

Dariusz Biskup

Akademia Ekonomiczna we Wrocławiu

ANALIZA PRZESTRZENNO-CZASOWA POZIOMU ZANIECZYSZCZENIA ŚRODOWISKA

1. Wstęp

Poziom zanieczyszczenia środowiska jest zjawiskiem, które można analizować w wymiarze zarówno przestrzennym jak i czasowym. W wymiarze przestrzennym zakłada się na ogół, że wielkość korelacji poziomu zanieczyszczenia między punktami pomiarowymi jest zależna od odległości między nimi. W analizie czasowej korelacja może być przedstawiona jako funkcja różnicy momentów pomiaru.

Wiele zbiorów danych dotyczących zanieczyszczenia środowiska jest dostępnych w ujęciu przestrzenno-czasowym. Łączna analiza takich danych w obydwu tych ujęciach pozwala na znacznie skuteczniejsze przewidywanie poziomów zanieczyszczenia w porównaniu z analizą uwzględniającą tylko jeden z tych czynników.

W referacie przedstawione zostaną wyniki analizy z użyciem modelu przestrzenno-czasowego dla danych o zanieczyszczeniu zmierzonych na Dolnym Śląsku. Parametry opisywanego modelu będą szacowane metodami bayesowskimi z wykorzystaniem algorytmu Gibbsa.

2. Proces przestrzenno-czasowy

W referacie przedstawiony zostanie jeden ze sposobów przestrzenno-czasowej analizy danych. W opisaney metodzie [1] zakłada się, że poziom zanieczyszczenia modelowany będzie jako proces gaussowski $Y(s, t)$, gdzie s oznacza współrzędne punktu pomiarowego ($s \in R^2$), a t moment pomiaru.

Wartość oczekiwana procesu $Y(s, t)$ jest równa $\mu_s(\beta) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ ($s = (s_1, s_2)^T$, $\beta = (\beta_0, \beta_1, \beta_2)^T$). Z kolei funkcja kowariancji $\text{Cov}(Y(s, t), Y(s', t'))$ w opisywanym modelu zależy od odległości między lokalizacjami s oraz s' i od

różnicy czasu $|t - t'|$, co oznacza, że $\text{Cov}(Y(s, t), Y(s', t')) = c(\|s - s'\|, |t - t'|)$. Ponadto przyjęte jest założenie o jej separowalności, tzn.

$$\text{Cov}(Y(s, t), Y(s', t')) = \sigma^2 \rho^{(1)}(s - s'; \phi) \rho^{(2)}(t - t'; \psi), \quad (1)$$

gdzie σ^2 oznacza wariancję procesu, natomiast $\rho^{(1)}, \rho^{(2)}$ – funkcje korelacji zależne odpowiednio od parametrów ϕ oraz ψ .

Załóżmy, że dane pomiarowe zebrane zostały z I punktów pomiarowych w J okresach (dniach). Niech wektor $\mathbf{Y}_s = (\mathbf{Y}'(s_1), \dots, \mathbf{Y}'(s_I))^T$, gdzie $\mathbf{Y}(s_i) = (Y(s_i, t_1), \dots, Y(s_i, t_J))^T$. Założenie (1) pozwala wyrazić macierz kowariancji zmiennej $\mathbf{Y}_s$ następująco:

$$\Sigma_{\mathbf{Y}_s}(\sigma^2, \phi, \psi) = \sigma^2 \mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi), \quad (2)$$

gdzie $\otimes$ oznacza iloczyn Kroneckera, $\mathbf{H}_s(\phi)$ – macierz o wymiarach $I \times I$, której element (i, i') jest równy $\rho^{(1)}(s_i - s_{i'}; \phi)$, $\mathbf{H}_t(\psi)$ – macierz o wymiarach $J \times J$, której element (j, j') jest równy $\rho^{(2)}(t_j - t_{j'}; \psi)$.

