S - 4

Bayesian Conditional Autoregressive (CAR) Dalam Menaksir Resiko Relative Diare di Kota Bandung

Neneng Sunengsih[1],I Gede Nyoman Mindra Jaya[2], Zulhanif[3],Bertho Tantular[4]
[1,2,3,4]Departemen Statistika FMIPA UNPAD
Email: nenks.stat@gmail.com

Penyakit diare merupakan salah satu dari sepuluh kasus paling mengkawatirkan di Indonesia. Pada Tahun 2011 penyakit ini pernah menduduki peringkat pertama dengan jumlah kasus meninggal mencapai 1.289 kasus. Kota Bandung sebagai salah satu kota tersibuk di Jawa Barat dan pernah disoroti karena lingkungannya yang kurang bersih saat ini sedang berjuang mengurangi angka kasus diare. Angka prevelansi menurut data Tahun 2015 mencapai 2.11 sedikit lebih rendah dari angka kasus Jawa Barat dengan angka 2.21. Berbagai upaya serius harus dilakukan oleh dinas kesehatan untuk mengontrol penyebaran penyakit diare ini. Salah satunya adalah dengan lebih mengedepankan upaya *early warning system* melalui deteksi dini daerah daerah yang memiliki resiko tinggi penyebaran penyakit diare. Salah satu peran akademisi adalah membantu pemerintah membangun model *early warning system* melalui perhitungan angka resiko relative diare di masing-masing kecamatan di Kota Bandung. Berbagai metode telah dikembangkan untuk menghitung angka resiko relative diantaranya adalah *Standardized Morbidity/Mortality Ratio* (SMR). Metode ini dikembangkan dari model regresi Poisson. Namun metode ini gagal dalam mengatasi adanya masalah overdispersi. Sehingga diperlukan pendekatan yang lebih valid dalam menaksir angka resiko relative diare di Kota Bandung. Penulis mengusulkan metode *Bayesian Conditional Autoregressive* (CAR) Model. Metode ini mampu mengatasi permasalahan overdisversi melalui pemilihan parameter prior yang tepat.

Kata kunci: Bayesian, CAR Model, Resiko Relative

I. PENDAHULUAN

Penyakit diare merupakan salah satu dari sepuluh kasus paling mengkawatirkan di Indonesia. Penyakit ini tercatat sebagai salah satu penyakit endemis di Indonesia dan juga merupakan penyakit potensial KLB yang sering disertai dengan kematian. Menurut hasil Riskesdas 2007, diare merupakan penyebab kematian nomor satu pada bayi (31,4%) dan pada balita (25,2%), sedangkan pada golongan semua umur merupakan penyebab kematian yang ke-empat (13,2%). Pada tahun 2012 angka kesakitan diare pada semua umur sebesar 214 per 1.000 penduduk dan angka kesakitan diare pada balita 900 per 1.000 penduduk (Kajian Morbiditas Diare 2012).[1]

Pada Tahun 2011 penyakit ini pernah menduduki peringkat pertama dengan jumlah kasus meninggal mencapai 1.289 kasus. Kota Bandung sebagai salah satu kota tersibuk di Jawa Barat dan pernah disoroti karena lingkungannya yang kurang bersih saat ini sedang berjuang mengurangi angka kasus diare. Angka prevelansi menurut data Tahun 2015 mencapai 2.11 sedikit lebih rendah dari angka kasus Jawa Barat dengan angka 2.21. Berbagai upaya serius harus dilakukan oleh dinas kesehatan untuk mengotrol penyebaran penyakit diare ini. Salah satunya adalah dengan lebih mengedepankan uapaya *early warning system* melalui deteksi dini di daerah - daerah yang memiliki resiko tinggi penyebaran penyakit diare. Salah satu peran akademisi adalah membantu pemerintah membangun model *early warning system* melalui perhitungan angka resiko relative diare di masing-masing kecamatan di Kota Bandung. Berbagai metode telah dikembangkan untuk menghitung angka resiko relative diantaranya adalah *Standardized Morbidity/Mortality Ratio* (SMR) [2]. Metode ini dikembangkan dari model regresi Poisson. Namun metode ini gagal dalam mengatasi adanya masalah overdispersi. Sehingga diperlukan pendekatan yang lebih valid dalam menaksir angka resiko relative diare di Kota Bandung. Penulis mengusulkan metode Bayesian *Conditional Autoregressive* (CAR) Model [3]. Metode ini mampu mengatasi permasalahan overdisversi melalui pemilihan parameter prior yang tepat.

