

Inteligentny system monitoringu - modelowanie rozkładów przestrzenno-czasowych niektórych substancji w powietrzu

Autorzy:

Konrad Garanty*)
Mieczysław Sowiński
Bogumiła Mysłek-Laurikainen
Jolanta Wojtkowska
Marek Kowalski

1. WPROWADZENIE

Inspiracją do podjęcia tej pracy były międzynarodowe (UE) i polskie regulacje prawne, m.in. ustawa "Prawo ochrony środowiska" z dnia 27 kwietnia 2001 r. oraz "Rozporządzenie Ministra Środowiska z dnia 6 czerwca 2002 r. w sprawie dopuszczalnych poziomów niektórych substancji w powietrzu, alarmowych poziomów niektórych substancji w powietrzu oraz marginesów tolerancji dla dopuszczalnych poziomów niektórych substancji" (Dz.Ust. Nr 87 poz. 796) i "Rozporządzenie Ministra Środowiska z dnia 6 czerwca 2002 r. w sprawie oceny poziomów substancji w powietrzu" (Dz.Ust. Nr 87 poz. 798).

W opracowaniu [1] oraz referacie [2] zaprezentowano wstępne wyniki badań możliwości aplikacji metody sieci neuronowych do rozwiązywania różnych zadań dotyczących jakości powietrza atmosferycznego.

W roku 2003 kontynuowano prace związane z wykorzystaniem metody sieci neuronowych do zarządzania i kontroli jakości powietrza atmosferycznego, w oparciu o wieloletnią bazę wiarygodnych danych pomiarowych stężeń zanieczyszczeń powietrza w Aglomeracji Gdańskiej (emisji SO_2 / NO / NO_2 / NO_x / CO / O_3 / Pm_{10} (pyłu) oraz parametrów meteorologicznych (prędkość i kierunek wiatru, temperatura, ciśnienie, wilgotność powietrza, opad, nasłonecznienie) tworzoną przez ARMAAG. Szczególnie obiecujące okazało się zastosowanie analizy falkowej [3] do modelowania rozkładów przestrzenno-czasowych stężeń substancji.

*Institut Problemów Jądrowych
) Uniwersytet Warszawski

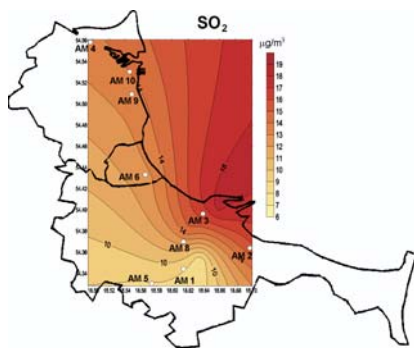
2. CEL I ZAKRES PRACY

Celem pracy jest opracowanie systemu modelowania rozkładów przestrzenno-czasowych niektórych substancji w powietrzu. Zakres pracy obejmuje:

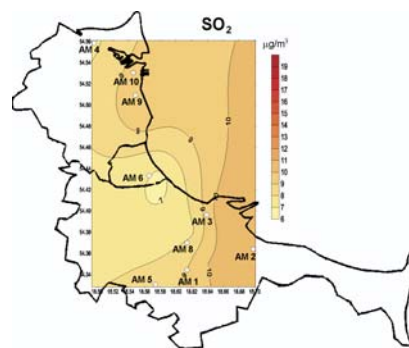
- Analizę wpływu zmian poszczególnych parametrów meteorologicznych (dostępnych w bazie danych ARMAAG), korelacji pomiędzy stężeniami różnych substancji (pierwotnych i wtórnych), rozmieszczenia źródeł substancji w skali lokalnej (pojedyncza stacja pomiarowa), w skali całej aglomeracji albo jej części na dokładność wyników modelowania interpolacji i ekstrapolacji przestrzennej lub czasowej stężeń substancji. Do modelowania ekstrapolacji zanieczyszczeń przewiduje się również możliwość wykorzystania wyników pochodzących z pomiarów manualnych i wskaźnikowych a w szczególnych przypadkach - wyników uzyskanych za pomocą mobilnej stacji pomiarów automatycznych.
- Opracowanie modeli predykcji krótko- i długookresowej rozkładów przestrzenno-czasowych stężeń ww. substancji opartych na metodach sieci neuronowych oraz metodach analizy falkowej.
- Testowanie i optymalizowanie tych modeli przy z wykorzystaniem wieloletniej bazy wiarygodnych danych pomiarowych z Aglomeracji Gdańskiej.

