

An Examination of Spatial Correlation in Hurricane Wind Fields

Shurong Fang¹, Yue Li², Weichiang Pang³, Fangqian Liu⁴

1. Department of Mathematical Sciences, Michigan Technological University
2. Department of Civil and Environmental Engineering, Michigan Technological University
3. Glenn Department of Civil Engineering, Clemson University
4. Glenn Department of Civil Engineering, Clemson University

Abstract

Each year hurricanes cause extensive economic loss and social disruption all around the world. Annual hurricane economic loss in the United States has been \$10 billion in recent years. Various hurricane wind field models have been proposed, and hurricane loss has been estimated based on these models. However, spatially correlated wind fields have not been considered yet. This paper examines spatial correlation in hurricane wind fields. In this paper, we describe the spatial correlation structure of hurricane wind fields and introduce the calculation of the spatial correlation using software R. The spatial correlation of wind fields is obtained by computing the semivariance of the standardized wind speed residuals between observed data from Real-time Hurricane Wind Analysis System (H*wind) and wind field model data. We consider the Exponential and Gaussian semivariogram models to fit the empirical semivariogram. As an illustrative example, the data from Hurricane Ivan (2004) is used to quantify the spatial correlations in wind field. Our analysis qualitatively determines the spatial correlation in the hurricane wind fields.

Key words: spatial correlation, semivariogram, hurricane wind fields, software R

Introduction

Hurricanes responsible for approximate \$10 billion USD economic loss each year in the United States (Pielke et al. 2005). Damages due to hurricanes are usually caused by the direct effects of hurricane induced strong winds. Various hurricane wind field models (Batts et al. 1980; Georgiou 1985; Vickery and Twisdale 1995; Vickery et al. 2000) have been developed over the years to model the spatial distribution of hurricane wind speeds. Most of these existing hurricane wind field models are deterministic and the modeling errors in the wind fields are not considered in these models. Uncertainty in hurricane wind fields is important for damage risk assessment and needs to be properly identified and included (Li and Ellinwood 2006). Risk assessment of spatially distributed systems such as networks of lifelines and portfolios of buildings require the modeling of hurricane wind hazard as a matrix of spatially distributed correlated wind speeds. In this paper, a methodology to quantify the spatial correlation of wind speed uncertainties of hurricane wind field model is introduced. As an illustrative example, the wind speed data from Hurricane Ivan (2004) is used to quantify the spatial correlations of a wind field model.

Hurricane Wind Fields Model

In this paper, the Georgiou (1985) wind field model is used to estimate the wind speeds. In order to compute the gradient wind speed (V), seven modeling parameters are required: the latitude, longitude, translational speed (V_T), heading angle (θ), central pressure (P_c), radius from hurricane eye to maximum wind location (R_{\max}), and air pressure profile parameter (B).

The gradient wind speed (V) at a point of interest (r, α) can be computed using the following equation (Georgiou 1985):

$$V = \frac{1}{2}(V_T \sin \alpha - fr) + \sqrt{\frac{1}{4}(V_T \sin \alpha - fr)^2 + \frac{B(P_a - P_c)}{\rho_a} \left(\frac{R_{\max}}{r}\right)^B e^{-(R_{\max}/r)^B}},$$

where $P_a - P_c$ is central peripheral pressure, ρ_a is the air density, r is the distance from the hurricane eye to the location of interest, and α is the angle between the location and the hurricane heading direction (counter-clockwise is positive). $f = 2\Omega \sin \phi$ is the Coriolis parameter in which Ω is the angular rotation rate of the earth (7.272×10^{-5} rad/s) and ϕ is the latitude at the location.

Figure 1 displays parameters in the wind field model and the track of Hurricane Ivan (2004) as it was making landfall in Alabama.

