

PREDIKSI DAN INTERPOLASI MELALUI ORDINARY KRIGING: STUDI KASUS KEMISKINAN DI PROVINSI JAWA TIMUR

Rokhana Dwi Becti

Mathematics & Statistics Department, School of Computer Science, Binus University
Jl. K.H. Syahdan No. 9, Palmerah, Jakarta Barat 11480
rokhana_db@binus.ac.id

ABSTRACT

Prediction and interpolation in geostatistics are kind of spatial statistical analysis to estimate the data in locations which can not be sampled (missing data) based on existing data yet still need a model. Geostatistics is also highly correlated with data autocorrelation. In data about poverty, prediction and interpolation are useful to provide information of poverty relationship patterns among locations and predict level of poverty in certain locations. Therefore, this study analyzes spatial autocorrelation through Moran's I as well as the prediction and interpolation by semivariogram and ordinary kriging on poverty Data. Then, it compares other prediction and interpolation models (Gaussian, Exponential, and Spherical). A case study is conducted in East Java. The analysis results show that through moran's I test appears a spatial autocorrelation on some percentages of poor citizens. Gaussian model is a model that gives better predictions than other models.

Keywords: *spatial autocorrelation, semivariogram, ordinary kriging, poverty*

ABSTRAK

Prediksi dan interpolasi dalam geostatistik merupakan analisis statistik spasial untuk menduga data pada suatu lokasi yang tidak dapat dilakukan sampling (missing data) berdasarkan data yang sudah ada dan tetap membutuhkan suatu model. Geostatistik juga sangat berhubungan dengan autokorelasi data. Dalam hal data kemiskinan, prediksi dan interpolasi bermanfaat untuk memberikan informasi pola hubungan kemiskinan antar lokasi dan memprediksi angka kemiskinan pada lokasi-lokasi tertentu. Oleh karena itu, penelitian ini melakukan analisis autokorelasi spasial melalui moran's I serta prediksi dan interpolasi melalui semivariogram dan ordinary kriging pada data kemiskinan. Selanjutnya membandingkan model prediksi dan interpolasi yang berbeda (Gaussian, Exponential, dan Spherical). Studi kasus dilakukan di Propinsi Jawa Timur. Hasil analisis menunjukkan bahwa melalui uji moran's I terdapat autokorelasi spasial pada persentase jumlah penduduk miskin. Model Gaussian adalah model yang memberikan prediksi yang lebih baik dibandingkan model lainnya.

Kata kunci: *autokorelasi spasial, semivariogram, ordinary kriging, kemiskinan*

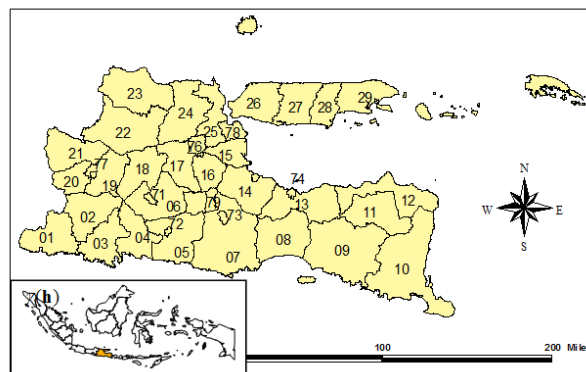
PENDAHULUAN

Geostatistik merupakan kumpulan teknik numerik yang berhubungan dengan karakteristik atribut spasial (Bohling, 2007:2). Geostatistik juga sangat berhubungan dengan autokorelasi data. Sementara itu, analisis data spasial tersebut merupakan analisis yang berhubungan dengan pengaruh lokasi. Hal ini didasarkan pada hukum pertama tentang geografi dikemukakan oleh Tobler menyatakan bahwa segala sesuatu saling berhubungan satu dengan yang lainnya, tetapi sesuatu yang dekat lebih mempunyai pengaruh daripada sesuatu yang jauh (Anselin dan Rey, 2010:17). Pada data spasial, seringkali pengamatan di suatu lokasi bergantung pada pengamatan di lokasi lain yang berdekatan (*neighboring*). Statistik spasial dan geostatistik sendiri telah dikembangkan untuk menggambarkan dan menganalisis variasi dalam kedua fenomena alam dan buatan manusia pada, baik di atas atau di bawah permukaan tanah (Fischer dan Getis, 2010:5).