W referacie wykorzystana zostanie wykładnicza funkcja korelacji przestrzennej postaci [1]:

$$\rho^{(1)}(s_i - s_{i'}; \phi) = \exp(-\phi \|s_i - s_{i'}\|), \quad \phi > 0. \quad (3)$$

Z kolei funkcję korelacji czasowej przyjęto postaci:

$$\rho^{(2)}(t_j - t_{j'}; \psi) = \frac{\psi^{|j-j'|}}{1 - \psi^2}, \quad \psi \in (-1, 1). \quad (4)$$

Parametr ϕ określa, jak szybko zanika korelacja między wartościami procesu wraz ze wzrostem odległości między punktami pomiarowymi. Jego wartość zależy również od przyjętej dla odległości jednostki miary. Z kolei parametr ψ określa współczynnik autokorelacji 1 rzędu.

Funkcja (4) odpowiada funkcji autokorelacji procesu autoregresji rzędu 1 (AR(1)). Przyjęcie takiej funkcji korelacji czasowej wynika z przeprowadzonej analizy funkcji autokorelacji i autokorelacji cząstkowej poziomów zanieczyszczeń na Dolnym Śląsku.

3. Rozkład *a priori* i *a posteriori* parametrów

Przyjęte zostaną następujące rozkłady *a priori* parametrów modelu [1]:

$$\beta \sim N(0, \Sigma_0), \quad \Sigma_0 = \sigma_0 I, \quad (5)$$

$$p(\psi) = 1, \quad \psi \in \langle 0, 1 \rangle, \quad (6)$$

$$\phi \sim G(\kappa_1; \kappa_2), \quad (7)$$

$$\sigma^{-2} \sim G(\varepsilon, \varepsilon), \quad (8)$$

gdzie $G(\cdot, \cdot)$ oznacza rozkład Gamma. W dalszej analizie przyjęto $\sigma_0 = 1000$, $\varepsilon = 0,001$, $\kappa_1 = 0,03$ i $\kappa_2 = 0,01$.

Logarytm funkcji wiarygodności w opisanym modelu ma postać:

$$\begin{aligned} \ln p(\mathbf{Y}_s | \beta, \sigma^2, \phi, \psi) = \\ = -\frac{1}{2} \ln |\sigma^2 \mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi)| - \frac{1}{2\sigma^2} (\mathbf{Y}_s - \mu_s(\beta))^T (\mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi))^{-1} (\mathbf{Y}_s - \mu_s(\beta)), \end{aligned} \quad (9)$$

gdzie:

$$\mu_s(\beta) = \begin{bmatrix} \mu_{s_1}(\beta) \\ \vdots \\ \mu_{s_l}(\beta) \end{bmatrix} \otimes \mathbf{J}, \quad \mathbf{J} = \underbrace{[1 \quad \dots \quad 1]}_J^T.$$

Z własności iloczynu Kroneckera wiadomo (por. (20) i (17)), że

$$|\sigma^2 \mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi)| = (\sigma^2)^J |\mathbf{H}_s(\phi)|^J |\mathbf{H}_t(\psi)|^J,$$

oraz że

$$(\mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi))^{-1} = \mathbf{H}_s^{-1}(\phi) \otimes \mathbf{H}_t^{-1}(\psi),$$

co oznacza, że do wyznaczenia macierzy odwrotnej do macierzy kowariancji (2) o wymiarze $IJ \times IJ$ wymagane jest obliczanie wyznacznika i macierzy odwrotnej dla macierzy rzędu $I \times I$ oraz $J \times J$. Oznacza to, że w przykładzie przedstawionym w referacie nie jest wymagane odwracanie macierzy rzędu 990×990 , a jedynie rzędu 90×90 i 11×11 , co jest tym bardziej istotne, że operacja ta musi być wykonana 10 000 razy.

Rozkład *a posteriori* parametrów $p(\beta, \sigma^2, \phi, \psi | \mathbf{Y}_s)$ jest proporcjonalny do iloczynu

$$p(\beta) p(\phi) p(\psi) p(\sigma^2) p(\mathbf{Y}_s | \beta, \sigma^2, \phi, \psi).$$

