

Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA  
Fakultas MIPA, Universitas Negeri Yogyakarta, 16 Mei 2009

## MEMPREDIKSI CURAH HUJAN (DATA SPATIO-TEMPORAL) DENGAN METODE BAYESIAN NETWORKS

Dewi Retno Sari Saputro

Jurusan Matematika, FMIPA UNS,  
Jl. Ir. Sutami 36A Solo Telp./Fax (0271) 663375  
E-mail: dewi\_rss@yahoo.com

### Abstrak

Curah hujan merupakan salah satu parameter yang berkaitan dengan dampak perubahan iklim. Oleh karena itu, dalam penelitian klimatologi dan meteorologi, parameter tersebut sering dilibatkan. Di Indonesia yang berada di kawasan tropis, simulasi curah hujan merupakan proses yang sukar untuk disimulasikan. Bahkan hingga saat ini, belum ada suatu model iklim yang mampu mensimulasikan curah hujan di Indonesia dengan baik. Beberapa penelitian yang telah dilakukan belum mampu memberikan hasil yang cukup memadai. Namun begitu, beberapa metode telah dikembangkan untuk menduga curah hujan. Salah satu metode yang sering digunakan adalah ARIMA (*Auto Regressive Integrated Moving Average*) untuk data musiman. Kelemahannya, metode tersebut hanya dapat dipergunakan pada pendugaan curah hujan berdasarkan data curah hujan yang ada di stasiun penakar hujan (*temporal*) tanpa mempertimbangkan lokasi stasiunnya (*spatial*). Dengan demikian, diperlukan suatu metode yang mempertimbangkan keduanya, salah satu metode yang dimaksudkan adalah *Bayesian Networks (BNs)*. Struktur *BNs* dibangun berdasarkan model peluang ketergantungan antar stasiun yang dinyatakan dalam *directed acyclic graph (DAC)* dan untuk mengoptimalkan peluangnya dipergunakan algoritma K2. Dalam paper ini, disampaikan tentang *BNs* dalam pendugaan curah hujan, kelemahan dan kelebihanannya.

Metode yang dipakai dalam penelitian ini yakni deduksi dan induksi, dengan melakukan penelusuran terhadap materi tentang iklim dan metode prediksi curah hujan, utamanya tentang *BNs*. Penelusuran dilakukan terhadap penelitian-penelitian terdahulu serta pada beberapa *textbook* dan jurnal ilmiah serta beberapa diskusi ilmiah. Selanjutnya, mengkaji dan *me-review* beberapa teoritisnya yang dikaitkan dengan pemodelan curah hujan.

Berdasarkan kajian ini, dapat dinyatakan bahwa *BNs* merupakan salah satu metode yang dapat dipertimbangkan untuk memprediksi data curah hujan karena dalam metode ini, *BNs* memperhitungkan nilai peluang untuk setiap stasiun penakar hujan. Namun, *BNs* lemah, saat menentukan urutan stasiun dalam membentuk suatu jaringan Bayes.

**Kata kunci** : curah hujan, *spatio-temporal*, *Bayesian Network (BNs)*, *directed acyclic graph (DAC)*, algoritma K2

### PENDAHULUAN

Terjadinya perubahan iklim akibat naiknya temperatur di bumi yang diprediksi mencapai 1<sup>0</sup>C sampai dengan 3<sup>0</sup>C (pemanasan global/*global warming*) pada abad ini berpotensi mengubah pola cuaca secara ekstrem.<sup>1</sup> Naiknya temperatur tersebut disebabkan oleh meningkatnya konsentrasi gas rumah kaca di atmosfer sebagai hasil aktivitas manusia. Indonesia merupakan negara agraris dengan tingkat produktivitas yang rawan mengalami gangguan akibat dampak pemanasan global.

