STATISTICAL AND GEOSTATISTICAL ANALYSIS OF WIND: A CASE STUDY OF DIRECTION STATISTICS

Tetsuya Shoji[a]

[a] Department of Environment Systems, The University of Tokyo, Tokyo 113-0033, Japan
shoji@k.u-tokyo.ac.jp

Commission IV, WG IV/1

KEY WORDS: Wind, Direction, Vector, AMeDAS, Power Generation, Exponential Distribution, Variogram, Central Japan

ABSTRACT:

To study the applicability of geostatistics for vector data, wind velocity data have been analyzed statistically and geostatistically. The study area consists of two districts, the mountainous Chubu and plain Kanto districts in central Japan. For the distribution of wind speeds, exponential distribution was fitted well in both districts. Temporal experimental variograms of wind speeds, directions and velocities suggest daily duration, and wind is stronger by day than by night. While some spatial experimental variogram of wind speeds are traditional spherical schemes showing clear nugget effects, sills and ranges whose values vary 50–130 km, other variograms are not traditional schemes illustrating flat or gradually increasing curves. If variograms of momentary wind speeds have no range an empirical variograms of temporally averaged wind speeds does not also show a range.

1. INTRODUCTION

In order to find a location suitable for wind power generation, it is necessary to know spatial distribution of wind. Geostatistics is a powerful tool to estimate a spatial distribution of geosciences variables. Accordingly, geostatistical tools are applied to the estimation of wind distribution in space. However, we must overcome two difficulties for the applications. The first one is that wind data are measured at a moment in time at a spatial location. Generally, averaging in time is applied to reduce data variation. We have to check the effect of such averaging along a time axis on analytical results of originally momentary wind data. The other problem is that wind velocities are vector. Geostatistical tools have been developed for commonly scalar data. A vector datum always consists of two values at least and very different mathematical characters than the scalar data.

Vector data such as not only wind velocity, but also topological slope, folded geological structure, underground water flow, and others are very common in geoscience fields. In this paper, several new geostatistical tools are proposed for the treatment of the vector data. The analytical results from the new tools suggest several viewpoint of assessing the locations of the wind power generation stations.

2. DATA AND AREAS

The Meteorological Agency of Japan has established an automated meteorological observation system named AMeDAS (Automated Meteorological Data Acquisition System). AMeDAS has 1536 stations in the whole area of Japan (377,800 km^2). Each station records rain precipitation, wind velocity, temperature and other meteorological data at every hour. The station density is about 1/250 km^{-2} (41 stations in 100・100 km^2). If the stations are arranged on a tetragonal grid, the average distance between the neighboring stations is 16 km. If they are arranged on a trigonal grid, the average distance is 17 km. Meteorological Agency of Japan have published data of AMeDAS every year as a CD-ROM. The present study uses the data in 1999.

A wind velocity ($\vec{v}$) is a vector in a 2-D space, and is described by the combination of a speed and a direction as follows:

$$\vec{v} = s \cdot \vec{u} \qquad (1)$$

where $\vec{v}$ = velocity (vector)
$s = |\vec{v}|$ = speed (scalar)
$\vec{u}$ = direction (unit vector).

Since the direction (θ) is defined as an angle measured clockwise from the north, the unit vector is given as follows:

$$u_{NS} = \cos\theta$$
$$u_{EW} = \sin\theta \qquad (2)$$

where u_{NS}, u_{EW} = north-south and east-west components of the direction.

In order to compare with wind patterns in mountainous area and plain, the Chubu district characterized by many mountains, and the Kanto district characterized by the widest plain in Japan are selected for the analysis (Figure 1). Both districts are next of each other, and the former is situated west of the latter. Accordingly, wind blows sequentially or contemporarily in both districts.

The Chubu district consists of many folding mountains and volcanoes including Mt. Fuji, which is the highest in Japan. In order to obtain the characteristics of wind in the mountainous district, only the stations whose elevations are higher than 100 m above the sea level are selected in the district. 170 stations are included in this district, and about 60 % of them recorded wind data. The elevation of the highest worked station is 1350 m (the elevation of the highest station is 2730 m). The density of stations is about 1/510 km^{-2}. This means that the nearest station distance is 23 km in a tetragonal grid, or 24 km in a trigonal grid.