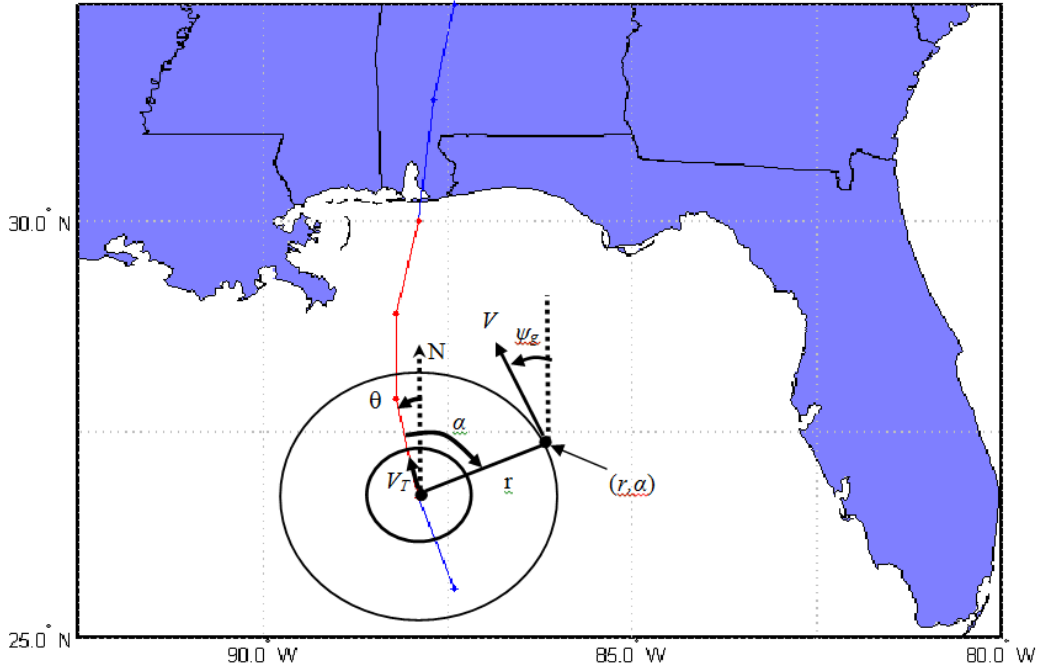


Figure 1: Hurricane Ivan (2004) track and gradient wind field model.

The error of a wind field model at the i^{th} location is defined as:

$$e_i = V_i - \tilde{V}_i,$$

where V_i is the observed wind speed (i.e. H*wind) and \tilde{V}_i is the estimated surface wind speed (i.e. $\tilde{V}_i \cdot F$) at the i^{th} location. The gradient-to-surface conversion factors (F) developed by Sparks and Huang (1999) are used to convert the gradient wind speed at the top of the boundary layer to the 10-m surface wind speed.

Spatial Correlation Methodology

In the real world, observations at different locations may not be independent. Measurements at nearby locations may be closer in value than those at locations farther apart. This phenomenon is called spatial correlation. Spatial correlation is the correlation of a variable with itself through space. In hurricane wind fields, we expect that the spatial correlation of observations at different locations is strong and positive when the locations are close, but it towards to zero as the locations are far away.

In theoretical geostatistics, the spatial correlation structure can be calculated through the semivariogram, which provides a measure of variance of the difference between observations at two locations across realizations of the field (Cressie 1993). Let u_l and $u_{l'}$ denote observations at two locations l and l' , respectively. In spatial statistics, the theoretical semivariogram $\gamma(l, l')$ is defined as

$$\gamma(l, l') = \frac{1}{2} \text{Var}(u_l - u_{l'}).$$

In Appendix, we show the empirical stationary semivariogram from data:

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{(l, l') \in N_h} (u_l - u_{l'})^2,$$

where N_h denotes the number of location pairs (l, l') with a distance $h = \|l - l'\|$. The empirical semivariogram for a particular distance is calculated by averaging one half of the difference squared of the values over all pairs of observations separated by the same distance. It is unlikely that actual location pairs would exactly have the distance h . So we consider a range of distances $[h - \delta, h + \delta]$ to modify N_h and calculate $\hat{\gamma}(h)$.

In applied geostatistics, the empirical semivariogram is often approximated by semivariogram models (Chiles and Delfiner 1999). There have been many positive definite semivariogram models proposed. Two important models are (Chiles and Delfiner 1999, Cressie 1993):

(1) Exponential model:

$$\hat{\gamma}(h) = c_0 + c_1 \left[1 - e^{(-h/a)} \right],$$

where c_0 denotes nugget effect, c_1 denotes the partial sill of the semivariogram, which is the variance of the observations, and a denotes the effective range (Duetsch and Journel 1992), which is the distance at which achieves 95% of the sill, $(c_0 + c_1)$.