Terdapat beberapa jenis CAR prior yaitu model CAR Besag-York-Mollie, dimana model ini mengasumsikan bahwa ada struktur dependensi yang sangat kuat $\rho = 1$, CAR LEROUX yaitu model yang memandang struktur dependensi $0 < \rho < 1$ dan CAR LOCAL dengan mengasumsikan adanya struktur kluster dalam data [4]

Terdapat delapan model yang diusulkan dalam penelitian ini yaitu model 1) model dengan prior independen tanpa kovariat, 2) model dengan prior independen dengan kovariat, 3) model dengan prior BYM tanpa kovariat, 4) model dengan prior BYM dengan kovariat, 5) model dengan prior LEROUX tanpa kovariat, 6) model dengan prior LEROUX dengan kovariat, 1) model dengan prior LOCAL tanpa kovariat, dan 8) model dengan prior LOCAL dengan kovariat. Delapan model ini akan diperbandingkan untuk mencari model terbaik dalam memodelkan penyakit diare di Kota Bandung.

II. METODE PENELITIAN

Penelitian ini dilakukan untuk memetakan resiko penyakit diare di Kota Bandung. Data yang digunakan adalah data diare tahun 2015 yang diperoleh dari dinas kesehatan kota Bandung. Terdapat dua variabel penelitian yang digunakan seperti yang disajikan pada Tabel 1 berikut:

TABEL 1. VARIABEL PENELITIAN

Variabel	Satuan	Min	Maksimum	Rata-Rata
Angka Kasus Diare	Orang	747	4231	1916
Persentase Hidup Bersih dan Sehat (PHBS)	Persentase (%)	58.00	92.10	107.00
Jumlah Penduduk	Orang	25856	85849	153260

A. Data Spatial Dalam Kajian Kesehatan

Spatial Epidemiolgy dalam kajian kesehatan menyatakan metode yang mengkaji mengenai pola spatial dari kejadian penyakit dan tingkat kematian. Aspek yang menjadi fokus perhatian dalam *spatial epidemiology* adalah faktor-faktor potensial yang berpengaruh pada kejadian penyakit di suatu lokasi [5]

B. Autokorelasi Spasial

Landasan dalam analisis spasial adalah adanya autokorelasi spasial. Variabel pada lokasi yang berdekatan diyakini memiliki nilai yang relatif lebih mirip dibandingkan dengan lokasi yang berjauhan. Hukum Tobler menyatakan "Everything is related to everything else, but near things are more related than distant things'' [6]

Spatial autocorrelation index mengukur asosiasi spatial yang secara simultan mempertimbangkan informasi lokasi dan atribut. Terdapat dua tipe indeks yaitu ukuran global dan lokal. Ukuran global menyatakan ringkasan dari asosiasi spasial untuk keseluruhan wilayah. Sedangkan ukuran lokal menyatakan hubungan satu lokasi dengan lokasi tentangganya.

Spatial autocorrelation positif menunjukkan bahwa nilai pada satu lokasi tertentu mirip dengan lokasi tetangganya. Sedangkan *spatial autocorrelation* negatif menunjukkan nilai pada lokasi tertentu berbeda dan bahkan sangat berbeda dengan lokasi tetanggnya.

Statistik yang umumnya digunakan dalam pengukuran autokorelasi spasial global adalah Statistik Moran's I. Ukuran moran I dapat dituliskan sebagai berikut [7]

$$\mathrm{I} = \frac{\mathrm{n}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_{ij}}} \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_{ij} z_i z_j}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{z_i^2}} \tag{1}$$

Dengan $\mathrm{z_i} = (\mathrm{x_i} - \bar{\mathrm{x}})^2$ dimana x menyatakan variabel pengamatan, dan w_{ij} adalah ukuran kedekatan antara data pada lokasi ke-i dan ke-j. Untuk nilai w_{ij} yang distandarkan menurut baris maka $\sum_{\mathrm{i,j}} \mathrm{w_{ij}} = n$. Jika nilai I positif menunjukkan adanya autokorelasi positif dan negative jika sebaliknya.