The Kanto district consists of many fields and cities including Tokyo, which has the largest population in Japan. The stations lower than elevation 200 m are selected in this district. 81 stations are included in this district, but about 50 % of them recorded wind data. The density of stations is about 1/490 km^{-2}. This means that the nearest station distance is 22 km or 24 km.

3. STATISTICS

3.1 Statistical Distribution of Speeds

First, basic statistics of speed, direction and velocity were obtained from wind data in the Chubu and Kanto districts, and the Chubu-Kanto total area defined by the combination of both districts. We have empirically fitted three distribution functions: normal, lognormal, and exponential distributions. Table 1 summarizes sample correlation coefficients between wind speeds and empirical cumulative frequency distribution functions. All sample correlation coefficients are high. Especially, exponential functions (Figure 2) show high correlation coefficients: –0.999 for the Chubu district, –0.998 for the Kanto district, and –0.9996 for the Chubu-Kanto total area. Figure 2 means that wind is stronger in the plain Kanto district than the mountainous Chubu district. This feature has been consistent, even if the data are treated in each month.

The fact that the correlation coefficient between wind speeds and the cumulative frequency is extremely high ($r^2 > 0.99$) in an exponential distribution is very important for assessing wind power generation, because this means that we can estimate accurately wind speeds in a simulation based on an assumption of a random process (e.g. the Monte Carlo method).

3.2 Rose Diagram of Directions (Wind Rose)

Frequencies of wind directions are presented by a rose diagram named “wind rose”. Figure 3 is an example. This diagram is most deviated among treated data. Let us define the average direction as follows:

$$\langle \vec{u} \rangle = \sum \vec{u}_i / n \qquad (3)$$

where $\langle \vec{u} \rangle$ = average direction

$\vec{u}_i$ = i ’th datum

n = number of data.

If direction is presented by Equation (2), then Equation (3) gives the following equations:

Table 1. Correlation coefficients between wind speed and frequency cumulated from the high value side in some statistical distribution models.

District	Number of	Statistical Distribution Model		
		Normal	Lognormal	Exponential
Chubu	504027	−0.9668	−0.9923	−0.9992
Kanto	779097	−0.9696	−0.9909	−0.9983
Total	1230815	−0.9657	−0.9929	−0.9996

“Total” means the total area of Chubu and Kanto.

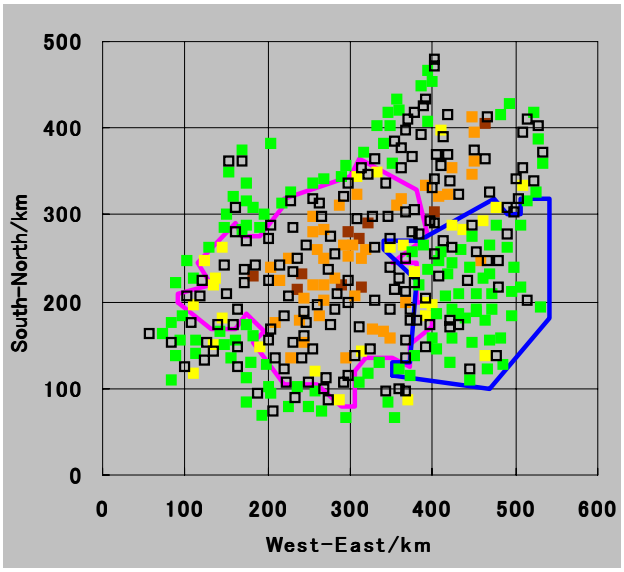

Figure 1. A map showing the locations and the elevations of the station points used in this study (open squares indicate did not work for wind data). The areas bounded by red lines and green lines represent the Chubu district and the Kanto district, respectively. Elevations: brown ≥ 1000 m, orange ≥ 500 m, yellow ≥200 m, yellowish green ≥ 100 m, and green < 100 m.