Poniżej, na Rys.1-6 przedstawiono sezonowe (okres grzewczy, okres letni) rozkłady przestrzenno-czasowe wybranych substancji: SO_2 / NO_2 / Pm_{10} (pyłu). Na Rys.7-9 pokazano przykładowo wybrane godzinne rozkłady przestrzenno-czasowe NO_2 a na Rys.10-12 dobowe rozkłady przestrzenno-czasowe pyłu dla wybranych dni w okresie grzewczym i okresie letnim. Okresy uśredniania wyników pomiarów (1 godzina dla NO_2 i 24 godziny dla pyłu) są zgodne z Zał. 1 do wyżej wspomnianego Rozporządzenia Ministra Środowiska (Dz.Ust. Nr 87 poz. 796).

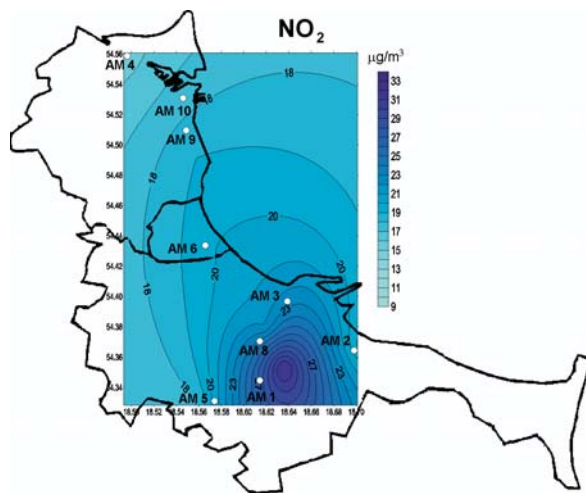
Przytoczone tu przykłady rozkładów przestrzenno-czasowych niektórych substancji wskazują, że już w latach 2005 -2006 mogą zaistnieć pewne trudności z dotrzymaniem poziomów dopuszczalnych niektórych substancji i dopuszczalnej częstotliwości ich przekraczania, nie tylko w obrębie stacji pomiarowych ale również na znacznie większych obszarach, szczególnie w Gdańsku. Dotyczy to zwłaszcza PM_{10} / NO_x / NO_2 a być może nawet SO_2 .



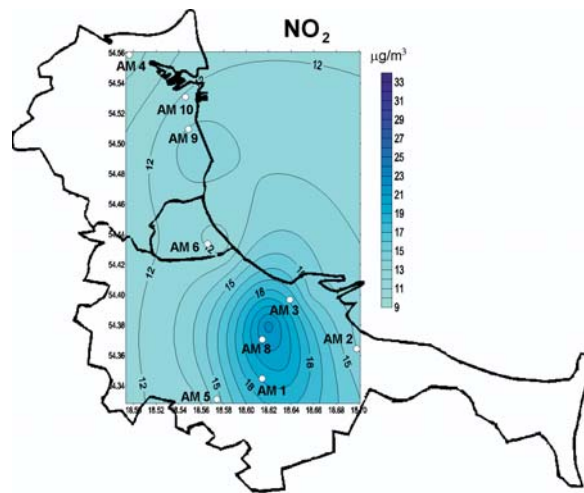
Rys.1: Sezonowy rozkład przestrzenno-czasowy SO_2 w okresie grzewczym 10.2002-03.2003



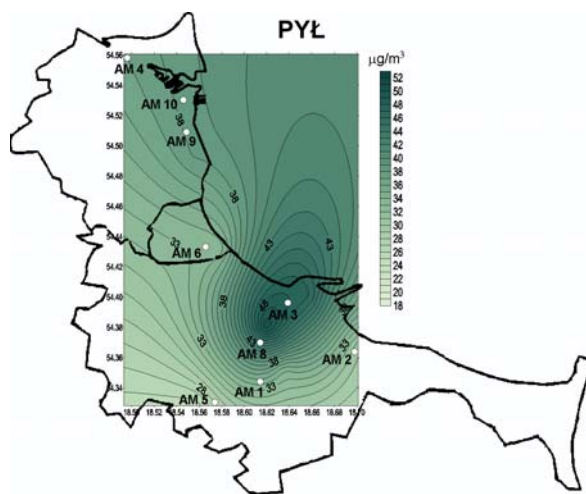
Rys.2: Sezonowy rozkład przestrzenno-czasowy SO_2 w okresie letnim 04.2003-09.2003



