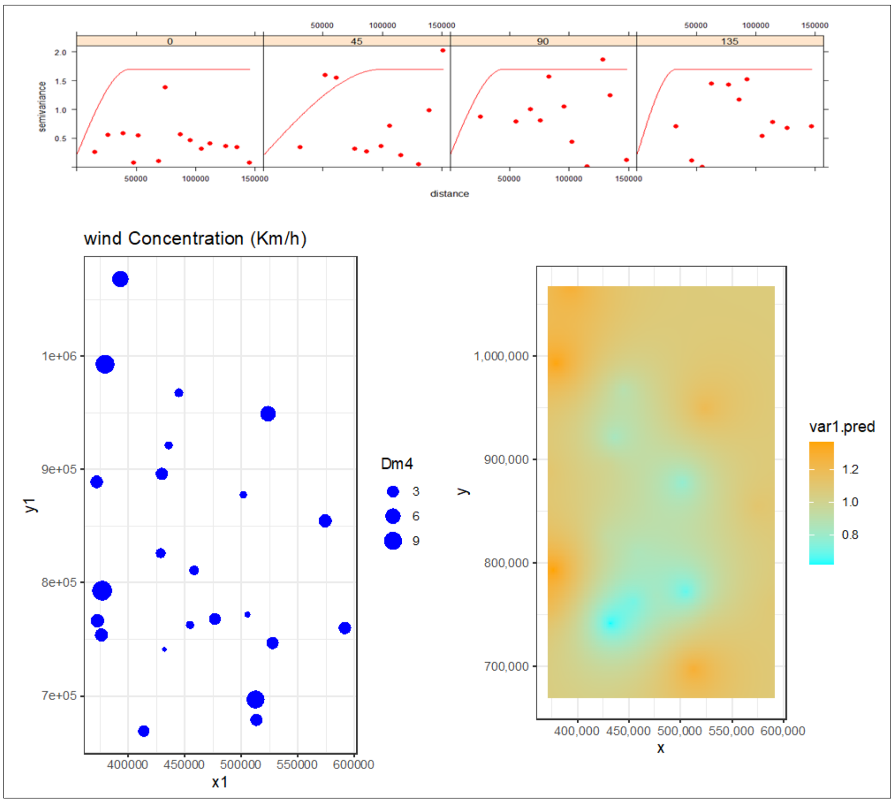

Fig. 9: Northeast monsoon, morning hours, Anisotropic Variograms and the initial wind predictions, with an exponential fit, N 90° E, Partial Sill-1.5, Range-10 km, Nugget-0.2.

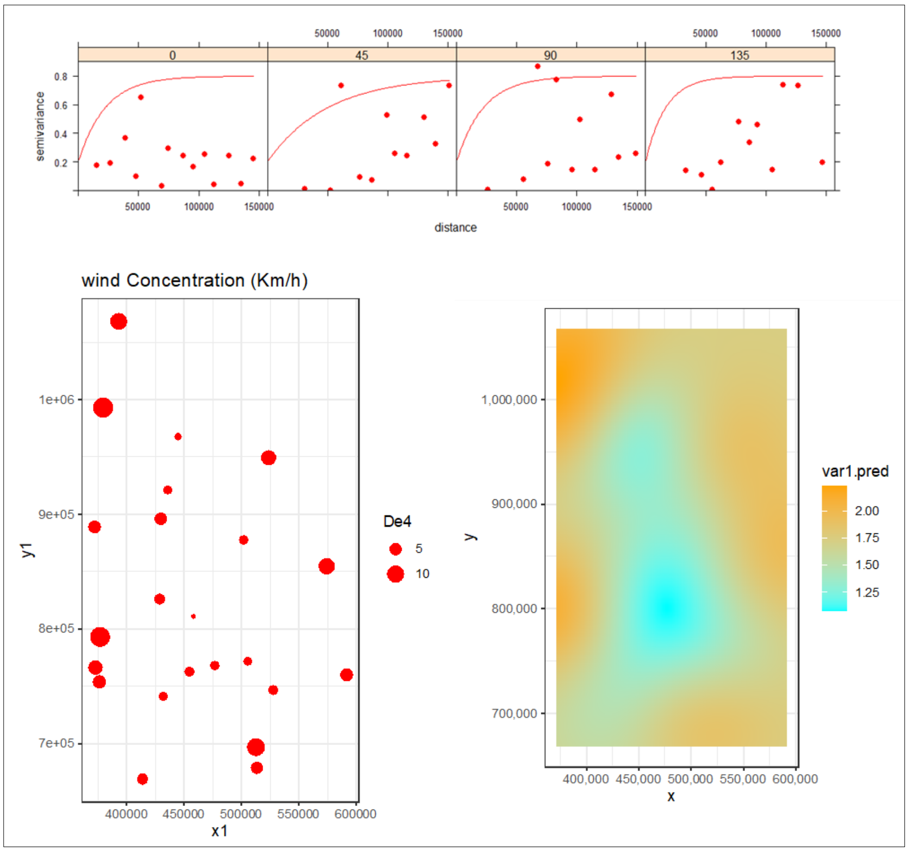

Fig. 10: Northeast monsoon, evening hours, Anisotropic Variograms and the initial wind predictions, with an exponential fit, N 135° E, Partial Sill-0.6, Range-50 km, Nugget-0.2.