Wyznaczenie charakterystyk rozkładu *a posteriori* wymaga zastosowania metod numerycznych. W artykule zastosowany zostanie algorytm Gibbsa–Metropolisa (ang. *Metropolis within Gibbs* – por. np. [4]). Algorytm ten polega na losowaniu z rozkładów warunkowych tylko dla tych składowych rozkładu *a posteriori*, dla których można je analitycznie wyznaczyć (tak jak w standardowym algorytmie Gibbsa). Jeżeli dla pewnej lub pewnych składowych nie można wyznaczyć rozkładu warunkowego, stosuje się dla nich algorytm Metropolisa. W opisywanym modelu nie można analitycznie wyznaczyć rozkładów warunkowych dla parametrów ψ oraz ϕ . Rozkłady warunkowe dla parametrów β oraz σ^2 są następujące [2]:

$$\beta | \phi, \psi, \sigma^2, \mathbf{Y}_s \sim N \left(\left(\mathbf{F}^T \Sigma_{\mathbf{Y}_s}^{-1} \mathbf{F} + \Sigma_0^{-1} \right)^{-1} \mathbf{F}^T \Sigma_{\mathbf{Y}_s}^{-1} \mathbf{Y}_s, \left(\mathbf{F}^T \Sigma_{\mathbf{Y}_s}^{-1} \mathbf{F} + \Sigma_0^{-1} \right)^{-1} \right), \quad (10)$$

$$\sigma^2 | \beta, \phi, \psi, \mathbf{Y}_s \sim G \left(\varepsilon + \frac{IJ}{2}, \varepsilon + \frac{(\mathbf{Y}_s - \mathbf{F}\beta)^T \sigma^2 \Sigma_{\mathbf{Y}_s}^{-1} (\mathbf{Y}_s - \mathbf{F}\beta)}{2} \right), \quad (11)$$

gdzie

$$\mathbf{F} = \begin{bmatrix} 1 & \mathbf{s}_1^T \\ \vdots & \vdots \\ 1 & \mathbf{s}_I^T \end{bmatrix} \otimes \mathbf{J}, \quad \mathbf{J} = \underbrace{[1 \quad \dots \quad 1]}_J^T. \quad (12)$$

4. Prognozowanie poziomów zanieczyszczenia

Załóżmy, że celem jest prognozowanie poziomu zanieczyszczenia w K nowych lokalizacjach $\mathbf{s}'_1, \dots, \mathbf{s}'_K$ w okresach t_j ($j=1, \dots, J$). Niech $\mathbf{Y}_s^T = (\mathbf{Y}(\mathbf{s}'_1), \dots, \mathbf{Y}(\mathbf{s}'_K))$, gdzie $\mathbf{Y}(\mathbf{s}'_i) = (Y(\mathbf{s}'_i, t_1), \dots, Y(\mathbf{s}'_i, t_J))^T$. Wyznaczenie prognoz będzie polegać na znalezieniu charakterystyk (np. wartości oczekiwanej) rozkładu $p(\mathbf{Y}_{s'} | \mathbf{Y}_s)$.

Rozkład łączny wektora $(\mathbf{Y}_s, \mathbf{Y}_{s'})$ ma postać [1]:

$$(\mathbf{Y}_s, \mathbf{Y}_{s'}) \sim N((\boldsymbol{\mu}_s(\boldsymbol{\beta}), \boldsymbol{\mu}_{s'}(\boldsymbol{\beta})), \boldsymbol{\Sigma}_{\mathbf{Y}_s, \mathbf{Y}_{s'}}), \quad (13)$$

gdzie

$$\boldsymbol{\Sigma}_{\mathbf{Y}_s, \mathbf{Y}_{s'}} = \sigma^2 \begin{pmatrix} \mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi) & \mathbf{H}_{s,s'}(\phi) \otimes \mathbf{H}_t(\psi) \\ \mathbf{H}_{s,s'}^T(\phi) \otimes \mathbf{H}_t(\psi) & \mathbf{H}_{s'}(\phi) \otimes \mathbf{H}_t(\psi) \end{pmatrix}. \quad (14)$$