Proses pendugaan data pada suatu lokasi yang tidak dapat disampling (*missing data*) membutuhkan suatu model. Namun beberapa penelitian memiliki permasalahan, di antaranya tidak ada model, hanya ada satu sampel data atau tidak ada teknik inferensia yang dapat digunakan untuk memprediksi data yang tidak dapat disampling. Geostatistik sangat berperan dalam hal tersebut, yaitu menggunakan metode estimasi dengan tetap didasarkan pada model. Salah satu teknik geostatistik untuk prediksi dan interpolasi adalah *semivariogram* dan *ordinary kriging*. Beberapa model *semivariogram* dan *ordinary kriging* tersebut adalah *Gaussian*, *Exponential* dan *Spherical*. Model-model tersebut juga memberikan prediksi yang berbeda-beda. Di kasus kemiskinan, karakteristik sosial, ekonomi, sumberdaya alam, penduduk, dan pendidikan yang berbeda di setiap lokasi yang berbeda menimbulkan permasalahan kemiskinan yang berbeda pula. Lokasi yang memiliki karakteristik yang sama akan memiliki permasalahan kemiskinan yang hampir sama. Hal ini menunjukkan bahwa kasus kemiskinan juga perlu dikaji secara spasial. Beberapa penelitian yang menganalisis kemiskinan dengan metode spasial adalah Kam, Bose, dan Villano (2005:1) dan Bakti (2011:1). Berdasarkan kajian teori dan permasalahan yang ada, dalam penelitian ini dilakukan analisis autokorelasi spasial, prediksi dan interpolasi kemiskinan. Hal ini bertujuan untuk memberikan informasi pola hubungan karakteristik kemiskinan antar lokasi di Jawa Timur. Selanjutnya, didapatkan informasi prediksi kemiskinan pada lokasi-lokasi tertentu, sehingga dapat direkomendasikan untuk menentukan suatu kebijakan yang terpadu antar wilayah.

METODE

Data yang digunakan dalam penelitian ini adalah jumlah penduduk miskin Jawa Timur 2007 yang diperoleh dari data Badan Pusat Statistika (BPS). Peta lokasi dan kode kabupaten/kota disajikan pada Gambar 1 dan Tabel 1.



Gambar 1 Lokasi penelitian

Tabel 1 Kode Kabupaten/Kota di Jawa Timur

No	Kab/Kota	No	Kab/Kota	No	Kab/Kota	No	Kab/Kota
01	Kab. Pacitan	11	Kab. Bondowoso	21	Kab. Ngawi	72	Kota Blitar
02	Kab. Ponorogo	12	Kab. Situbondo	22	Kab. Bojonegoro	73	Kota Malang
03	Kab. Trenggalek	13	Kab. Probolinggo	23	Kab. Tuban	74	Kota Probolinggo
04	Kab. Tulungagung	14	Kab. Pasuruan	24	Kab. Lamongan	75	Kota Pasuruan
05	Kab. Blitar	15	Kab. Sidoarjo	25	Kab. Gresik	76	Kota Mojokerto
06	Kab. Kediri	16	Kab. Mojokerto	26	Kab. Bangkalan	77	Kota Madiun
07	Kab. Malang	17	Kab. Jombang	27	Kab. Sampang	78	Kota Surabaya
08	Kab. Lumajang	18	Kab. Nganjuk	28	Kab. Pamekasan	79	Kota Batu
09	Kab. Jember	19	Kab. Madiun	29	Kab. Sumenep		
10	Kab. Banyuwangi	20	Kab. Magetan	71	Kota Kediri		

Metode analisis yang digunakan adalah: (1) menganalisis autokorelasi spasial melalui Moran's I; (2) membagi data menjadi dua bagian, yaitu 90% (34 lokasi) data aktual untuk pemodelan dan 10% (4 lokasi) untuk prediksi dan validasi; (3) membentuk *semivariogram* dan model variogram; (4) melakukan prediksi dan interpolasi melalui *ordinary kriging*, dengan model variogram *gaussian*, *exponential*, dan *spherical*; (5) membandingkan hasil prediksi model variogram *gaussian*, *exponential*, dan *spherical* melalui MSE.