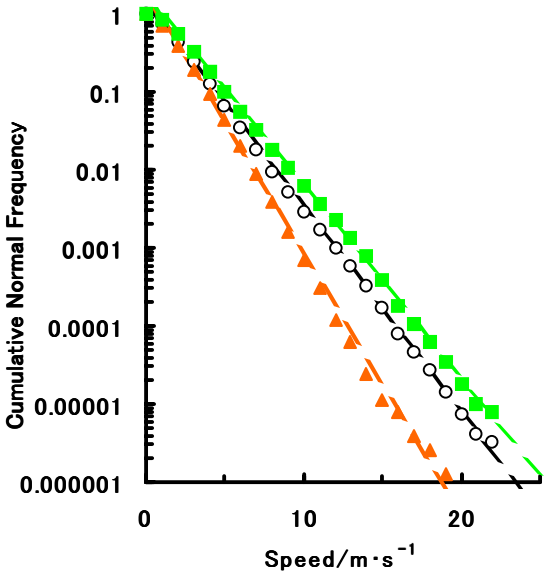

Figure 2. A diagram of fitting an exponential distribution for wind speeds in Chubu (orange triangle) and Kanto (green square) districts, and the Chubu-Kanto total area (black open circle).

$$\langle u_{NS} \rangle = \sum u_{NSi} / n$$
$$\langle u_{EW} \rangle = \sum u_{EWi} / n \quad (4)$$

where $\langle u_{NS} \rangle$, $\langle u_{EW} \rangle$ = north-south and east-west components of the average direction

u_{NSi}, u_{EWi} = north-south and east-west components of i'th datum.

A large $|\langle \vec{u} \rangle|$ value means that wind directions are largely deviated. The largest value is observed in January in the Kanto district. The value and the azimuth in this month and district are 0.39 and 323°, respectively (Figure 3).

4. VARIOGRAM

4.1 Variogram Equations

Generally, empirical variogram for a scalar variable is defined as follows:

$$\gamma(h) = \sum \{s(x_i) - s(x_i + h)\}^2 / n(h) \quad (5)$$

where $\gamma(h)$ = variogram

$s(x_i)$ = scalar value at point whose coordinate is x_i

$n(h)$ = number of point pairs whose distances are h.

Speed (s) is scalar, and hence this definition is applicable. Direction ($\vec{u}$) and velocity ($\vec{v}$) are vector, and hence Equation (5) has to be expanded as follows:

$$\gamma(h) = \sum \{\vec{u}(x_i) - \vec{u}(x_i + h)\}^2 / n(h) \quad (6)$$

where $\vec{u}(x_i)$ = direction or velocity at point whose coordinate is x_i.

If direction is given as Equation (2), Equation (6) is written as follows:

$$\gamma(h) = \frac{\sum \left[\Delta u_{NS}(x_i, h)^2 + \Delta u_{EW}(x_i, h)^2\right]}{n(h)} \quad (7)$$

where $\Delta u_{NS}(x_i, h) = u_{NS}(x_i) - u_{NS}(x_i + h)$

$\Delta u_{EW}(x_i, h) = u_{EW}(x_i) - u_{EW}(x_i + h)$,

and

$$\gamma(h) = \frac{\sum \left[\Delta v_{NS}(x_i, h)^2 + \Delta v_{EW}(x_i, h)^2\right]}{n(h)} \quad (8)$$

where

$\Delta v_{NS}(x_i, h) = s(x_i) \cdot u_{NS}(x_i) - s(x_i + h) \cdot u_{NS}(x_i + h)$

$\Delta v_{EW}(x_i, h) = s(x_i) \cdot u_{EW}(x_i) - s(x_i + h) \cdot u_{EW}(x_i + h)$

$s(x_i) = |\vec{v}(x_i)|$ = speed at point whose coordinate is x_i

$s(x_i + h) = |\vec{v}(x_i + h)|$ = speed at point whose coordinate is $x_i + h$.