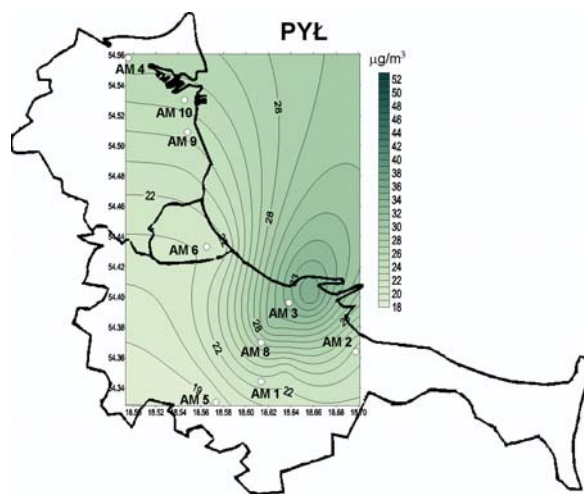
Rys.3: Sezonowy rozkład przestrzenno-czasowy NO_2 w okresie grzewczym 10.2002-03.2003



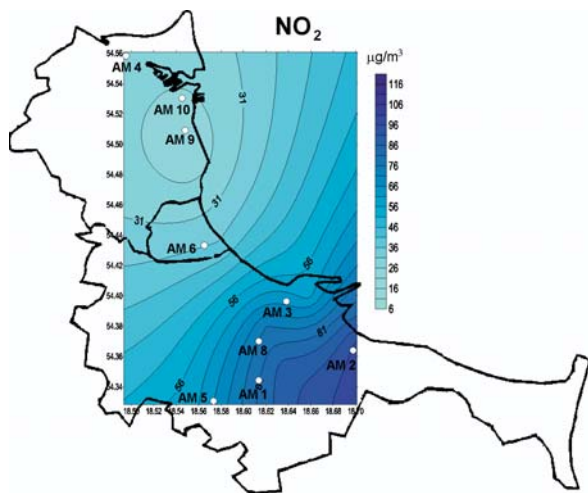
Rys.4: Sezonowy rozkład przestrzenno-czasowy NO_2 w okresie letnim 04.2003-09.2003



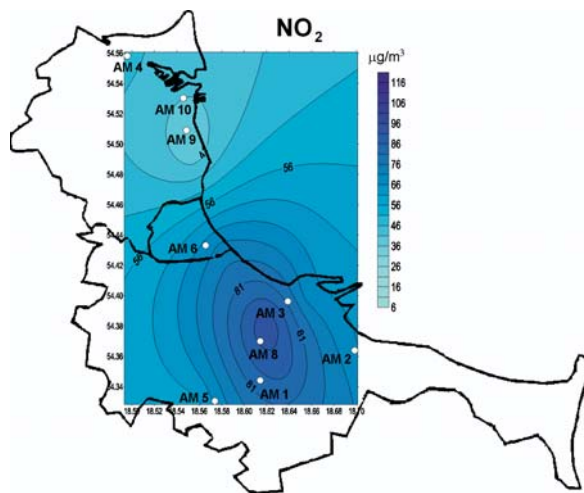
Rys.5: Sezonowy rozkład przestrzenno-czasowy pyłu w okresie grzewczym 10.2002-03.2003



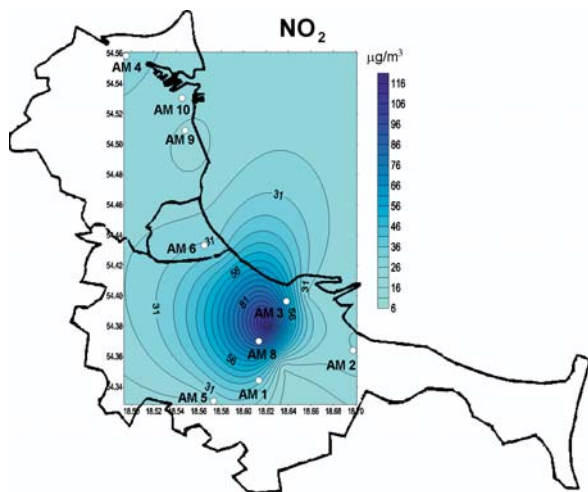
Rys.6: Sezonowy rozkład przestrzenno-czasowy pyłu w okresie letnim 04.2003-09.2003