(2) Gaussian model:

$$\hat{\gamma}(h) = c_0 + c_1 \left[1 - e^{(-h^2/a^2)} \right].$$

Assume that the nugget effect $c_0 = 0$, that is, the geological microstructure and measurement errors are ignored. The semivariogram is always 0 at distance $h = 0$.

The theoretical spatial correlation ρ can be estimated as $\hat{\rho}(h) = 1 - \gamma(h)/c_1$. Typically, the spatial correlation decreases as the distance h between the locations increases.

The spatial correlation structure in this paper refers to the correlation structure of the residuals in the wind field model. We explore the correlation structure of residuals at spatially distributed locations.

Results and Discussions

The observed wind speed data are from Real-time Hurricane Wind Analysis System (H*wind) (Powell et al. 1998). The spatial correlation is estimated based on the wind field records of Hurricane Ivan on September 16, 2004 at 1200 UTC time. To simplify

the computation, let $\tilde{e}_i = \frac{e_i}{\sigma}$, where σ is the standard deviation of the residual. Then the variance of \tilde{e}_i is 1. We only show the results of 400 randomly selected locations in this paper. Figure 2 shows the distribution of errors \tilde{e}_i at 400 locations.

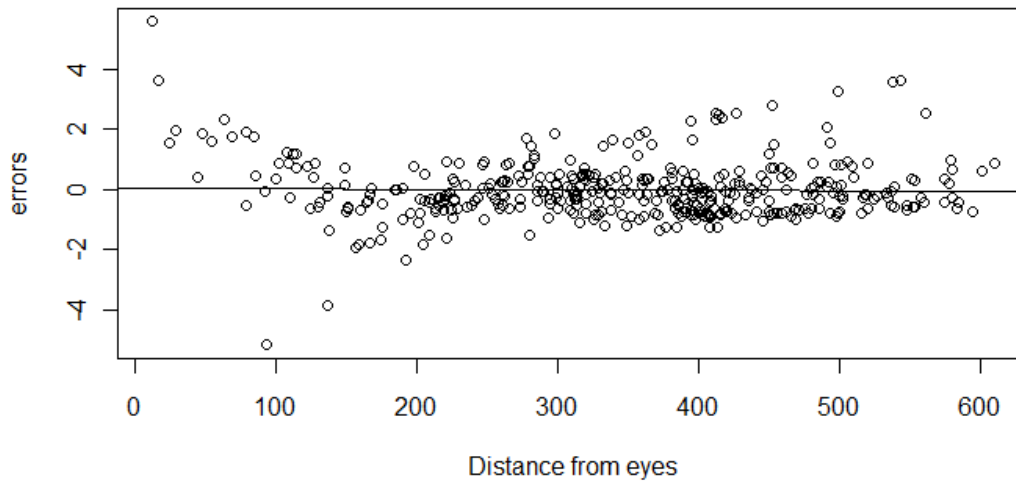


Figure 2. Distribution of errors \tilde{e}_i at 400 randomly selected locations. The x-axis denotes the distance between the location and the wind eye.

The software R (<http://www.r-project.org/>) is free to public. The package `geoR` in the software R provides geostatistical analysis on spatially distributed data (Diggle and Ribeiro, 2007; Ribeiro and Diggle, 2001). The command `plot(geodata)` gives distribution of the 400 locations and the errors \tilde{e}_i (Figure 3). The left-top panel displays the distribution of locations. The minimum distance between two locations is 6km. Different colors represent the different values of errors. The right-top and the left-bottom panels display the data errors versus coordinates in terms of longitudes and latitudes, respectively. The right-bottom panel is the histogram of errors.