Pengaruh pemanasan global telah menyebabkan perubahan iklim, antara lain terlihat dari curah hujan di bawah normal, sehingga masa tanam terganggu, dan meningkatnya curah hujan di sebagian wilayah. Indonesia merupakan negara kepulauan dengan bentuk topografi yang sangat beragam, dilewati oleh garis katulistiwa, diapit oleh dua benua dan dua samudera. Posisi ini menjadikan Indonesia sebagai daerah pertemuan sirkulasi meridional (Utara-Selatan) yang dikenal sebagai sirkulasi *Hadley* dan sirkulasi zonal (Timur-Barat) yang dikenal sebagai sirkulasi *Walker*,

<sup>1</sup> Kompas Sabtu 1 Desember 2007, Fokus : Mencermati Adaptasi

dua sirkulasi yang sangat mempengaruhi keragaman iklim Indonesia. Posisi semu matahari yang berpindah dari 23.5<sup>0</sup> Lintang Utara ke 23.5<sup>0</sup> Lintang Selatan sepanjang tahun dan gangguan siklon tropis juga ikut mempengaruhi keragaman iklim Indonesia. Semua aktivitas dan sistem ini berlangsung secara bersamaan sepanjang tahun dan menyebabkan sistem golongan lokal cukup dominan. Oleh karena itu, Indonesia sebagai negara tropis memiliki pola hujan dengan variasi yang besar dibandingkan unsur cuaca dan iklim yang lain serta memiliki curah hujan yang berbeda antara satu tempat dengan tempat lainnya. Dan peubah penting mengenai dampak perubahan iklim yang sering digunakan adalah curah hujan.

Curah hujan mempunyai keragaman yang besar, baik dalam skala ruang maupun dalam skala waktu (Sekarjaya 1986). Menurut Bruce & Clark (1966) dan Seyhan (1993), keragaman dalam skala ruang di samping ditentukan oleh keberadaan uap air juga ditentukan oleh garis lintang, ketinggian tempat, jarak dari sumber air, posisi di dalam dan ukuran massa tanah atau benua, arah angin, hubungan dengan deret pegunungan, suhu nisbi tanah dan samudra yang berbatasan. Keragaman curah hujan terjadi juga secara lokal di suatu tempat yang menyebabkan penyebaran hujan tidak merata. Keragaman dalam skala waktu dapat dipandang dalam hubungannya dengan batasan waktu terjadinya hujan yaitu tahunan, musiman atau jangka pendek, maupun hubungannya dengan peluang statistik seperti nilai ekstrim, frekuensi hujan.

Secara umum, ada tiga pendekatan yang dapat dilakukan untuk mengurangi resiko iklim atau mengurangi dampak kejadian iklim ekstrem, yakni pendekatan strategis, taktis dan operasional (Boer (2006) dalam Peragi dan Perhimpri (1994)). Dalam paper ini akan dieksplorasi tentang pendekatan taktisnya, yakni upaya untuk meningkatkan ketepatan ramalan iklim/musim di suatu daerah dalam bentuk pemodelan iklim.

Proses pembentukan hujan di kawasan tropis merupakan proses yang paling sukar disimulasikan (Ratag dalam Rahmat 2008). Hingga saat ini, belum ada suatu model iklim yang mampu mensimulasikan curah hujan di Indonesia dengan baik dan berdasarkan beberapa penelitian yang dilakukan belum memberikan hasil yang memuaskan (Kustiyo et al dalam Rahmat 2008). Beberapa metode telah dikembangkan untuk memprediksi curah hujan. Salah satu metode yang sering digunakan adalah ARIMA (*Auto Regressive Integrated Moving Average*) untuk data musiman. Namun, metode tersebut hanya digunakan pada pendugaan curah hujan berdasarkan data curah hujan yang ada di stasiun penakar hujan (*temporal*) tanpa memperhatikan pengaruh lokasi stasiunnya (*spatial*). Oleh karena itu, diperlukan suatu metode untuk memprediksi curah hujan yang mempertimbangkan pengaruh data curah hujan pada stasiun dan lokasi stasiunnya (*spatio-temporal*). Dan metode *Bayesian Networks* yang ideal untuk menyelesaikan masalah ini (Chan Li, Montsera F, & Jerry MD 2005). Metode *BNs* merupakan metode yang tidak hanya melibatkan pengaruh *temporal* dalam pendugaan curah hujan, namun juga menunjukkan adanya pengaruh *spatial* antara stasiun pengamatan curah hujan. Dalam paper ini diuraikan bagaimana *BNs* menjadi suatu metode yang praktis dan efisien untuk menduga curah hujan.