Moran's I

Autokorelasi spasial merupakan salah satu analisis spasial untuk mengetahui pola hubungan atau korelasi antar lokasi (amatan). Beberapa pengujian dalam spasial autokorelasi spasial adalah Moran's I, Rasio Geary's, dan Local Indicator of Spatial Autocorrelation (LISA). Metode ini sangat penting untuk mendapatkan informasi mengenai pola penyebaran karakteristik suatu wilayah dan keterkaitan antar lokasi didalamnya. Selain itu, metode ini juga digunakan untuk identifikasi pemodelan spasial. Beberapa penelitian yang telah menggunakan metode autokorelasi spasial adalah Kissling dan Carl (2008:3) di bidang pemodelan ekologi, Bektid dan Sutikno (2011:1) dalam analisis autokorelasi data kemiskinan.

Koefisien Moran's I merupakan pengembangan dari korelasi pearson pada data univariate series. Koefisien Moran's I digunakan untuk uji dependensi spasial atau autokorelasi antar amatan atau lokasi. Perhitungan Morans'I (Lee dan Wong, 2001):

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Keterangan:

- x_i = data variabel lokasi ke-i ($i = 1, 2, \dots, n$)
- x_j = data variabel lokasi ke-j ($j = 1, 2, \dots, n$)
- \bar{x} = rata-rata data
- w = matrix pembobot

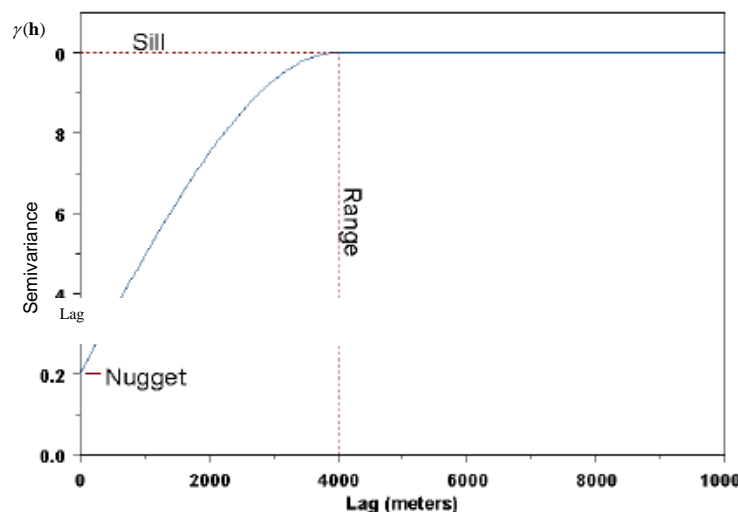
$$E(I) = I_o = -\frac{1}{n-1} \quad (2)$$

Nilai dari indeks I adalah antara -1 dan 1. Apabila $I > I_o$, data memiliki autokorelasi positif. Jika $I < I_o$, data memiliki autokorelasi negatif.

Semivariogram

Semivariogram merupakan plot antara semivariance dan lag yang menunjukkan karakteristik korelasi spasial. Bentuk umum *semivariogram* disajikan pada Gambar 2, dengan komponen-komponen:

- Sill* : nilai semivariance ketika level semivarians mencapai konstant. Juga digunakan untuk menunjukkan “amplitude” dari komponen tertentu dari *semivariogram*.
- Range* : jarak lag Di mana *semivariogram* (atau komponen *semivariogram*) mencapai nilai sill.
- Nugget* : Menurut teori, nilai *semivariogram* pada titik asal (lag 0) seharusnya nol. Jika nilai tersebut secara signifikan berbeda dengan nol pada lag yang mendekati nol, maka nilai *semivariogram* tersebut dinamakan nugget. Nugget menunjukkan variabilitas sampel pada jarak yang lebih kecil dari sampel tertentu, termasuk juga kesalahan pengukuran.



Gambar 2 *Semivariogram*.