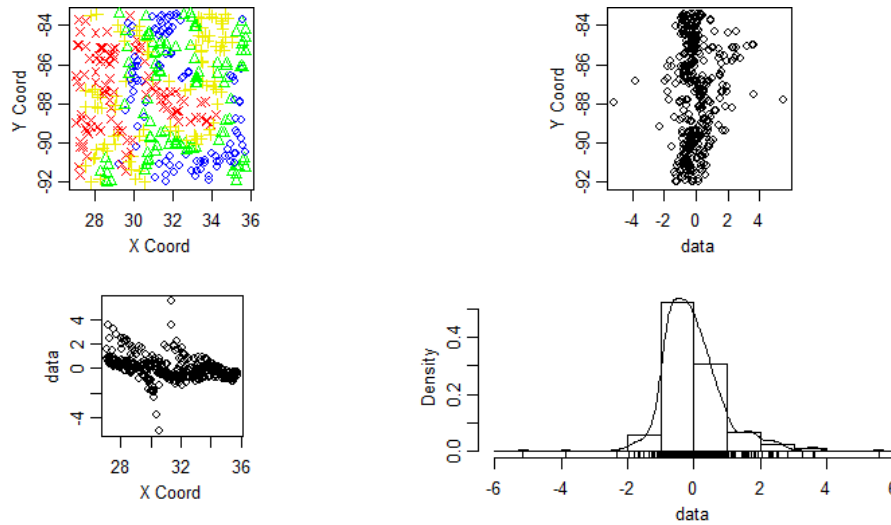


Figure 3. Distribution of the 400 locations and the data errors.

The function *variog* is used to estimate the empirical semivariogram. It accepts a large number of arguments. Since the correlation of errors becomes small and towards to 0 as the distance between two locations is large, the locations with distance greater than 200km are ignored. The command $v = \text{variog}(\text{geodata}, \text{max.dist}=1.8, \text{breaks}=\text{seq}(0, 1.8, l=11))$ returns the semivariogram of errors and the corresponding range of distance between different locations (Table 1). The argument *max.dist* specifies the maximum distance used in constructing the semivariogram. The argument *breaks* specifies the endpoints of the intervals of distances.

Semivariogram	Range of distance (km)	Correlation ρ
0.1527481	0.00~19.93	0.84725194
0.1842556	19.93~39.85	0.81574442
0.2776118	39.85~59.78	0.72238823
0.3381029	59.78~79.70	0.66189713
0.5674614	79.70~99.63	0.43253856
0.5974252	99.63~119.56	0.40257475
0.6826387	119.56~139.48	0.31736127
0.7703401	139.48~159.41	0.22965992
0.8637183	159.41~179.34	0.13628167
0.9880528	179.34~200.00	0.01194723

Table 1. Semivariogram, range of distance between different locations, and spatial correlation of errors.

The correlation ρ can be obtained based on the empirical semivariogram in Table 1. A large distance between two locations corresponds to a large semivariogram (Figure 4), and subsequently a small correlation.

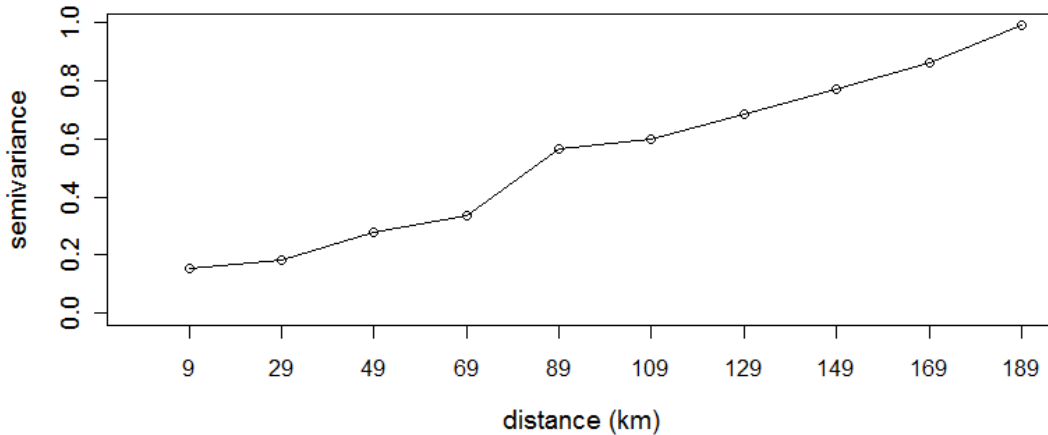


Figure 4. Empirical semivariogram plot.

In practice, the parameters of semivariogram, such as effective range a , are unknown. We use the Exponential and Gaussian semivariogram models to fit the empirical semivariogram. In this paper, the Weighted Least Squares (WLS) is used to obtain the estimates of parameters. In the package `geoR`, the function `variofit` can estimate the parameters. Since the nugget effect $c_0 = 0$ and partial sill $c_1 = 1$, we only need to estimate the effective range a . The following commands give us the estimated effective range by WLS under the Exponential and Gaussian semivariogram models.

```
wlsexp = variofit( v, ini=c(1,1), cov.model="exp", fix.nugget=TRUE )
```

```
wlsgaus = variofit( v, ini=c(1,1), cov.model="gaus", fix.nugget=TRUE )
```

The argument `ini` specifies the initial values of parameters c_1 and a . The argument `cov.model` specifies the semivariogram model. The argument `fix.nugget=TRUE` specifies the nugget effect is fixed to 0. We can also use the argument `pars.limits` to give limits of parameters. The effective ranges under the Exponential and Gaussian models are 103km and 110km, respectively. Figure 5 shows the plot of fitted models against the empirical semivariogram.