Larger lag spacing and tolerance increase the number of data pairs for estimation but decrease the amount of detail in the semivariogram, as shown by the variogram clouds.

The longitudinal span of the measuring stations is approximately 219 km, while the latitude span is approximately 398 km. The bin size for the variogram was estimated using the bounding box of the data points and it was set to 10 km. In this study, the spatial continuity of the wind speed measurements was determined by using fitted variograms in four different directions: N, N 45° E, N 90° E, and N 135° E. An experimental semivariogram of average wind speed during the morning hours of the Southwest monsoon season has been fitted by a spherical model in four different azimuthal directions. The semivariogram in the direction of N 90° E was used for the weight determination for the interpolation. The semivariogram levels off to the sill almost at a distance of 5 km. The partial sill is about 0.75. In the second case, an experimental semivariogram of average wind speed during the evening hours of the Southwest monsoon season has been fitted by an exponential model in four different azimuthal directions. As for the morning, the semivariogram in the direction of N 90° E was used in the weight determination for the interpolation. The semivariogram levels off to the sill almost at a distance of 4 km. The partial sill is about 0.07. An exponential model in four different azimuthal directions has been fitted to the third experimental semivariogram of average wind speed during the morning hours of the Northeast monsoon season. As for the morning, the semivariogram in the direction of N 90° E was used in the weight determination for the interpolation. The semivariogram levels off to the sill almost at a distance of 4 km. The partial sill is about 1.5. Finally, for the 4th monsoon, an experimental semivariogram of average wind speed during the evening hours of the Northeast monsoon season has been fitted by an exponential model in four different azimuthal directions. As for the morning, the semivariogram in the direction of N 135° E was used for the weight determination for the interpolation. The semivariogram levels off to the sill at a distance of almost 5 km. The partial sill is about 0.6. The resulting wind speed maps for the 4 seasons are shown in Fig. 11. The surfaces interpolated by OK provide a more accurate representation of the variation of wind speeds during the monsoon seasons in Sri Lanka. The maximum and minimum wind speeds are broadly represented according to the original data range. Overall model parameters experimented on in the study are summarized in Table 3. It could be seen that the variogram models shift between the exponential and spherical, while a Gaussian model, resulted in the 2nd inter monsoon mesurments. Spherical and exponential models exhibit linear behavior at the origin and are quite appropriate to represent properties with a higher level of short-range variability. The anisotropy resulted in the study is due to the different autocorrelation structures of the wind in different directions. This step in the study is the most significant and the most critical stage of the whole modelling process. An additional validation by using a larger number of ground measurements focusing purely on validation should be further executed.

Table 3: Authorized variogram models over empirical variogram and their parameters.

Wind Season	Time of Day	Estimated Semivariogram Model	Anisotropy	Variogram Parameters
1st Inter-Monsoon (March-April)	Morning	Spherical	N 0° E	Partial Sill-0.5 Range-10Km Nugget-0.1
	Evening	Exponential	N 135° E	Partial Sill-0.6 Range-10Km Nugget-0.1
Southwest Monsoon (May-September)	Morning	Spherical	N 90° E	Partial Sill-0.75 Range-50Km Nugget-0.05
	Evening	Exponential	N 90° E	Partial Sill-0.7 Range-15Km Nugget-0.25
2nd Inter-Monsoon (October-November)	Morning	Exponential	N 90° E	Partial Sill-1.1 Range-50Km Nugget-0.1
	Evening	Gaussian	N 0° E	Partial Sill-0.2 Range-5Km Nugget-0.2

Northeast Monsoon (December-February)	Morning	Exponential	N 90˚ E	Partial Sill-1.5 Range-10Km Nugget-0.2
	Evening	Exponential	N 135˚ E	Partial Sill-0.6 Range-50Km Nugget-0.2