Nilai *semivariogram* diperoleh dari persamaan berikut:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_{\alpha} + \mathbf{h}) - z(\mathbf{u}_{\alpha})]^2 \quad (3)$$

di mana:

- \mathbf{u} : vektor dari koordinat spasial (terdiri dari komponen x dan y atau *easting* dan *northing*)
- $z(\mathbf{u})$: variabel
- \mathbf{h} : vektor lag antara dua amatan (lokasi)
- $z(\mathbf{u}+\mathbf{h})$: nilai variabel setelah dilag kan.
- $N(\mathbf{h})$: jumlah pasangan amatan pada lag h

Pola *semivariogram* diperlukan pada analisis kriging. Pola *semivariogram* tersebut selain diperoleh dari pendekatan bentuk empiris juga perlu dari pendekatan numerik, yaitu model-model *semivariograms*. Persamaan 4, 5, dan 6 adalah beberapa model *semivariogram*, dengan a range, c sill, dan h lag distance.

$$\text{Spherical} \quad : \quad g(\mathbf{h}) = \begin{cases} c \left(1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right) & \text{jika } h \leq a \\ c & \text{lainnya} \end{cases} \quad (4)$$

$$\text{Exponensial} \quad : \quad g(\mathbf{h}) = c \left(1 - \exp \left(\frac{-3h}{a} \right) \right) \quad (5)$$

$$\text{Gaussian} \quad : \quad g(\mathbf{h}) = c \left(1 - \exp \left(\frac{-3h^2}{a^2} \right) \right) \quad (6)$$

Kriging

Kriging merupakan salah satu metode prediksi dan interpolasi dalam geostatistika, terdiri dari dua jenis yaitu *ordinary kriging* ketika hanya satu variabel dan *cokriging* ketika terdapat lebih dari satu variabel yang diamati. Definisi interpolation dalam hal ini adalah metode untuk menghasilkan sebuah prediction surface yang bersifat kontinyu dari sekelompok sampel data. *Interpolation analysis* diperlukan karena data tidak mungkin diambil dari semua lokasi yang ada. Teknik interpolasi mengambil data di sebagian lokasi dan menghasilkan nilai prediksi untuk lokasi lainnya.

Suatu sampeal data pada lokasi $1, 2, \dots, n$ adalah $V(x_1), V(x_2), \dots, V(x_n)$, maka untuk menduga $V(x_0)$ adalah:

$$\hat{V}(x_0) = \sum_{i=1}^n w_i V(x_i) \quad (7)$$

di mana w_i diduga melalui matrik berikut:

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{D} \quad (8)$$

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \\ \mu \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \dots & \bar{C}_{1n} & 1 \\ \bar{C}_{21} & \bar{C}_{22} & \dots & \bar{C}_{2n} & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \bar{C}_{n1} & \bar{C}_{n2} & \dots & \bar{C}_{nn} & 1 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} \bar{C}_{10} \\ \bar{C}_{20} \\ \dots \\ \bar{C}_{n0} \\ 1 \end{bmatrix}$$