In this paper, empirical semivariograms were calculated in all cases, and are shown in figures. However, "variogram" is used instead of "empirical semivariogram", because their values are not important.

4.2 Temporal Variograms

AMeDAS records a wind velocity (speed and direction) at a moment at every hour. Wind velocities vary every moment. For this reason, in order to know temporal continuity of wind velocities, experimental variograms as a function of time were calculated using all annual data. Figures 4, 5 and 6 show

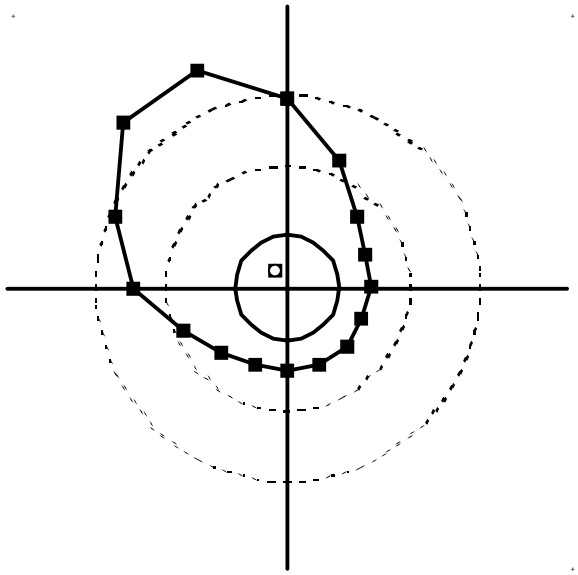

Figure 3. A wind rose showing frequencies of wind directions in the Kanto district (January, 1999). The polygon represents frequencies of directions, and the area of the central circle shows the proportion of calm (wind speed is 0). The open square shows the gravity center, when frequencies of directions are represented by columns standing at the rim of the calm's circle.

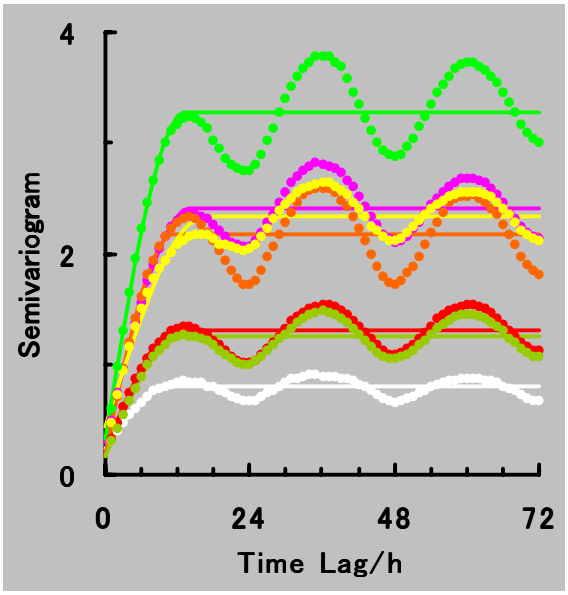

Figure 4. Temporal experimental semivariograms of wind speeds and fitted models at selected stations in the Kanto district. Hole effects showing a daily period are remarkable. Locations: white = Yorii, red = Kumagaya, brown = Kuki, orange = Hatoyama, yellow = Urawa, light green = Koshigaya, and green = Tokorozawa.

temporal variograms of speeds, directions and velocities in Saitama Prefecture, the Kanto district.

The most remarkable point observed in the temporal experimental variograms is a hole effect representing a daily period. This period is not caused by sea, because every station is located more than 20 km from seashore. Figure 7 shows wind speeds from 1st to 10th, January. According to this diagram, wind was strong by day, and weak by night. The same feature is also observed in temporal variograms of wind speeds, directions and velocities in Yamanashi Prefecture, the Chubu district.

Every temporal variogram is well approximated by a traditional spherical model. The ranges of the models vary from 8 to 19 hours (Figures 4-6). This means that a wind condition continues generally a few hours.