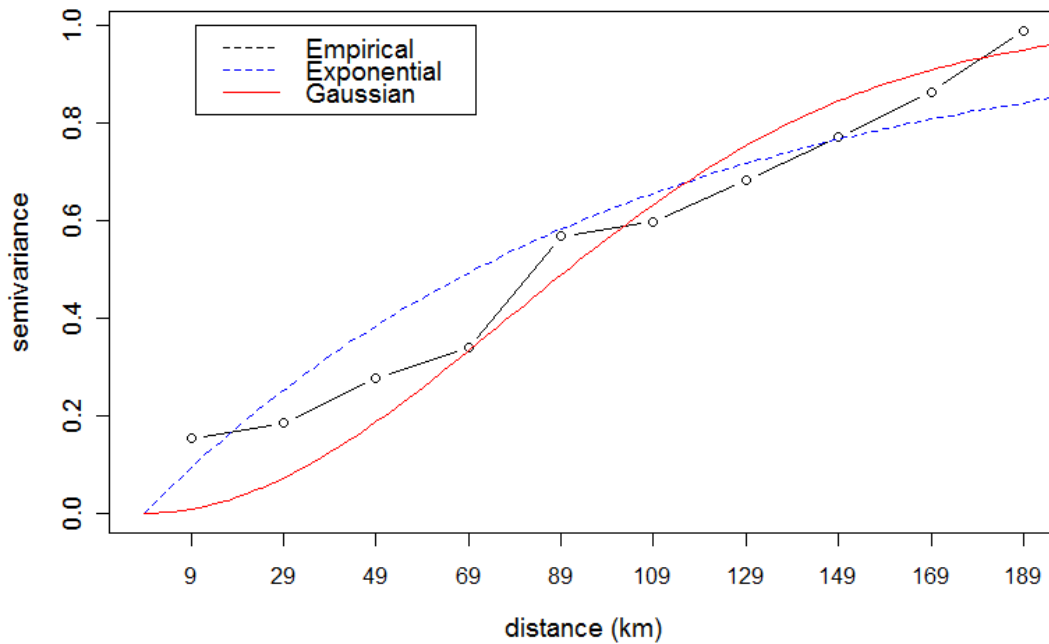


Figure 5. Plot of empirical semivariogram and two fitted semivariogram models: Exponential and Gaussian models.

We investigate the variability of the empirical semivariogram over the Exponential and Gaussian semivariogram models through an envelope analysis. The function *variog.mc.env* computes the envelopes for empirical semivariograms by permutation. For each simulation, data \tilde{z}_i 's are randomly allocated to the spatial locations. Then, the empirical semivariogram is computed using the same distance ranges as for the semivariogram originally computed for the data. The envelopes are calculated by taking the maximum and minimum semivariogram values from each distance range of the simulated data. The argument *obj.var* is typically an output of the function *variog*. The function *variog.model.env* computes the envelopes for a semivariogram by simulating data under a given model. The argument *model* specifies the parameters under a semivariogram model. The envelopes are computed under a semivariogram model. Simulated data values are generated at the spatial locations, given the model parameters. Then, in each simulation, the semivariogram is computed using the same distance ranges. The envelopes are the maximum and minimum values of the semivariogram of each distance range. For both functions, we can use the argument *nsim* to specify the number of simulations used to compute the envelopes. The default is 99. The commands as follow are used to calculate the envelope without spatial correlation assumption and the envelopes based on the Exponential and Gaussian semivariogram parameters.

```
env = variog.mc.env( geodata, obj.var=v )
```

```
envexp = variog.model.env( geodata, obj.var=v, model=wlsexp )
```

```
envgaus = variog.model.env( geodata, obj.var=v, model=wlskaus )
```