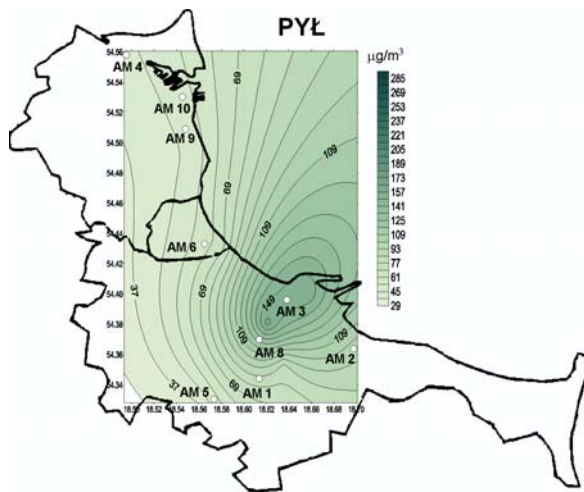
Rys.7: Wybrany godzinny rozkład przestrzenno-czasowy NO_2 w okresie grzewczym, dnia 23.12.2002 (dane z godz. 12:00, 12:30)



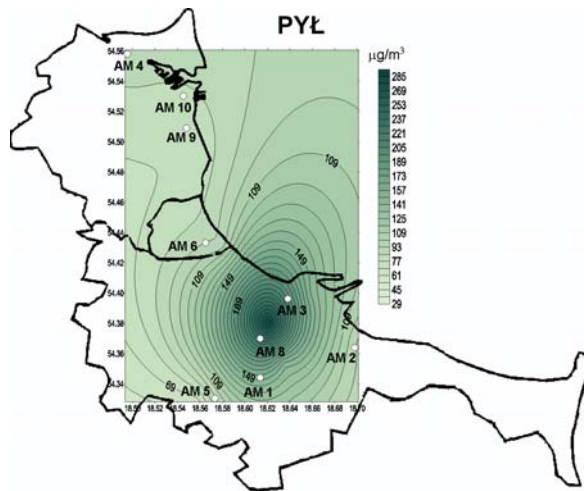
Rys.8: Wybrany godzinny rozkład przestrzenno-czasowy NO_2 w okresie grzewczym, dnia 07.01.2003 (dane z godz. 19:30, 20:00)



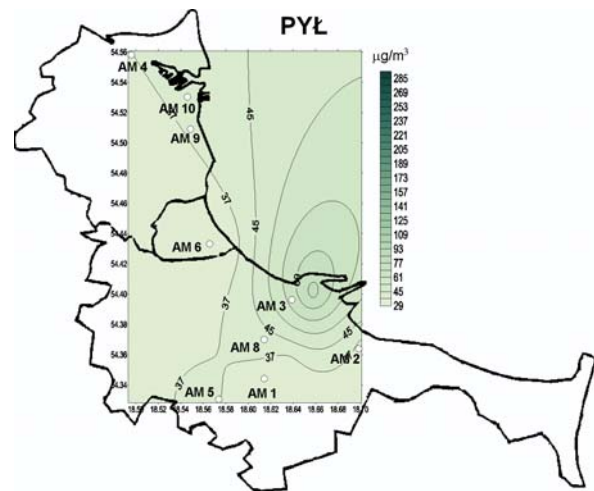
Rys.9: Wybrany godzinny rozkład przestrzenno-czasowy NO_2 w okresie letnim, dnia 29.08.2003 (dane z godz. 08:00, 08:30)



Rys.10: Wybrany dobowy rozkład przestrzenno-czasowy pyłu w okresie grzewczym 23.12.2002



Rys.11: Wybrany dobowy rozkład przestrzenno-czasowy pyłu w okresie grzewczym 07.01.2003



Rys.12: Wybrany dobowy rozkład przestrzenno-czasowy pyłu w okresie letnim 19.08.2003

3. PODSUMOWANIE

Dotychczasowe wyniki naszych analiz pokazują, że modelowanie rozkładów przestrzenno-czasowych z wykorzystaniem sieci neuronowych w połączeniu z metodami analizy falkowej może stanowić unikatowe narzędzie do wyznaczania stref zanieczyszczeń powietrza (oddzielnie dla każdej substancji), prognozowania (predykcji) ich stężeń, identyfikacji źródeł substancji, optymalizacji czynników ekologicznych, technologicznych i ekonomicznych. Ponadto, może ono być komplementarne z innymi metodami modelowania, np. metodą dyfuzyjną.