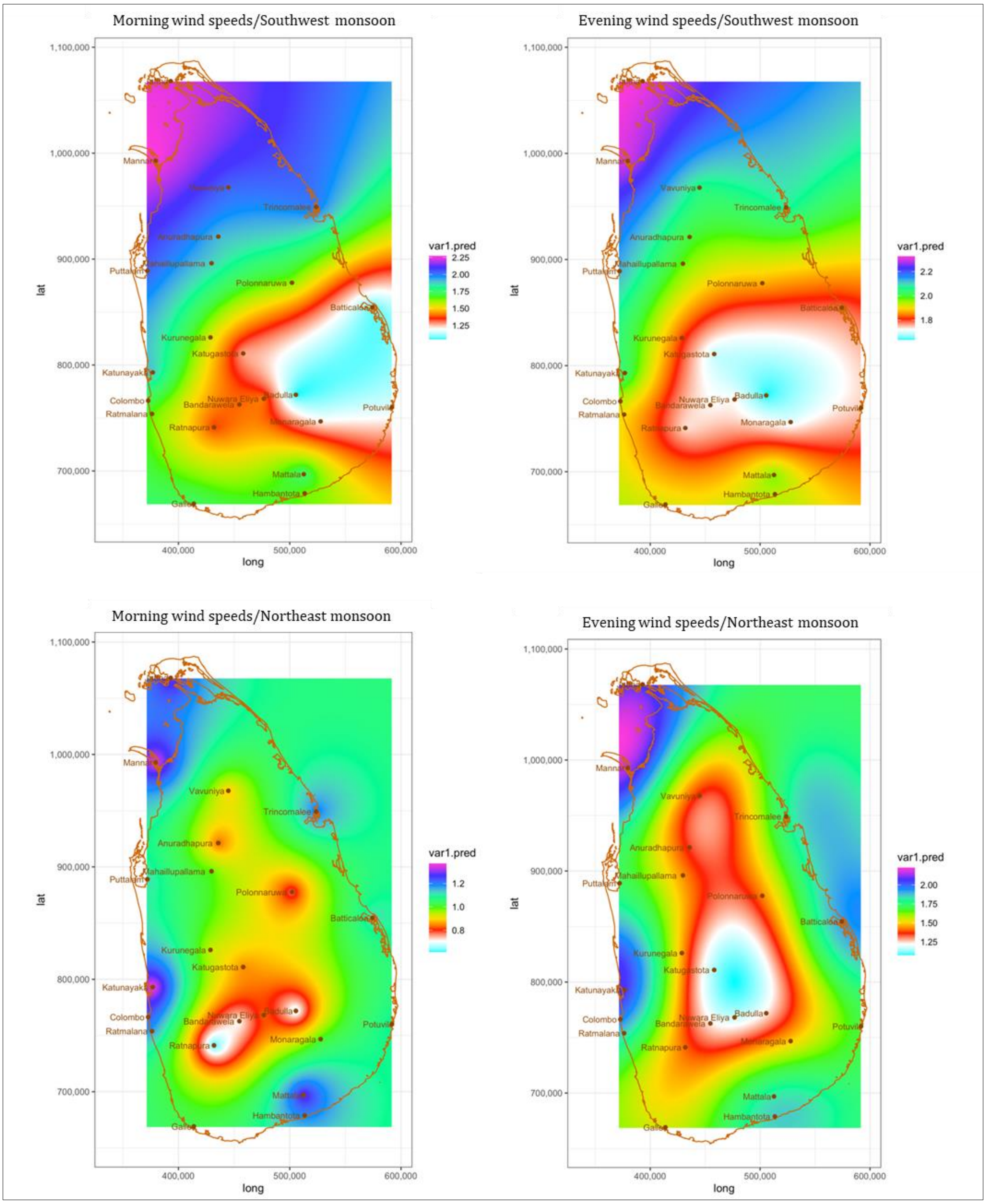

Fig. 11: Ordinary Kriging based interpolated wind speed surfaces, left to right, and top to bottom; morning of the Southwest monsoon, evening of the Southwest monsoon, morning of the Northeast monsoon, and evening of the Northeast monsoon seasons respectively.

The limitations of previous work using variograms to model the wind in the study area focused on this work does not provide possibilities for comparison. Cross validation was implemented to evaluate each interpolation result. The Sum of Squares of Error (SSE) was calculated for the validation according to the following equation.

$$SSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Z}\,(x_i) - Z(x_i)\right)^2 \quad \text{Eq. (6)}$$

where $\hat{Z}\,(x_i)$ is the predicted value and $Z(x_i)$ is the measured value.

The OK prediction standard deviations are well represented in Fig. 11. In general, kriging standard deviations are higher for the Northeast parts of Sri Lanka as can be seen with the prediction results of the Southwest monsoon wind speed. This variance is subjected to change further due to the scarceness of the measuring stations in the region. Low standard deviations of the kriging predictions can be seen in the central regions of Sri Lanka during the Northeast monsoon seasons, while large daily mean wind speeds occur in the regions near to shore areas all around the country.