4.3 Spatial Variograms

Figures 8, 9 and 10 show spatial variograms in the Kanto district at 1 to 10 o'clock on 1st, January. We can see three patterns of variograms. The first one is traditional spherical variograms showing clear ranges and sills. The second one is flat (i.e. shows only sills). The third type is linearly increasing with increasing lag. It is not clear at present why these three types appear in this study.

Rain precipitation data are always accumulated values. Different to this, every wind datum is a record of a momentary condition. This seems to imply that an accumulated or averaged value has no meaning. However, the variograms in Figures 4-6 indicate some continuity existing in the temporal distribution of wind, and therefore suggests a possibility that moving averages along time has meaning. Figure 11 shows variograms of temporarily averaged wind speeds. The data were obtained from 14 to 23 o'clock on 7th, January. No temporal variogram of momentary data has a clear range and sill in this duration. The variograms shown in Figures 11 were calculated for speed data averaged during 2 to 9 hours. Any of the diagrams does not show a clear range. This point is quite different to the rain precipitation case reported by Shoji and Kitaura (2001), where variograms of accumulated data show clear ranges and sills.

The facts that some variograms of momentary wind data are not typical, and that they are not improved (i.e. do not become typical) by accumulation means that kriging will not be able to give a good estimate for wind data.

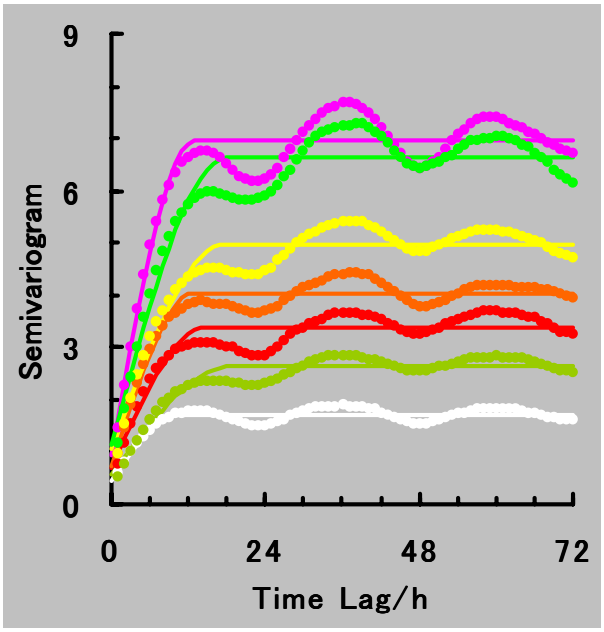

Figure 6. Temporal experimental semivariograms of wind velocities and fitted models at selected stations in the Kanto district. Symbols of locations are the same as Figure 4.

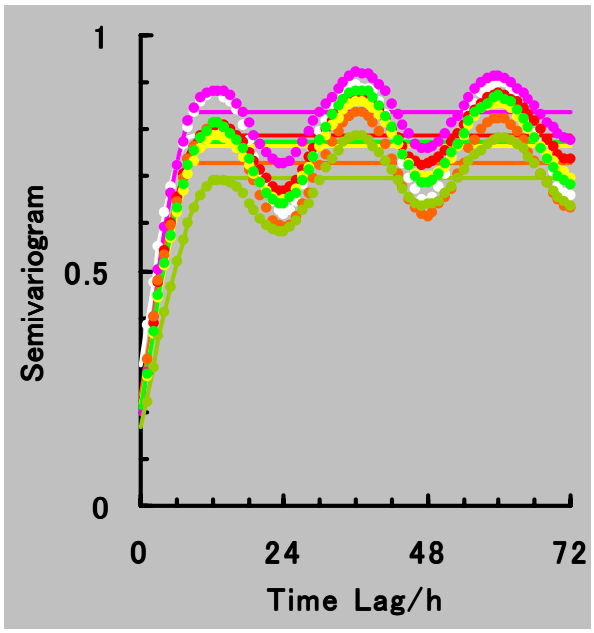

Figure 5. Temporal experimental semivariograms of wind directions and fitted models at selected stations in the Kanto district. Symbols of locations are the same as Figure 4.