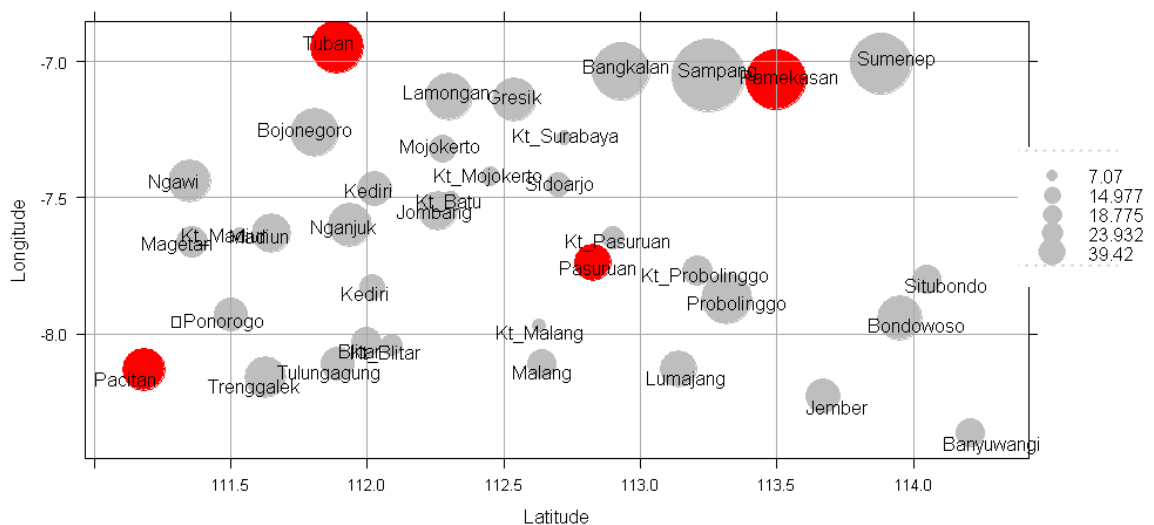
\bar{C}_{mn} merupakan nilai kovarian antar masing-masing amatan dan \bar{C}_{n0} merupakan nilai kovarian antar amatan dengan amatan yang akan diduga.

HASIL DAN PEMBAHASAN

Kemiskinan sebagai salah satu permasalahan utama di negara Indonesia perlu terus dikaji untuk upaya penyelesaian. Dari tahun ke tahun, jumlah penduduk miskin mengalami penurunan, seperti yang ditunjukkan oleh data BPS, Di mana jumlah penduduk miskin di Indonesia 2011 adalah 30.018.930 jiwa (12,47%). Angka ini menurun dibandingkan tahun sebelumnya, yaitu 31.023.400 jiwa (13,33%). Namun berbagai upaya perlu terus dilakukan untuk meningkatkan kesejahteraan masyarakat.

BPS Propinsi Jawa Timur menyebutkan bahwa penduduk miskin Jawa Timur pada bulan September 2011 sebanyak 5,227 juta (13,85 persen) atau turun 2,41 persen dibandingkan dengan jumlah penduduk miskin pada bulan Maret 2011 yang sebesar 5,356 juta (14,23 persen). Di beberapa lokasi di Jawa Timur, seperti pada beberapa kabupaten di kepulauan Madura memiliki jumlah penduduk miskin relatif sama dan lebih tinggi dibandingkan dengan kabupaten/kota lainnya. Berdasarkan hal tersebut, dapat dikatakan bahwa penyebaran angka kemiskinan di beberapa lokasi menunjukkan pembentukan pola-pola tertentu. Adanya pola-pola ini juga menunjukkan adanya hubungan antar lokasi.

Pola penyebaran persentase penduduk miskin di Jawa Timur 2007 disajikan di Gambar 3. Jumlah penduduk miskin di Jawa Timur tahun 2007 adalah 19,98 persen (7.155,3 juta jiwa). Angka ini mengalami penurunan dibandingkan tahun 2006 yang berjumlah 21,09 persen (7.678,1 juta jiwa). Persentase penduduk miskin tersebut membentuk pola-pola mengelompok pada lokasi-lokasi tertentu. Pengelompokan kabupaten/kota yang ada juga hampir sama. Kabupaten Sampang masih memiliki persentase yang tinggi, yaitu 39,42 persen. Selanjutnya Kabupaten Bangkalan, Sumenep, dan Pamekasan di Pulau Madura, yang berlokasi saling berdekatan dan memiliki persentase penduduk miskin lebih tinggi dibandingkan kabupaten/kota lainnya. Seperti halnya di beberapa kabupaten lainnya seperti Kabupaten Tuban, Bojonegoro, Lamongan, dan Gresik. Daerah tersebut memiliki persentase yang hampir sama. Warna merah menunjukkan empat lokasi (Kabupaten Pacitan, Tuban, Pamekasan, dan Pasuruan) yang digunakan sebagai prediksi dan validasi. Pemilihan lokasi ini dilakukan secara random.