TABEL 2. INDEKS MORAN'S DIARE

```
        Moran I test under randomisation

data:  Kasus
weights: BANDUNG

Moran I statistic standard deviate = 1.1603, p-value = 0.123
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
       0.09718359       -0.03448276        0.01287744
```

Perhitungan indeks Moran's mengasilkan nilai 0.097 dengan p.value = 0.123 lebih besar dari 0.05 sehingga disimpulkan tidak adanya spatial dependensi. Namun demikian, strutkur spatial dependensi yang diuji adalah strutkur global. Sehingga nilai ini bisa signifikan untuk pendekatan struktur lokal.

C. *Resiko Relatif*

Risiko Relatif θ_i merupakan ukuran yang digunakan dalam pemetaan penyakit untuk mengidentifikasi lokasi yang berisiko tinggi terjangkit penyakit. Ukuran Resiko relatif yang paling umum digunakan dalam pemetaan penyakit adalah Standadized Morbility/Mobidity Ratio (SMR) [8]

Standardized Morbidity Ratio (SMR) merpakan rasio antara kejadian / kasus pada lokasi ke-i (y_i) yang diasumsikan berdistribusi Poisson dengan nilai harapan terjadinya kasus pada lokasi ke-i (E_i). Nilai harapan terjadi umumnya diperoleh dari hasil perkalian jumlah populasi yang beresiko dengan peluang terjadinya satu kasus pada keseluruhan lokasi [9]

SMR digunakan dalam studi epidemiologi untuk mengamati lokasi dengan jumlah kasus lebih tinggi dibandingkan nilai harapannya sehingga dapat memberikan informasi mengenai sebaran geografis suatu penyakit.

$$\mathrm{SMR} = \frac{\mathrm{y_i}}{\mathrm{E_i}} \tag{2}$$

SMR tidak tepat digunakan untuk *small area* atau lokasi dengan jumlah kasus sangat sedikit dan jumlah populasi yang berisiko sangat kecil. Penerapan SMR pada *small area* akan memberikan varians taksiran yang semakin besar dengan semakin kecilnya jumlah kasus dan populasi pada lokasi ke-*i*. Sehingga taksiran resiko relatif yang didasarkan pada formulasi SMR dinyatakan kurang handal. Selain itu penggunaan SMR dalam menaksir resiko relatif tidak memungkinkan memasukkan informasi ketergantungan spasial.

D. *Bayesian Conditional Autoregressive Model (CAR)*

Kelemahan SMR dalam penaksiran resiko relatif dapat diatasi dengan membuat model pemulusan dan memasukkan informasi spasial ke dalam model. Karena ketergantungan spasial umumnya terjadi pada kasus penyakit menular seperti DBD. *Bayesian Conditional Autoregressive* (CAR) merupakan teknik dalam pemetaan penyakit yang memodelkan resiko relatif dengan memperhatikan pemulusan nilai taksiran resiko relatif dan memasukkan informasi spatial untuk mengurangi kekeliruan dari taksiran parameter resiko relatif sehingga diperoleh taksiaran resiko relatif yang lebih handal [10]. Istilah Bayesian merujuk pada konsep model pemulusan sedangkan *Conditional Autoregressive* merujuk pada model yang memungkinkan memasukkan informasi spasial dalam pemodelan. *Conditional Autoregresive Model* pertama kali diperkenalkan oleh [11].

Misalkan variabel acak $\mathrm{y} = \{\mathrm{y_1}, .., \mathrm{y_m}\}$ merupakan *vector univariate* yang menyatakan jumlah kasus pada lokasi ke-i (*i =1, 2,..,m*). Maka model umum CAR adalah:

$$\mathrm{y_i}|\mu_\mathrm{i} \sim f(\mathrm{y_i}|\mu_\mathrm{i}, \mathrm{v}^2)$$
$$g(\mu_\mathrm{i}) = \mathrm{x_i^T}\beta + \phi_\mathrm{i} + \mathrm{E_i} \tag{3}$$

$g(.)$ adalah *link function* untuk kasus *Generalized Linear Model*, *x* adalah covariate, β adalah vektor parmeter regresi dan E adalah nilai *expected Rate*.