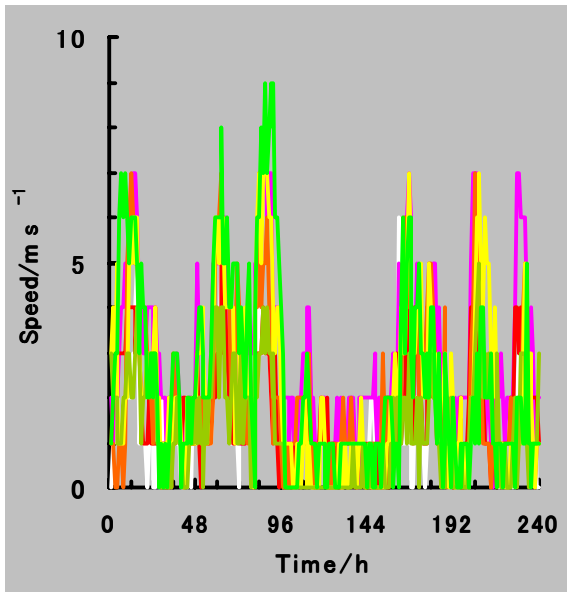

Figure 7. Wind speeds from 1st to 10th, January, 1999. Note that wind is generally strong by day, and weak by night. Symbols of locations are the same as Figure 4.

5. CONCLUSIONS

Statistical and geostatistical analyses of wind data in the mountainous Chubu and plain Kanto districts in central Japan have given the following conclusions: 1) wind speeds show exponential distributions independent of the districts, and wind is stronger in Kanto than in Chubu; 2) all temporal variograms of speeds, directions and velocities suggest daily duration, and wind is stronger by day than by night; 3) spatial variograms are classified into three types: the traditional type defined by a clear range (50–130 km) and sill, the flat type having only a sill, and a linear type where variogram values increase with increasing lag; and 4) the accumulation cannot change untypical type variograms to typical type ones.

Acknowledgments: The author would like to thank Dr. Chang-Jo F. Chung of Geological Survey of Canada for his critical

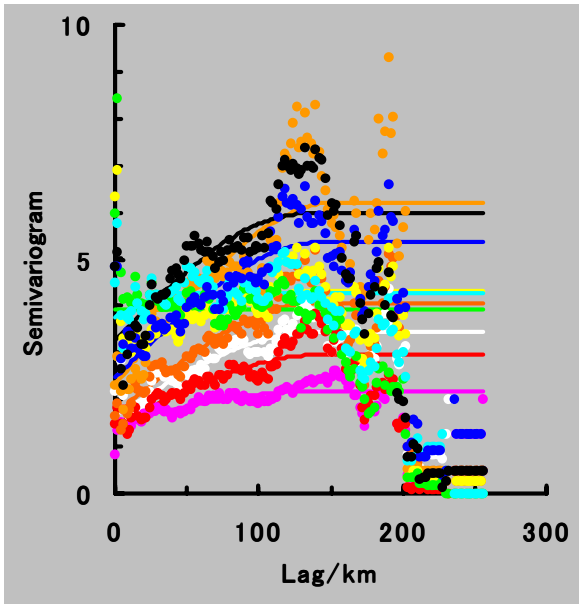

Figure 8. Spatial experimental semivariograms of wind speeds and fitted models in the Kanto district at every hour from 1 to 10 o'clock on 1st January 1999. The time sequence corresponds to the order of white-magenta-red-brown-orange-yellow-green-cyan-blue-black.