Gambar 3 Pola penyebaran persentase penduduk miskin Jawa Timur 2007

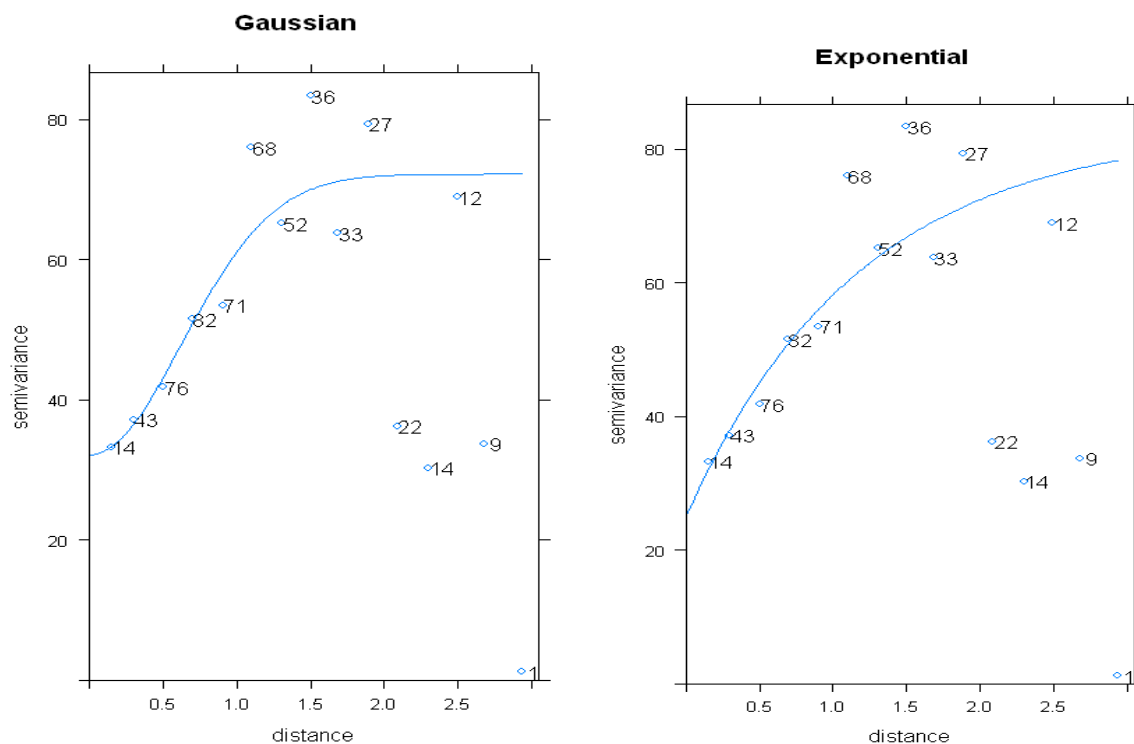
Autokorelasi Spasial Moran's I

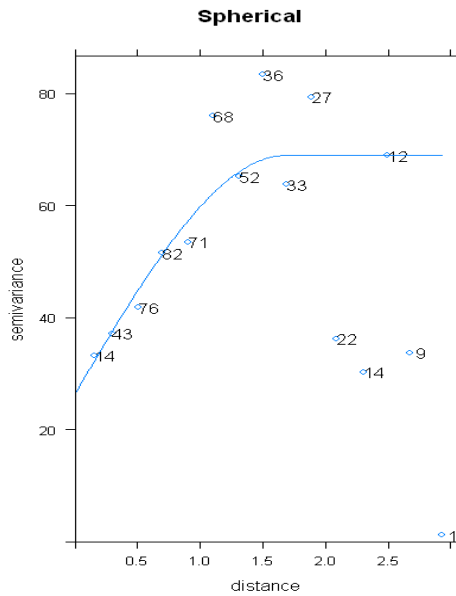
Autokorelasi spasial ini menunjukkan bahwa ada keterkaitan atau hubungan nilai persentase penduduk miskin antar kabupaten/kota di Jawa Timur. Seperti pada identifikasi sebelumnya, bahwa ada pengelompokan pada beberapa lokasi. Angka Moran's I pada tahun 2007 adalah 0.5434 yang lebih besar dari $I_0 = -0,027$. Hal ini menunjukkan ada autokorelasi positif atau pola yang mengelompok dan memiliki kesamaan karakteristik pada lokasi yang berdekatan.