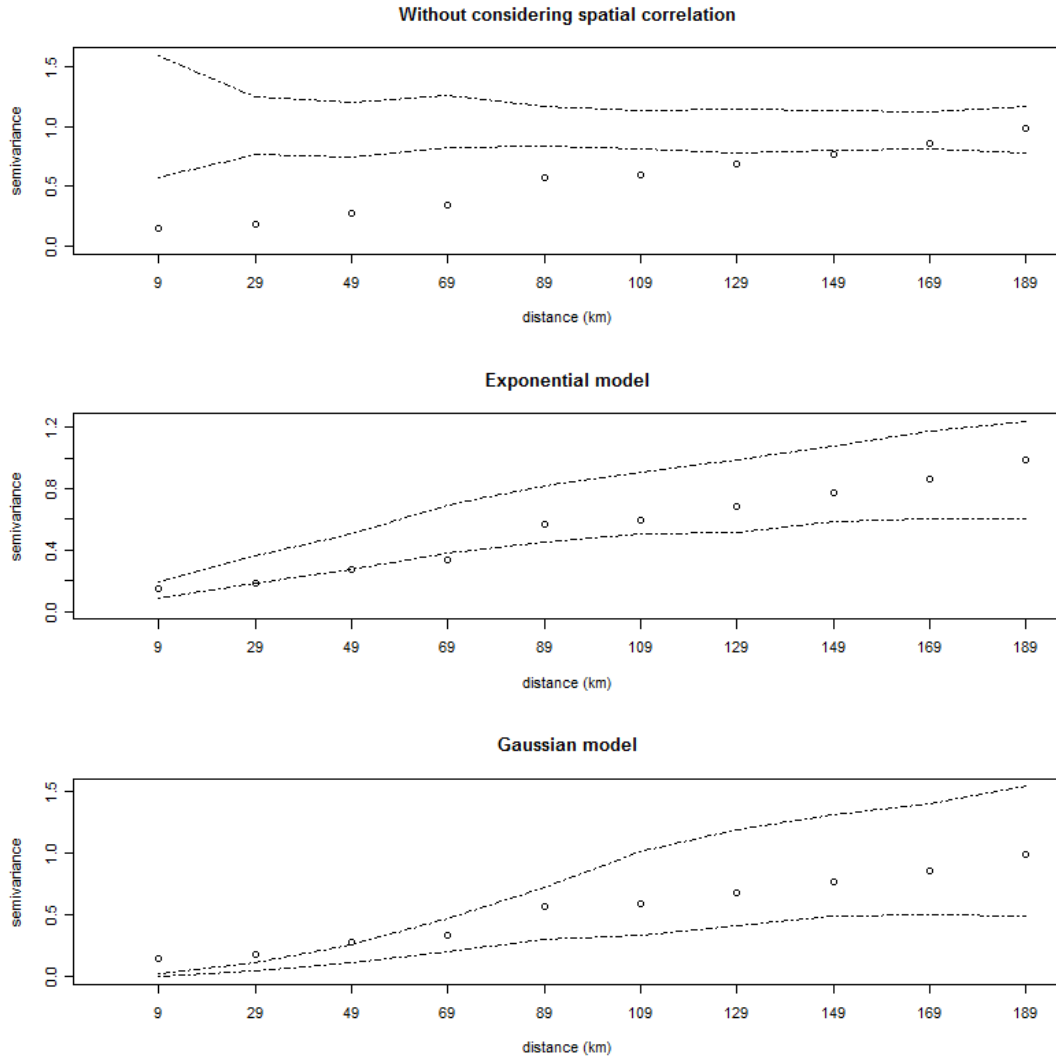



Figure 6: Envelopes of semivariogram models.

The top panel of Figure 6 illustrates the empirical envelope for semivariogram under the assumption of no spatial correlation between two locations. It is based on the permutation over the spatial locations. It is obvious that there is spatial dependence as most empirical semivariogram points fall outside of the envelope. The middle panel shows the envelope based on simulations from Exponential semivariogram model. The envelope of Exponential semivariogram includes almost all points. The bottom panel displays the envelope based on simulations from Gaussian semivariogram model. The envelope of semivariogram under the Gaussian model contains most semivariogram points. Based on the analysis of semivariogram envelopes, the Exponential model with a range of 103km appears to be the best model based on these data.