## PEMBAHASAN

Lebih dari 10 tahun, *BNs* menjadi alat yang populer untuk memodelkan permasalahan-permasalahan statistik (Mittal, Asraf & Tele 2007). Menurut Pearl (1988), *BNs* merupakan model grafik dari sebaran peluang. Dalam kerangka ini, *BNs* merupakan model *spatial-temporal* dimana terdapat ketergantungan dan ketidaktergantungan antara stasiun penakar hujan satu dengan lainnya dalam bentuk grafik (*directed acyclic graph/DAG*) yang mudah dimengerti dan diinterpretasikan (Cofino et al. 2002). Bahkan, proses perhitungannya pun sangat efisien (Mittal et al 2007). Dalam literatur yang lain, dikatakan bahwa *BNs* dapat memberikan informasi peluang dengan sederhana dan lengkap serta merupakan teknik yang populer dalam menyelesaikan suatu masalah jika ketidakpastiannya sangat kompleks (Mittal et al 2007). Menurut Cooper & Herskovits (1992); Cano, et al (2004), berdasarkan komponennya, *BNs* terdiri dari *Bayesian Structure (structure learning)* dan *Bayesian Parameter (parametric learning)*.

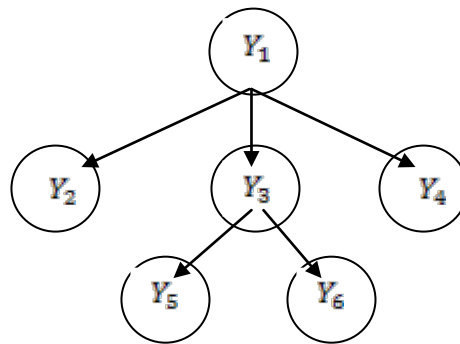
Sementara itu, menurut Cofino et al (2002) dan Cano et al (2004), komponen dari *BNs* tersebut yang disebut sebagai penilaian dan pencarian terdiri dari

1. ukuran kualitas, untuk menentukan kualitas struktur grafik dan pendugaan parameter bagi kandidat *BNs*
2. algoritma pencarian, untuk pencarian secara efisien ruang *BNs* yang mungkin dan menemukan satu kualitas yang terbaik.

Oleh Cano et al (2004) dipergunakan algoritma yang disebut sebagai algoritma K2, hingga saat ini algoritma tersebut belum ada yang menyempurnakan dari sisi kompleksitas algoritmanya. Tentang algoritma ini akan diuraikan pada submateri berikutnya.

Mekanisme teknik ini dengan menggunakan grafik, sebagai spesifikasi sebaran peluang bersama dari suatu himpunan peubah acak (model stokastik). Dasar *BNs* adalah teori probabilitas (Baclawski). Adapun komponen dari *BNs* (Pollino 2006) adalah himpunan peubah acak dan himpunan arah yang terhubung di *node*.

Peubah acak dari model stokastik direpresentasikan dengan *node* dari suatu graf yakni *Directed Acyclic Graph (DAG)* yang menggambarkan ketergantungan antar setiap peubah random. *Bayesian parameter* merupakan himpunan parameter dari sebaran peluang bersyarat setiap peubah berdasarkan graf tersebut. *Bayesian structure* terdiri dari *node* yang merepresentasikan peubah-peubah dan *edge* yang merepresentasikan hubungan ketergantungan antar *node* seperti pada Gambar 1. Setiap *node* yang dihubungkan secara langsung, menunjukkan hubungan ketergantungan



Gambar 1 Direct Acyclic Graph

Misalkan hubungan dari *node* dinyatakan dengan  $\{Y_1, \dots, Y_n\}$ , jika terdapat *edge* dari *node*  $Y_j$  ke *node*  $Y_k$ , dikatakan bahwa  $Y_j$  adalah *parent* dari  $Y_k$ , dan  $Y_k$  adalah *child* dari  $Y_j$ . Himpunan *parent* dari *node*  $Y_i$  dinotasikan dengan  $\Pi_i$ .