Pada kasus penyakit menular, diasumsikan bahwa distribusi peluang bersyarat Y_i bergantung pada nilai dari observasi selain dari lokasi ke-*i* dengan peluang $\mathrm{P}(\mathrm{y_i}|\mathbf{y}_{-\mathrm{i}})$ dimana $\boldsymbol{Y}_{-i}$menyatakan vektor semua

observasi selain observasi pada lokasi ke-*i* [5]. Pada penelitian ini mengasumsikan bahwa distribusi peluang bersyaratnya adalah Gaussian sehingga model *Conditional autoregression* (CAR) untuk menaksir resiko relatif adalah [11]

$$E(\phi_i|\phi_j, j \in S_i) = \mu_i + \sum_{j \in S_i} C_{ij} h_i(\phi_j - \mu_i), \quad i = 1, \dots, m \tag{4}$$

dengan S_i menyatakan himpunan lokasi yang bersinggungan dengna lokasi ke-i, $\beta_i = \ln\theta_i$, dan μ_i menunjukkan rata-ratan resiko relatif pada lokasi ke-i dengan varians $V(y_i|\mathbf{y}_{-i}) = \tau^{-2}$.

Vektor μ_i dapat dinyatakan dalam model regresi $\mathbf{X\alpha}$ dengan $\mathbf{X}$ adalah matrik covariate dan $\boldsymbol{\alpha}$ adalah vektor koefisien regresi. Kovariat dalam penelitian ini adalah Angka Bebas Jentik, Kepadatan Penduduk dan Rumah Sehat.

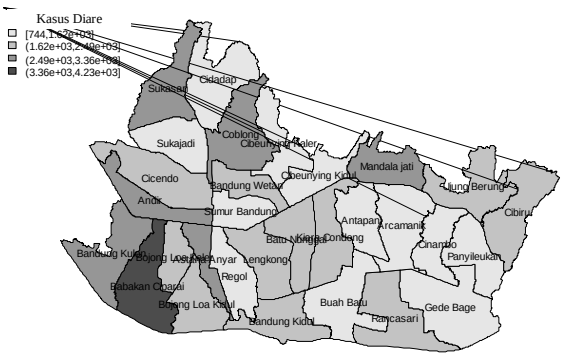
Untuk mensolusikan persamaan di atas digunakan pendekatan model *Hierarchical Bayes* (HB) dengan penaksiran parameter model menggunakan pendakatan *Markov Chain Monte Carlo* (MCMC).

Model yang akan diuji dengan struktur pior sebagai berikut:

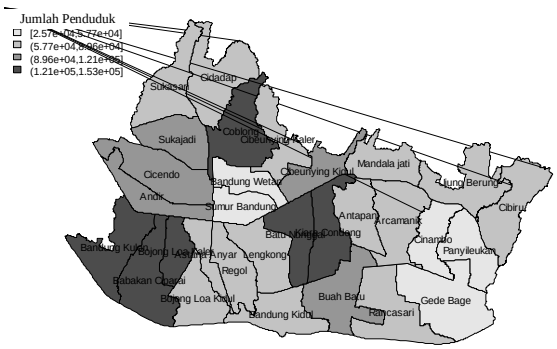
TABEL 3. MODEL PRIOR

Metode	Prior
BYM	$\phi_i\|\phi_{-i} \sim N\left(\frac{\sum_{j=1}^{m} w_{ij}\phi_i}{\sum_{j=1}^{m} w_{ij}}, \frac{\tau^2}{\sum_{j=1}^{m} w_{ij}},\right)$
LEROUX	$\phi_i\|\phi_{-i} \sim N\left(\frac{\rho\sum_{j=1}^{m} w_{ij}\phi_i}{\rho\sum_{j=1}^{m} w_{ij} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{j=1}^{m} w_{ij} + 1 - \rho},\right)$
LOCAL	$COR(\phi_i, \phi_k\|\phi_{-ik}) = \frac{\rho w_{ij}}{\sqrt{(\rho\sum_{j=1}^{m} w_{ij} + 1 - \rho)(\rho\sum_{j=1}^{m} w_{kj} + 1 - \rho}}$

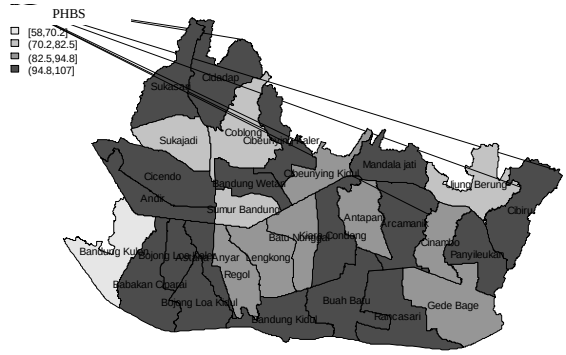
Pola ditribusi variabel penelitian beserta nilai *expected rate* (E) dan nilai *Standardized Morbidity Ratio* (SMR) disajikan di bawah ini:



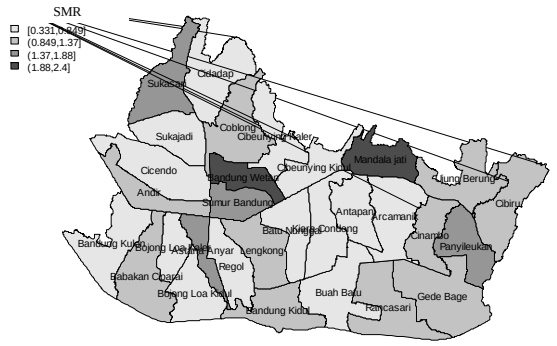
A. KASUS DIARE



B. POPULASI



C. PHBS



D. SMR

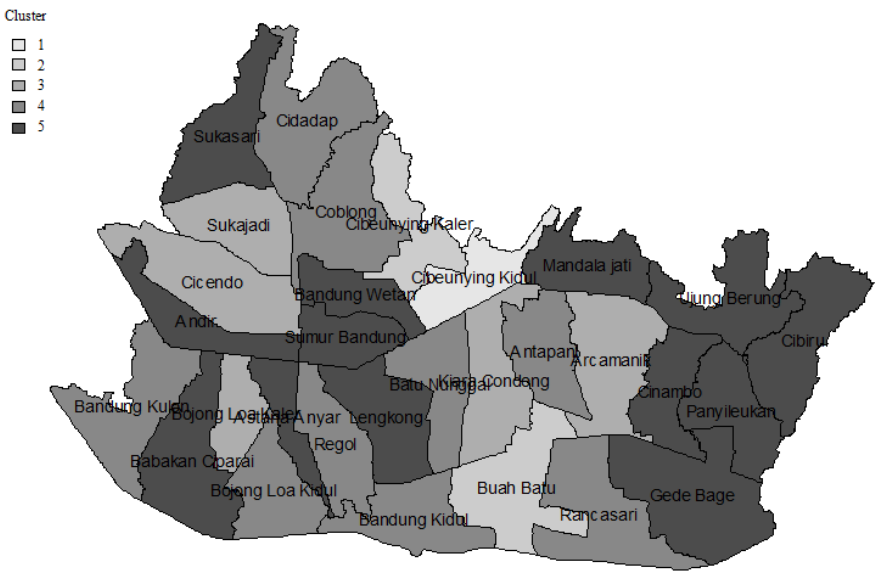
GAMBAR 1. STATISTIK DATA PENELITIAN

Terlihat ada pola yang cukup mirip antara angka kasus diare dengan jumlah penduduk dimana pada umumnya angka kasus tinggi untuk jumlah penduduk yang tinggi dan angka kasus relatif rendah untuk jumlah penduduk yang kecil. Namun tidak begitu dengan angka PHBS, hanya kecamatan Babakan Ciparay yang memiliki pola yang sama. Jika dibandingkan dari nilai SMR, tidak menunukkan bahwa angka kasus tinggi selaras dengan SMR yang tinggi.

TABEL 4. VARIABEL PENELITIAN

Metode	Model	DIC	R^2
INDEPENDENT	$\hat{\eta}_i = 0.015 + \log(E_i)$	338.6428	0.9969422
INDEPENDENT-X	$\hat{\eta}_i = -0.2192 + 0.0024PHBS + \log(E_i) + v_i$	338.9849	0.9976073
BYM	$\hat{\eta}_i = 0.015 + \log(E_i) + v_i$	339.2018	0.9973311
BYM-X	$\hat{\eta}_i = -1.0999 + 0.0120PHBS + \log(E_i) + v_i$	338.5093	0.9966633
LEROUX	$\hat{\eta}_i = 0.015 + \log(E_i) + v_i$	338.4259	0.9977621
LEROUX-X	$\hat{\eta}_i = 0.0012 + 0.0001PHBS + \log(E_i) + v_i$	338.3796	0.9973868
LOCAL	-	337.8243	0.9991088
LOCAL-X	$\hat{\eta}_i = -0.0042\text{PHBS} + \log(E_i) + v_i$	336.6985	0.9983505