Rozkład warunkowy $\mathbf{Y}_{s'} | \mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \phi, \psi$ jest rozkładem normalnym z wartością oczekiwaną:

$$\begin{aligned} \boldsymbol{\mu}_{s'}(\boldsymbol{\beta}) + (\mathbf{H}_{s,s'}^T(\phi) \otimes \mathbf{H}_t(\psi))(\mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi))^{-1}(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})) = \\ = \boldsymbol{\mu}_{s'}(\boldsymbol{\beta}) + (\mathbf{H}_{s,s'}^T(\phi) \mathbf{H}_s^{-1}(\phi) \otimes I_{J \times J})(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})) \end{aligned} \quad (15)$$

i macierzą kowariancji:

$$\begin{aligned} \sigma^2 (\mathbf{H}_{s'}(\phi) \otimes \mathbf{H}_t(\psi) - (\mathbf{H}_{s,s'}^T(\phi) \otimes \mathbf{H}_t(\psi))(\mathbf{H}_s(\phi) \otimes \mathbf{H}_t(\psi))^{-1}(\mathbf{H}_{s,s'}(\phi) \otimes \mathbf{H}_t(\psi))) = \\ = \sigma^2 (\mathbf{H}_{s'}(\phi) - \mathbf{H}_{s,s'}^T(\phi) \mathbf{H}_s^{-1}(\phi) \mathbf{H}_{s,s'}(\phi)) \otimes \mathbf{H}_t(\psi). \end{aligned} \quad (15)$$

We wzorze (15) wykorzystano własności (17) i (18), we wzorze (16) dodatkowo własność (19).

Aby na podstawie rozkładu $p(\mathbf{Y}_{s'} | \mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \phi, \psi)$ wyznaczyć rozkład $p(\mathbf{Y}_{s'} | \mathbf{Y}_s)$ w ramach algorytmu Gibbsa–Metropolisa, należy generować próbki $(\boldsymbol{\beta}^{(i)}, \sigma^{2(i)}, \phi^{(i)}, \psi^{(i)})$ najpierw z rozkładu *a posteriori* $p(\boldsymbol{\beta}, \sigma^2, \phi, \psi | \mathbf{Y}_s)$, a następnie z rozkładu $p(\mathbf{Y}_{s'} | \mathbf{Y}_s, \boldsymbol{\beta}^{(i)}, \sigma^{2(i)}, \phi^{(i)}, \psi^{(i)})$.

5. Analiza poziomu zanieczyszczenia na Dolnym Śląsku

Opisana procedura zostanie wykorzystana do prognozowania poziomu zanieczyszczenia dwutlenkiem siarki (SO_2) na Dolnym Śląsku. Wykorzystano dane ze stycznia, lutego i marca 2006 r. otrzymane w lokalizacjach wymienionych w tab. 1. Ich położenie przedstawiono na rys. 1. Dane te są dostępne na stronie internetowej Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu (<http://www.wroclaw.pios.gov.pl/>). Zgodnie z przyjętymi oznaczeniami liczba stacji pomiarowych $I = 11$, natomiast liczba okresów $J = 90$.

Tabela 1. Lista stacji pomiarowych

Lp.	Lokalizacja	Długość geograficzna (E)	Szerokość geograficzna (N)
1	Czarna Góra	16°48'51"	50°15'17"
2	Działoszyn	14°56'34"	50°58'17"
3	Dzierżoniów	16°38'56"	50°43'59"
4	Jelenia Góra – Cieplice	15°44'06"	50°51'30"
5	Legnica	16°10'50"	51°12'15"
6	Nowa Ruda	16°30'53"	50°34'44"
7	Wałbrzych	16°16'13"	50°46'08"
8	Wleń	15°40'14"	51°00'45"
9	Wrocław – Karłowice	17°01'46"	51°07'46"
10	Wrocław – Wiśniowa	17°00'46"	51°05'11"
11	Zgorzelec	15°01'30"	51°08'40"

Źródło: opracowanie na podstawie danych Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu.

Na podstawie długości i szerokości geograficznej obliczono macierz odległości (tab. 2). Odległość między dwiema lokalizacjami obliczono według następującego wzoru [5]:

$$d = 6371 \cdot \arccos(B) [km],$$

gdzie $B = \sin(x_1)\sin(x_2) + \cos(x_1)\cos(x_2)\cos(y_1 - y_2)$, x_1, x_2 – szerokość geograficzna lokalizacji 1 i 2 (w radianach), y_1, y_2 – długość geograficzna lokalizacji 1 i 2 (w radianach).

Wszystkie współrzędne geograficzne oraz odległości przeliczone zostały na kilometry.