Struktur ketergantungan yang digambarkan dengan DAG tersebut dapat diinterpretasikan ke fungsi kepekatan bersama dari peubah-peubah dengan jalan mengalikan semua peluang berdasarkan parentnya.

$$P(y_1, \dots, y_n) = \prod_{i=1}^n P(y_i | \pi_i) \quad (1)$$

Berdasarkan persamaan tersebut dapat ditunjukkan bahwa DAG merupakan dekomposisi dari fungsi peluang berdimensi besar ke dalam sebaran lokal berdimensi rendah. Menurut Cano et al (2004), terdapat dua tipe penting dalam *BNs* yaitu *Multinomial BNs* untuk peubah diskrit dan *Gaussian BNs* untuk peubah kontinu. Dua tipe tersebut adalah

#### A. *Multinomial Bayesian Networks*

Dalam *multinomial BNs* diasumsikan bahwa semua peubah, diskrit, di mana setiap peubah memiliki himpunan nilai yang terbatas, dan bahwa peluang bersyarat untuk setiap peubah berdasarkan *parent*-nya menyebar multinomial.

#### B. *Gaussian Bayesian Networks (GBNs)*

Dalam *Gaussian BNs*, semua peubah diasumsikan menyebar Normal ganda, yaitu  $f(y) \sim N(\mu, \Sigma)$  dengan  $\mu$  rata-rata vektor dimensi  $n$ ,  $\Sigma$  adalah matriks kovariansi  $n \times n$ .

Fungsi kepekatan bersama

$$f(\mathbf{y}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2(\mathbf{y} - \mu)' \Sigma^{-1} (\mathbf{y} - \mu)} \right\}$$

Peluang bersyarat untuk setiap node berdasarkan *parent*-nya

$$f(\mathbf{y}_i | \pi_i) \sim N(\mu_i + \sum_{j=1}^{i-1} \beta_{ij} (\mathbf{y}_j - \mu_j), \mathbf{v}_i) \quad i = 1, 2, \dots, n; j = 1, 2, \dots, n-1$$

dengan  $\beta_{ij}$  adalah koefisien regresi dari  $\mathbf{y}_j$  dalam regresi  $\mathbf{y}_i$  pada *parent*  $\mathbf{y}_i$  dengan  $\pi_i$  dan  $\mathbf{v} = \Sigma_i - \Sigma_{i\pi_i} \Sigma_{\pi_i}^{-1} \Sigma_{\pi_i i}^{-1}$  adalah ragam bersyarat dari  $\mathbf{y}_i$ .  $\Sigma_{i\pi_i}$  merupakan kovariansi vektor di antara  $\mathbf{y}_i$  dan peubah dalam  $\pi_i$ . Jadi, *GBNs* diberikan oleh himpunan parameter  $\{\mu_1, \mu_2, \dots, \mu_n\}$ ,  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  dan  $\{\beta_{ij} | j < i\}$ .

### Pembentukan Struktur *BNs* dengan Algoritma K2

Struktur *BNs* ditentukan dengan memaksimalkan peluang. Karena fungsi dari banyaknya node (banyaknya struktur yang mungkin), jumlah struktur yang mungkin tumbuh secara eksponensial. Oleh karena itu, suatu perhitungan sempurna dari semua struktur *network* tidak *feasible* dalam banyak domain, sehingga oleh Robinson (1977) disusun perhitungan secara efisien fungsi recursif untuk menentukan banyaknya struktur jaringan yang terdiri dari  $n$  node:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i)$$

Berdasarkan rumusan tersebut, banyaknya struktur untuk  $n$  node dapat ditentukan, misalnya  $n = 2$ , banyaknya struktur yang mungkin ada 3;  $n = 5$ , banyaknya struktur yang mungkin ada 29.000; dan seterusnya.

Jika diasumsikan bahwa  $n$  peubah terurut, sedemikian sehingga jika  $\mathbf{y}_i$  mendahului  $\mathbf{y}_j$  dalam urutan, maka tidak dibolehkan adanya struktur (tanda panah) dari  $\mathbf{y}_j$  ke  $\mathbf{y}_i$ . Berdasarkan aturan

pengurutan tersebut, ada  $2^{\binom{n}{2}} = 2^{\frac{n(n-1)}{2}}$  kemungkinan struktur yang dapat dibentuk (Cooper & Herzkovits 1992), sehingga diperlukan algoritma yang memberikan struktur terbaik. Berdasarkan eksplorasi literatur yang ada, algoritma yang digunakan adalah algoritma K2 (Cano et al 2004).