Semivariogram

Plot *semivariogram* pola penduduk miskin dibentuk ke dalam model *spherical*, eksponensial, dan *gaussian* (Gambar 4). Titik-titik data menunjukkan *semivariogram* berdasarkan data experimental (aktual) dan garis menunjukkan *semivariogram* berdasarkan teori (model). Secara umum, *semivariogram* tersebut memiliki nilai sill 70, range 1.5, dan nugget 30. *Semivariogram* ini menunjukkan pasangan data, jarak antar data, nilai semivariance, serta korelasi spasial.

Melalui perbandingan model *spherical*, eksponensial, dan *gaussian* dapat diketahui bahwa *semivariogram* model *gaussian* lebih sesuai untuk data eksperimental. Hal ini ditunjukkan oleh plot data experimental yang mengikuti pola model *gaussian*. Pada semivariance terbentuk 15 pasangan data yang memiliki kesamaan karakteristik dan jarak yang relatif dekat antara satu lokasi dan lokasi lainnya. Pasangan data pertama terdiri dari 14 data dengan lag (distance) 0.151 dan semivariance 33.233. Nilai semivariance semakin konstan pada lag 1.5, yang menunjukkan bahwa pasangan data dengan lag di atas 1.5 dan semivariance sekitar 70 memiliki korelasi yang kecil atau dapat dikatakan tidak ada korelasi.



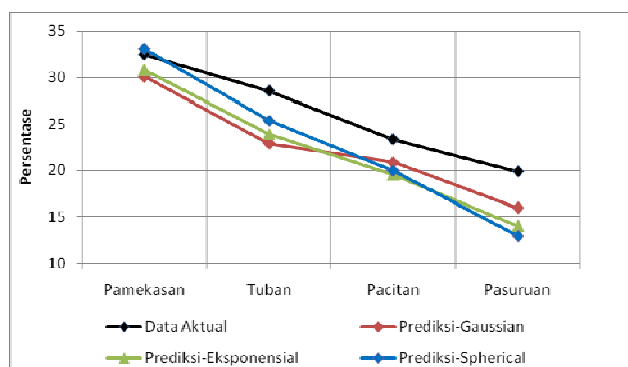


Gambar 4 Semivariogram model gaussian, eksponensial, dan spherical.

Ordinary Kriging

Hasil prediksi persentase penduduk miskin melalui ordinary kriging disajikan pada Gambar 5. Perhitungan prediksi dilakukan dengan menggunakan model *spherical*, *exponential*, maupun *gaussian*. Masing-masing model menghasilkan prediksi yang berbeda-beda. Persentase penduduk miskin di Kabupaten Pacitan tahun 2007 adalah 23.27%. Hasil prediksi melalui model *Gaussian*, *exponential*, dan *spherical* adalah 20.84%, 19.51%, dan 19.96%. Sementara itu, persentase penduduk miskin di Kabupaten Pasuruan adalah 19.88 dan hasil prediksi melalui model *Gaussian*, *exponential*, dan *spherical* adalah 15.97%, 14.07%, dan 12.95%.

Hasil prediksi kemiskinan di kedua lokasi Kabupaten Pacitan dan Pasuruan lebih baik dengan menggunakan model *gaussian*, karena mendekati nilai data aktualnya dan error yang kecil. Sementara itu hasil prediksi terbaik Kabupaten Pamekasan dan Tuban adalah *spherical*. Namun secara umum, berdasarkan perbandingan MSE dapat diketahui bahwa model *gaussian* memberikan prediksi yang lebih baik. Hal ini ditunjukkan oleh nilai MSE adalah paling kecil dibandingkan model *spherical* dan *gaussian* (Tabel 2).