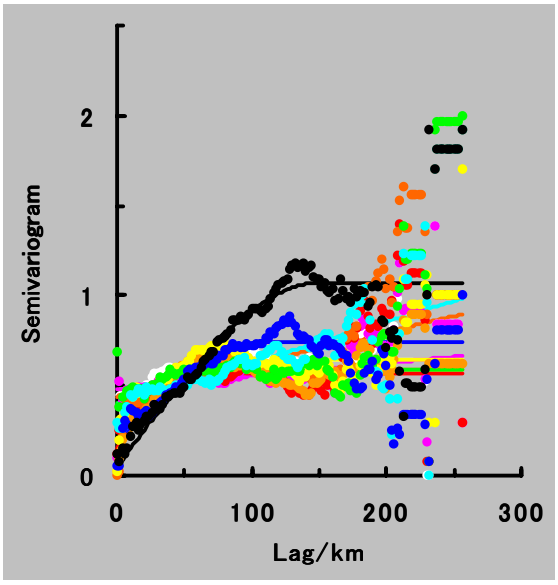

Figure 9. Spatial experimental semivariograms of wind directions and fitted models in the Kanto district at every hour from 1 to 10 o'clock on 1st January 1999t. Symbols of locations are the same as Figure 8.

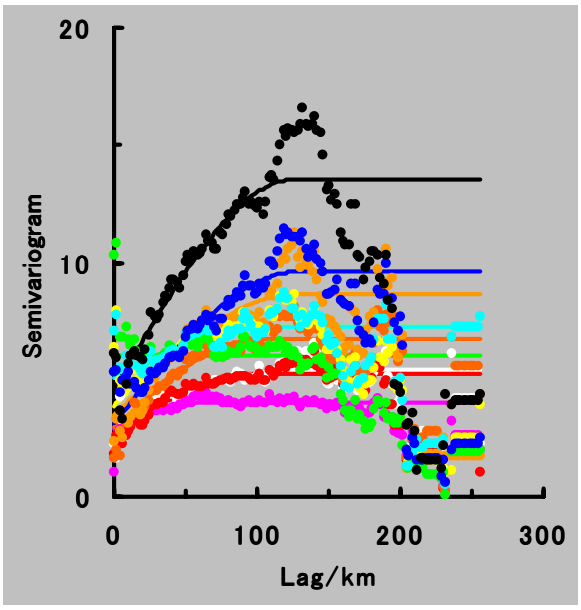

Figure 10. Spatial experimental semivariograms of wind velocities and fitted models in the Kanto district at every hour from 1 to 10 o'clock on 1st January 1999. Symbols of locations are the same as Figure 8.

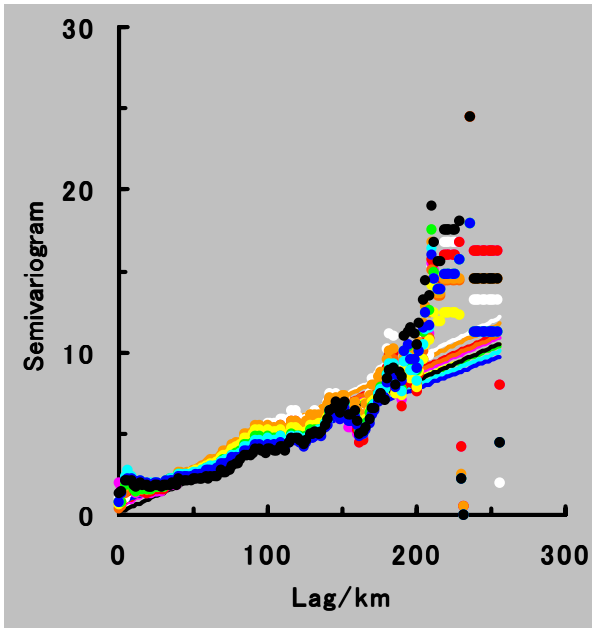

Figure 11. Spatial experimental semivariograms of averaged wind speeds and fitted models in the Kanto district from 14 to 23 o'clock on 7th January 1999. The durations from 1 to 9 hours are represented by the order of white-magenta-red-brown-orange-yellow-green-cyan-blue-black in color. Note that all variogram are not typical.

reading of the manuscript and valuable suggestions.

References

Shoji, T., and Kitaura, H., 2001. Statistical and geostatistical analysis of rain fall in central Japan. In Proc. IAMG2001 (Intern. Assoc. Math. Geol.), Cancum, Mexico, Sept. 6-12, 2001, CD-ROM (Session D).