Algoritma K2 merupakan algoritma pencarian greedy yang sederhana untuk menemukan *BNs* dengan peluang yang tinggi dalam waktu yang *reasonable* (Cooper & Herzkovits 1992, Cano et al 2004). Algoritma K2 diperlukan untuk menentukan struktur jaringan  $\mathbf{B}_s$  yang memaksimalkan  $P(\mathbf{B}_s, \mathbf{D})$  dengan mengasumsikan bahwa peubah tersebut terurut dan dilakukan dengan asumsi bahwa setiap *node* tidak memiliki *parent*. Selanjutnya, dilakukan penambahan *parent* dimana penambahan tersebut meningkatkan peluang dari hasil akhir struktur. Demikian seterusnya proses ini diulang hingga penambahan *parent* yang baru tidak meningkatkan nilai peluang atau *network* lengkap telah diperoleh. Jumlah maksimum *parent* untuk setiap node dapat juga dioptimumkan selama proses tersebut. Adapun fungsi yang menjadi acuan untuk meningkatkan nilai peluang dari strukturnya (Cooper & Herzkovits 1992) adalah

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)} \prod_{j=1}^{q_i} N_{ijk}!$$

Dan berikut merupakan prosedur untuk algoritma K2 (Cooper & Herzkovits 1992).

1. **Procedure** K2;
2. **For**  $i := 1$  **to**  $n$  **do**

3.  $\pi_i = \Phi$ ;
4.  $P_{old} = g(i, \pi_i)$ ;
5.  $OKToProceed := true$ ;
6. **While**  $OKToProceed$  and  $|\pi_i| \leq u$  **do**
7. *let*  $V$  *be the node in*  $Pred(y_i) - \pi_i$  *that maximizes*  $g(i, \pi_i \cup \{V\})$ ;
8.  $P_{new} = g(i, \pi_i \cup \{V\})$ ;
9. **If**  $P_{new} > P_{old}$  **then**
10.  $P_{old} = P_{new}$ ;
11.  $\pi_i := \pi_i \cup \{V\}$ ;
12. **else**  $OKToProceed := false$ ;
13. **End** {while};
14. **Write**("Node:", "parent of this nodes :",  $\pi_i$ );
15. **End** {for};
16. **End** {K2};

Lebih lanjut tentang kompleksitas dari algoritma K2 ini dapat dikaji pada paper Cooper & Herscovits (1992).

### Pemodelan Peluang Curah Hujan

Data curah hujan merupakan data yang berbasis *spatio-temporal* dan dengan *BNs* dapat diperoleh nilai peluang munculnya bulan basah, kering dan lembab untuk setiap stasiun. Peluang untuk setiap stasiun dapat berubah dengan dapat ditentukannya nilai peluang dari stasiun yang lain. Hal ini disebabkan karena adanya unsur ketergantungan antara stasiun satu dengan yang lain yang diinterpretasikan dengan nilai peluang bersyarat dalam jaringan.

### Implementasi *BNs*

Di antara yang telah melakukan implementasi *BNs* adalah Cofino (2002), *BNs* ini diimplementasikan pada 100 stasiun di North basin (Iberia) selama musim dingin tahun 1999 dan menghasilkan prediksi peluang cuaca yang efisien. Sekaligus dalam kajian tersebut ditunjukkan bahwa metode standar (*prediksi time series, neural network*) untuk prediksi cuaca merupakan kasus khusus dari *BNs* jika tidak ada ketergantungan secara spasial yang dipertimbangkan dalam model. Kajian dan implementasi lainnya, Rahmat (2008), *BNs* ini diimplementasikan pendugaan curah hujan pada 9 dari 31 stasiun penakar hujan di Indramayu, namun dia menggabungkannya dengan *ARIMA*, menjadi *BNARIMA* dan diperoleh simpangan baku yang lebih kecil daripada jika pendugaan tersebut digunakan model *ARIMA*. Hal ini dapat dipahami karena *ARIMA* berbasis *time series (temporal)* dan *BNARIMA* berbasis *spatio-temporal*.