Recently, several hurricane wind field models have been proposed. However, the spatial correlation of wind fields has been underdeveloped in these models. Our preliminary analysis provides obvious evidence of spatial correlation in the hurricane wind fields. We highly suggest that the spatial correlation structure of wind fields should be considered when hurricane risk assessment of spatially distributed systems, such as networks of

lifelines and portfolios of buildings, is conducted. Our findings have implications for future risk analysis and loss estimation in wind fields.

In the illustrative example, the observed hurricane wind fields for Hurricane Ivan (H^* wind), we fit the empirical semivariogram with the Exponential and Gaussian models, because they are commonly used continuous functions. Other permissible models can be utilized as well. One can also estimate model parameters using Maximum Likelihood (ML) or Ordinary Least Squares (OLS) methods.

The analysis shows that spatial correlation exists under the wind field model and the Exponential semivariogram model fits the best of our data. Note that the Exponential model may not be the best choice for any wind field model. When different data is considered, one should fit the empirical semivariogram and construct the envelopes to determine the best model based on those data.

A free software R is used to examine the spatial correlation structure of the hurricane wind fields. In this paper, we present how to use the package geoR to perform the examination of spatial correlation. Of course, other softwares such as Matlab can be considered as well.

References:

- Batts, M.E., Cordes, M.R., Russell, C.R., Shaver, J.R., Simiu, E., (1980). "Hurricane Wind Speeds in the United States," *National Bureau of Standards Report No. BSS-124*, US Department of Commerce, Washington, D.C.
- Chiles J.P. and Delfiner P. (1999) "Geostatistics: Modeling Spatial Uncertainty" *Wiley*.
- Cressie, N. (1993). "Statistics for Spatial Data" *Wiley*.
- Diggle, P.J. and Ribeiro, Jr. P.J. (2007) "Model Based Geostatistics" *Springer*, New York.
- Georgiou, P.N., (1985). *Design Wind Speeds in Tropical Cyclone-prone Regions*. PhD Dissertation, Department of Civil Engineering, University of Western Ontario, Canada.
- Li, Y. and Ellingwood, B.R. (2006) "Hurricane Damage to Residential Construction in the US; Importance of Uncertainty Modeling in Risk Assessment." *Engineering Structures*, 28(7), 1009-1018.
- Pielke, R.A., Gratz, J., Landsea, C.W., Collins, D., Saunders, M.A., and Musulin, R.(2005). "Normalized hurricane damage in the United States: 1900-2005," *Natural Hazards Review*, 9(1):29-42.
- Powell, M.D., Houston, S.H., Amat, L.R., and Morisseau-Leroy, H., (1998). "The HRD Real-time Hurricane Wind Analysis System," *J. of Wind Engineering and Indust. Aerodyn.*, 77&78, pp.53-64.
- Ribeiro, Jr. P. J. and Diggle, P. J. (2001) "geoR: a package for geostatistical analysis" *R-NEWS*, 1(2):15-18.

Sparks, P.R., and Huang, Z., (1999). “Wind Speed Characteristics in Tropical Cyclones,” *Proc., 10th International Conference on Wind Engineering*, Copenhagen, Denmark.

Vickery, P.J., Skerlj, P.F., Steckley, A.C., and Twisdale, L.A., (2000). “Hurricane Wind Field Model for use in Hurricane Simulations,” *ASCE Journal of Structural Engineering*, 126(10), pp. 1203-1221.

Vickery, P. J., and Twisdale, A. (1995). “Wind-field and filling models for hurricane wind-speed predictions.” *ASCE Journal of Structural Engineering*, 121(11), 1700–1709.

Appendix:

The theoretical semivariogram of two observations and two distinct locations is

$$\begin{aligned}\gamma(l, l') &= \frac{1}{2} \text{Var}(u_l - u_{l'}) \\ &= \frac{1}{2} E \left[\left((u_l - u_{l'}) - E(u_l - u_{l'}) \right)^2 \right] \\ &= \frac{1}{2} E \left[(u_l - u_{l'})^2 \right]\end{aligned}$$

Assume that the distance between locations l and l' is $h = \|l - l'\|$, and there are N_h such location pairs, then

$$E \left[(u_l - u_{l'})^2 \right] = \frac{1}{N_h} \cdot \sum_{(l, l') \in N_h} (u_l - u_{l'})^2.$$

Thus, the empirical semivariogram based on the data is

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{(l, l') \in N_h} (u_l - u_{l'})^2.$$