Penelitian ini menganalisis delapan model yang dibedakan menurut prior dan variabel kovariat dalam model. Model terbaik menurut DIC yang minimum dan nilai R^2 yang besar adalah model local dengan melibatkan variabel kovariat PHBS. Nilai DIC model Local-X sebesar 336.6985 dan nilai R^2 sebear 0.99983. Model ini tidak melibatkan *intercept* dengan *slope* koefisien regresi PHBS adalah sebesar -0.0042. Ini artinya peningkatan 10% nilai PHBS akan menurunkan resiko diare sebesar 0.042 satuan.



GAMBAR 2. POLA RESIKO RELATIF MENURUT MODEL LOCAL-X

Hasil pengolahan data menunjukkan bahwa kecamatan di bagian timur meliputi Mandala Jati, Ujung Berung, Cibiru, Cinambo, Penyileukan, dan Gede Bage adalah kecamatan dengan resiko relative diare yang tinggi masuk dalam kluster lima. Sedangkan beberapa kecamatan berada di pusat dan utara.

TABEL 1. PROFILE CLUSTER

Cluster	Min	Max	Rata-Rata
1	0.335762	0.335762	0.335762
2	0.499539	0.511832	0.505686
3	0.641464	0.831669	0.702443
4	0.764295	1.287772	0.896372
5	1.237115	2.380186	1.549367

Kluster lima adalah kluster dengan angka resiko relative yang tinggi dibandingkan cluster yang lain, nilai resiko relative paling rendah di kluster ini adalah 1.24 dan paling tinggi adalah 2.38 dengan rata-rata resiko relative mencapai 1.59. Angka ini menyatakan bahwa kejadian diare di kecamatan yang masuk kluster lima, 50% lebih tinggi dibandingkan dengan yang diharapkan. Sedangkan kluster dengan angka resiko relative paling rendah adalah kluster satu yang meliputi hanya satu kecamatan yaitu Buah Batu.

III. SIMPULAN DAN SARAN

Model yang terbaik utuk menaksir resiko relative penyakit diare di Kota Bandung adala model CAR dengan prior Local dan melibatkan PHBS sebagai variabel yang turut berkontribusi pada tinggi rendahnya resiko relative pada kecamatan-kecamatan di Kota Bandung.

UCAPAN TERIMA KASIH

Terimaksih kepada Rektor Universitas Padadjaran atas HIBAH Riset Kompetensi Dasar melalui DRPM Unpad.

DAFTAR PUSTAKA

[1] Yudianto, et al (2014). Profil Kesehatan Indonesia Tahun 2014. Kementerian Kesehatan Republik Indonesia
[2] Clayton, D., & Kaldor, J. (1987). Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. Biometrics , 671-681
[3] Mindra Jaya, IGDN et.al (2014) Conditional Autoregressive Model on Dengue Fever Disease Mapping In Bandung City. Conference paper. International Conference on Statistics and Mathematics 2014
[4] Lee, D (2014) CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. Journal of Statistical Software Volume 55, Issue 13
[5] Waller, L. A., & Gotway, C. A. (2004). Applied Spatial Statistics for Health Data. John Willey & Son.
[6] Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography , 46 (2), 234-240.
[7] Manfred, D., Fischer, & Wang, J. (2011). Spatial Data Analysis, Models Methods and Techniques. New York: Springer.
[8] Pringle, D. (1996). Mapping Disease Risk Estimate Based on Small Numbers : An Assesment of Empirical Bayes Techniques. The Economic and Social Review , 27 (4), 341-363.
[9] Wakefield, J. (2006). Disease mapping and spatial regression with count data. Biostatistics , 8 (2), 158-183.
[10] Venkatesan, P., Srinivasan, R., & Dharuman. (2012). Bayesian Conditional Auto Regressive Model For Mapping Tuberculosis Prevalence In India. International Journal of Pharmaceutical Studies and Research , 3 (1), 1-3.
[11] Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Latice System. Journal Royal Statistics , 36, 192-236.
[11] Maiti, T. (1998). Hierarchical Bayes estimation of mortality rates for Disease Mapping. Journal of Statistical Planning and Inference , 69, 339-348.