Selain itu, Jauriyah (2008) mengkombinasi Hidden Markov dan *BNs*. Hasil pengujian terhadap sistem peramalan temperatur udara menunjukkan ketepatan sistem dengan hasil akurasi yang berbeda-beda sesuai dengan pembagian daerah dan waktu dari data latih.

## PENUTUP

### Simpulan

Berdasarkan kajian ini, dapat ditunjukkan bahwa *BNs* memiliki keunggulan dengan memperhitungkan nilai peluang setiap stasiun penakar hujan. Namun, juga memiliki kekurangan yakni dalam pembentukan jaringan Bayes harus memiliki urutan tertentu, dimana urutan ini akan berpengaruh terhadap pemodelan peluangnya.

### Saran

Perlu kajian dan eksplorasi lebih mendalam tentang inisialisasi pembentukan jaringan yakni urutan stasiun penakar, karena urutan ini tergantung pada banyak faktor dan akan berpengaruh terhadap hasil prediksi peluang.

### Rekomendasi

*BNs* sebagai salah satu metode untuk menyusun model yang berbasis spatio-temporal (=curah hujan) dapat dipertimbangkan untuk memodelkan secara peluang. Tidak hanya dari sisi efisiensi, namun juga lebih realistik.

### DAFTAR PUSTAKA

- Baclawski K. Bayesian Network Development, College of Computer Science, Northeastern University Boston, Massachusetts USA.
- Boer, R. 2006. Pendekatan Dalam Mengelola Resiko Iklim. Makalah dalam Pelatihan Dosen Bidang Pemodelan dan Simulasi Komputer untuk Pertanian. Cisarua Bogor, 7-20 September 2006.
- Baclawski, K. Bayesian Network Development, College of Computer Science, Northeastern University Boston, Massachusetts USA.
- Cano, R, Carmen S & Jose MG. 2004. Applications of Bayesian Networks in Meteorology. Springer.
- Chen L, Montserrat F & Jerry MD. 2005. Spatial-temporal statistical modeling and prediction of environmental processes.
- Cofino, AS et al. 2002. Bayesian Networks for Probabilistic Weather Prediction. Proceedings of the 15th European Conference on Artificial Intelligence, IOS Press, 695-700.
- Coles, S.G & Tawn, J.A. 1996. A Bayesian Analysis of Extreme Rainfall Data. Applied Statistics, 45, hal. 463-478.
- Cooper, GF, Herskovits F. 1992. A Bayesian Method for The Induction of Probabilistic Networks from Data. Machine Learning Journal 9:308-347.
- Daly, C. 2006. *Guidelines For assessing The Suitability of Spatial Climate Data Sets. International Journal of Climatology* 26: 707-721.
- Jauhariyah KF. 2008. Analisis Sistem Peramalan Temperatur Udara Menggunakan Kombinasi Hidden Markov dan Bayesian Network [skripsi]. Bandung : ITT TELKOM.
- Koesmaryono Y, Rizaldi Boer, Hidayat P, Yusmin dan Irsal Las. 1999. Pendekatan IPTEK dalam Mengantisipasi Penyimpangan Iklim. Di dalam : Prosiding Diskusi Panel Strategi Antisipatif Menghadapi Gejala Alam El-Nino dan La-Nina. PERHIMPI. Bogor. Hal 43-58.
- Mittal, A, Ashraf K, & Tele Tan. 2007. Bayesian Network Technologies: Applications and Graphical Models. IGI Publishing.
- Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publisher, Inc, San Mateo, California.
- Pollino, CA. 2006. Quantifying Ecological Risks using Bayesian Networks: Modelling in an Uncertain World. MRC Ecological Risk Assessment Training Course. [30 Maret 2009].
- Rahmat, H.F. 2008. Pendugaan Curah Hujan dengan Bayesian Networks (Studi Kasus Curah Hujan di Daerah Indramayu) [tesis]. Bogor : Institut Pertanian Bogor.
- Ruiz, C. Ilustration of K2 Algoritm for Learning Bayes Net Structures. Departement of Computer Science, WPI.
- Sahu, SK & Kanti VM. 2005. Recent Trends in Modeling Spatio-Temporal Data.