Estymacja parametrów $\beta, \phi, \psi, \sigma^2$ modelu oraz wartości prognoz została wykonana z użyciem algorytmu Gibbsa–Metropolisa. Wykonano 20 000 iteracji, z czego ostatnie 10 000 wykorzystano do wyznaczenia rozkładów *a posteriori* parametrów. Czas obliczeń wyniósł ok. 30 minut (Intel PIV, 2,8 GHz). Charakterystyki rozkładu *a posteriori* parametrów zebrano w tab. 3.

Otrzymane wartości parametrów β_1 oraz β_2 oznaczają, że zmiana położenia stacji pomiarowej o 100 km na wschód zwiększa wartość oczekiwaną poziomu zanieczyszczenia o ok. 5,25 jednostki, natomiast zmiana położenia o 100 km na północ zwiększa wartość oczekiwaną o ok. 12,19 jednostki. Wartość $\phi = 0,17$ oznacza, że korelacja zanieczyszczenia między stacjami odległymi o 1 km wynosi 0,84, odległymi o 10 km jest równa 0,18 i praktycznie zanika w odległości 20 km.



Rys. 1. Mapa stacji pomiarowych

Źródło: <http://www.wroclaw.pios.gov.pl>.

Tabela 2. Macierz odległości między stacjami pomiarowymi (w km)

	Czarna Góra	Działoszyn	Dzierżoniów	Jelenia Góra – Cieplice	Legnica	Nowa Ruda	Wałbrzych	Wleń	Wrocław – Karłowice	Wrocław – Wiśniowa	Zgorzelec
Czarna Góra	0,00	154,22	54,46	101,57	114,61	41,83	68,90	116,64	98,44	93,53	160,20
Działoszyn	154,22	0,00	122,66	56,94	90,24	118,83	95,84	51,14	146,91	145,32	20,08
Dzierżoniów	54,46	122,66	0,00	65,73	61,80	19,58	26,93	75,35	51,52	46,85	122,63
Jelenia Góra – Cieplice	101,57	56,94	65,73	0,00	49,49	63,08	38,90	17,73	95,48	92,99	59,00
Legnica	114,61	90,24	61,80	49,49	0,00	73,37	48,81	41,50	59,77	59,51	80,83
Nowa Ruda	41,83	118,83	19,58	63,08	73,37	0,00	27,26	76,45	71,09	66,39	122,01
Wałbrzych	68,90	95,84	26,93	38,90	48,81	27,26	0,00	50,03	66,60	62,88	96,70
Wleń	116,64	51,14	75,35	17,73	41,50	76,45	50,03	0,00	95,83	94,18	47,42
Wrocław – Karłowice	98,44	146,91	51,52	95,48	59,77	71,09	66,60	95,83	0,00	4,93	139,86
Wrocław – Wiśniowa	93,53	145,32	46,85	92,99	59,51	66,39	62,88	94,18	4,93	0,00	138,90
Zgorzelec	160,20	20,08	122,63	59,00	80,83	122,01	96,70	47,42	139,86	138,90	0,00

Źródło: opracowanie własne.

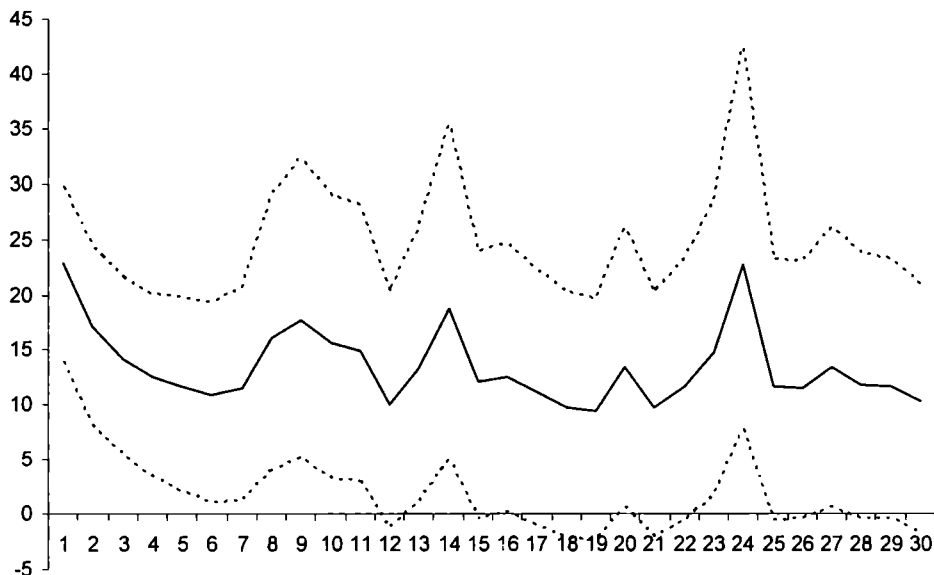
Wartość $\psi = 0,66$ oznacza, że korelacja poziomu zanieczyszczenia w dwóch kolejnych dniach jest równa 0,66. Oszacowanie tej korelacji jest jednak dosyć nieprecyzyjne – 95-procentowy przedział wiarygodności jest równy (0,44; 0,80).