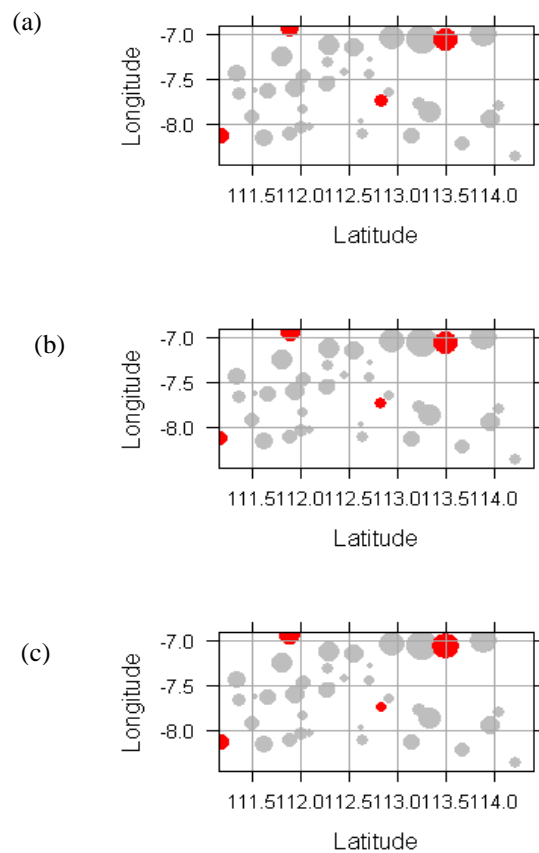


Gambar 5 Prediksi persentase penduduk miskin melalui ordinary kriging.

Tabel 2 Perbandingan Nilai MSE

Model	MSE
<i>Gaussian</i>	14.63071
Eskponensial	18.01143
<i>Spherical</i>	17.37903

Perbandingan hasil prediksi dengan lokasi-lokasi disajikan pada Gambar 6. Gambar tersebut menunjukkan bahwa hasil prediksi di keempat lokasi masih memiliki karakteristik yang sama dengan lokasi-lokasi lain yang bertetanggan atau berdekatan. Kabupaten Pasuruan dengan nilai prediksi kemiskinan kecil dikelilingi oleh kabupaten/kota lain yang kemiskinannya kecil pula.



Gambar 6 Perbandingan prediksi persentase penduduk miskin dengan lokasi lain pada model *gaussian* (a), *exponential* (b), dan *spherical* (c).

SIMPULAN

Berdasarkan hasil analisis didapatkan nilai moran's I sebesar 0.5434 yang menunjukkan bahwa terdapat autokorelasi spasial pada persentase jumlah penduduk miskin di Kabupaten/Kota di Jawa Timur. Sementara itu, pada hasil prediksi dan interpolasi, model *Gaussian* adalah model yang memberikan prediksi yang lebih baik dibandingkan model *exponential*, dan *spherical*.

DAFTAR PUSTAKA

- Anselin, L. dan Rey, S.J. (2010). *Perspectives on Spatial Data Analysis*. New York: Springer.
- Badan Pusat Statistik Jawa Timur. (2012). *Profil Kemiskinan di Jawa Timur September 2011*. Surabaya: BPS Provinsi Jawa Timur
- Bekti, R. D. dan Sutikno. (2011). Spatial Modeling on the Relationship between Asset Society and Poverty in East Java. *Jurnal Matematika dan Sains*, 16:140-146.
- Bekti, R.D. (2011). Autokorelasi Spasial Untuk Identifikasi Pola Hubungan Kemiskinan di Jawa Timur. *Jurnal Comtech*,3:217-227.
- Bohling, G. (2005). *Introduction to Geostatistics and Variogram Analysis*. Diakses 10 November 2010 dari <http://people.ku.edu/~gbohling/cpe940>.
- Fischer, M. M. dan Getis, A. (2010) *Handbook of Applied Spatial Snalysis: Software Tools, Method, and Applications*. New York:Springer.
- Kam, S. P, Hossain, M, Bose M. L, Villano, L. S. (2005). Spatial Patterns Of Rural Poverty and Their Relationship with Welfare-Influencing Factors In Bangladesh. *Food Policy*, 30:551-567.
- Kissling, W. D. dan Carl, G. (2008). Spatial Autocorrelation And The Selection Of Simultaneous Autoregressive Models. *Global Ecology and Biogeography*, 17: 59–71.
- Lee, J. dan Wong, D. W. S. (2001). *Statistical Analysis with Arcview GIS*. New York: John Wiley and Sons.