6 Conclusions and Recommendation

The main aim of the study was to use the spatial autocorrelation structure of irregularly measured wind speed data on a daily basis for a period of one year to model the wind patterns of Sri Lanka. The use of R open source software for such detailed modelling has also been highlighted. The results of this study show that the use of variograms and OK interpolation produces simple yet robust wind speed predictions. The overall statistical modelling was done using the R 4.0.2 open source programming environment. The possibility to handle spatial data with different projection coordinate systems within R has made the kriging-based interpolation more robust (Table 2). The irregularly distributed wind measuring stations introduce difficulties in the variogram modelling. One instance is that the variogram clouds did not become dense enough for estimations. As a solution, we suggest looking into antistrophic empirical variograms and choosing the most reasonable one out of them. The paper reports the predictions for the two main wind seasons in Sri Lanka (with the standard errors of predictions shown in Table 4). The SSE for all the predictions was generally very small, leading to reasonably reliable results. Yet it is necessary to carry out field data for comparisons in order to validate the interpolation results to a higher degree of certainty. A comparison with other interpolation methods, such as universal and co-kriging, could provide insight into a complete prediction mechanism for vague geographic phenomena such as wind, centered for this study. Further timeseries variability modelling, which is possible within R, could be implemented to reduce the uncertainties in the predictions due to a smaller number of measurements. Comprehensive validation of the prediction results is necessary in order to provide strong justification of the methods adopted. A seasonal Auto-Regressive Integrated Moving Average (SARIMA) model could be used in tandem to see the robustness of the kriging outcomes. The study will extend in the direction of producing mobile wind measuring data in the regions where the measurements have not been made in order to see the performance of the interpolations. Possibilities of using external drifts, such as altitude, will also be one further step of this study.

Table 4: The standard error estimations for respective model predictions

Wind Season	Weighted Sum of Square Error (SSE) for the fitted model
Morning of the Southwest monsoon	2.062047e-09
Evening of the Southwest monsoon	2.355453e-10
Morning of the Northeast monsoon	8.422471e-10
Evening of the Northeast monsoon	1.063742e-09

R programming environment is a comprehensive tool to apply higher order statistical methods for spatial data. Statistical pattern recognition for vague and dynamic geographic phenomena could become much more efficient with the use of R. This paper was much oriented in the direction of stressing the potential use of open source programming such as R for quick, yet robust, spatial data mapping on the basis of the detailed data availability.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Conceptualization, methodology, analysis, writing - original draft, review and editing, D.R.W. The author has read and agreed to the published version of the manuscript.

Conflict of Interest

The author declares no conflict of interest.

References

A comprehensive R Archive network, https://cran.r-project.org

Bargaoui, Z. K., Chebbi, A., 2009. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. Journal of Hydrology. 365:(1-2), 56-73.

Crawley, M. J., 2005. Statistics: an introduction using RJ Wiley. Chichester, West Sussex, England.

Crawley, M. J., 2007. The R book. John Wiley & Sons.
Cressie, N. A., 1993. Statistics for spatial data. John Willy and Sons. Inc., New York.
Dalgaard, P., 2002. Analysis of variance and the Kruskal-Wallis test. Introductory Statistics with R. 111-127.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press.
Isaaks, E., Srivastava, R., 1989. An introduction to applied geostatistics. Oxford University Press, Oxford.
Journel, A. G., Huijbregts, C. J., 1979. Mining Geostatistics. Academic Press, London & New York
Krige, D. G., 1951. A statistical approach to some basic mine problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy. 52:119-139.
Luo, W., Taylor, M., Parker, S., 2008. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. International Journal of Climatology: A Journal of the Royal Meteorological Society. 28:(7), 947-959.
Met-SL, 2016. Meteorological Department of Sri Lanka, http://www.meteo.gov.lk/index.php?lang=en
Ord, K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association. 70:(349), 120-126.
Pebesma, E., Bivand, R. S., 2005. S classes and methods for spatial data: the sp package. R news. 5:(2), 9-13.
Ripley, B. D., 1988. Statistical inference for spatial processes. Cambridge university press.
Ripley, B. D., 2001. Spatial statistics in R. R news. 1:(2), 14-15.
Stein, M. L., 1999. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.
Sterk, G., Stein, A., 1997. Mapping wind - blown mass transport by modeling variability in space and time. Soil Science Society of America Journal. 61:(1), 232-239.
Webster, R., Oliver, M. A., 2007. Geostatistics for environmental scientists. John Wiley & Sons.