Tabela 3. Charakterystyki rozkładu *a posteriori* parametrów

Charakterystyka	Parametr					
	β_0	β_1	β_2	ϕ	ψ	σ
Wartość oczekiwana	9,384562	0,052485	0,121973	0,172295	0,663868	15,32982
Odchylenie standardowe	1,840261	0,008218	0,013838	0,024145	0,094883	0,372223
Kwantyl rzędu 2,5%	5,029076	0,035385	0,100977	0,116397	0,441255	14,66676
Kwantyl rzędu 97,5%	11,98464	0,068428	0,153574	0,203019	0,803822	16,20331

Źródło: opracowanie własne.

Na rysunku 2 przedstawiono prognozę poziomu zanieczyszczenia dla lokalizacji Wrocław – Rynek w dniach 1-30 stycznia 2006. Linia ciągła oznacza prognozę punktową, natomiast linie przerywane wyznaczają 90-procentowy obszar wiarygodności dla prognozy.



Rys. 2. Prognoza poziomu zanieczyszczenia

Źródło: opracowanie własne.

Załącznik – definicja i własności iloczynu Kroneckera

Niech $\mathbf{A}$ oznacza macierz o wymiarach $m \times n$ oraz $\mathbf{B}$ macierz o wymiarach $p \times q$. Wówczas iloczyn Kroneckera $\mathbf{A} \otimes \mathbf{B}$ jest macierzą blokową o wymiarach $mp \times nq$ postaci:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{1n}\mathbf{B} \\ \vdots & \dots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix}.$$

Iloczyn ten ma m.in. następujące własności:

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}, \quad (17)$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}, \quad (18)$$

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}. \quad (19)$$

Ponadto, jeżeli $\mathbf{A}$ oraz $\mathbf{B}$ są macierzami kwadratowymi o wymiarach odpowiednio p oraz q , to:

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^q |\mathbf{B}|^p. \quad (20)$$

Twierdzenia i definicje dotyczące iloczynu Kroneckera można znaleźć np. w [3].

Literatura

- [1] Banerjee S., Carlin B.P., Gelfand A.E., *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton 2003.
- [2] Koop G., *Bayesian Econometrics*, Wiley, New York 2003.
- [3] Rao C.R., Rao M.B., *Matrix Algebra and Its Applications to Statistics and Econometrics*, World Scientific Publishing, Singapore 1998.
- [4] Robert C.P., Casella G., *Monte Carlo Statistical Methods*. Springer, New York 2004.
- [5] Sahu S.K., Mardia K.V., *A Bayesian Kriged Kalman Model for Short-term Forecasting of Air Pollution Levels*, „Journal of the Royal Statistical Society”, Series C (Applied Statistics) 2005, 54, 223-244.
- [6] Strona internetowa Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu <http://www.wroclaw.pios.gov.pl/>.

SPATIO-TEMPORAL POLLUTION ANALYSIS

Summary

Environmental pollution can be analysed both spatially and temporally. In the spatial context, it is usually assumed that pollution correlation between monitoring sites is dependent on the distance between them. In the temporal analysis the correlation depends on the time lag between the measurements.

The paper presents one of the methods of spatio-temporal modelling. It will be assumed that the pollution is a Gaussian process $Y(s, t)$, where s denotes the location of the measurement site ($s \in R^1$), and t is the measurement time.

The model is illustrated by the analysis of pollution in the Lower Silesia region. The parameters are estimated with the Bayesian approach using the Gibbs sampler.