[1] - Inteligentny system monitoringu, jako instrument zarządzania i/lub kontroli jakości powietrza (na przykładzie Aglomeracji Gdańskiej), M. Kowalski, B. Mysłek-Laurikainen, Z. Moroz, M. Sowiński, J. Wojtkowska. Stan zanieczyszczenia powietrza atmosferycznego w Aglomeracji Gdańskiej w roku 2002 i informacja o działalności Fundacji ARMAAG, Raport 2002, str. 69. Raport przygotował zespół T. Kołakowski, M. Sarafin, T. Waszczyk, E. Kachniarz pod kierunkiem K. Szymańskiej. Agencja Regionalnego Monitoringu Atmosfery Aglomeracji Gdańskiej 2003.

[2] - Inteligentny system monitoringu, jako instrument zarządzania i/lub kontroli jakości powietrza, Z. Moroz, M. Sowiński, J. Wojtkowska, B. Mysłek-Laurikainen, M. Kowalski, K. Szymańska, T. Kołakowski, referat wygłoszony przez M. Sowińskiego. Konferencja 29-30 maja 2003, Gdańsk. Sieć monitoringu ARMAAG - inicjatywa samorządowa jako narzędzie zarządzania środowiskiem w regionie u progu wejścia do Unii Europejskiej

[3] - Falki i aproksymacje, Jan T. Białasiewicz, Wydawnictwa Naukowo-Techniczne, Warszawa 2000

Statystyczna interpolacja braków danych pomiarowych

Autor:
Piotr Oskar Czechowski

Skróty:

BD - brak danych,

AMxx - oznaczenie automatycznej stacji pomiarowej AM, gdzie xx od 1 do 8,

BDi - seria braków danych,

BDi=xxx - długość serii braków danych, gdzie xxx oznacza długość serii.

1. Klasyfikacja braków danych pomiarowych

Występowanie braków danych w pomiarach sieci monitorującej stan atmosfery są poważnym problemem. Ciągi wyników pomiarów, traktowane jako realizacje pojedynczych procesów stochastycznych, znajdujących zastosowanie m.in. w prognozowaniu stanu zanieczyszczeń atmosfery. Braki danych powodują niemożność dokonywania poprawnych obliczeń a co za tym idzie prognoz.

Interpolacje braków dla wyników pomiarów godzinnych stężeń zanieczyszczeń (imisji) uzależniono od okresu wystąpienia braków oraz długości wystąpienia serii braków danych.

Kluczowym zagadnieniem jest długość ciągu, kolejno po sobie występujących, braków danych. Założono, że inne metody będą stosowane w przypadku krótkich ciągów braków, w których zakłada się, że upływ czasu jest głównym czynnikiem kształtującym, inne zaś dla długich ciągów, w których niekiedy zachodzi w istocie konieczność odtworzenia sytuacji imisyjnej. Odtworzenie tej sytuacji możliwe jest poprzez wykorzystanie informacji o dodatkowych czynnikach, spośród których zakłada się, że najważniejszymi będą:

- sytuacja imisyjna w bezpośrednim sąsiedztwie przestrzennym (wyniki pomiarów z innych stacji),
- wpływ warunków meteorologicznych,
- warunki emisji na wybranym obszarze reprezentatywności stacji pomiarowej, dla której przeprowadzane będą interpolacje (w miarę dostępności danych).

Przyjęto wstępnie następujące kategorie długości ciągów braków danych pomiarowych:

Kategoria	Długość ciągu braków danych (BDi)	Założony charakter modelu	Dziedzina modelu
A	1 - 3	Interpolacyjny	Czasu
B	4 - 12	Interpolacyjny	Czasowo - częstotliwościowa
C	13 - 24	Interpolacyjno - odtwarzający	Czasowo - częstotliwościowa
D	25 - 70	Interpolacyjno - odtwarzający	Czasowo - częstotliwościowa
E	71 - 400	Interpolacyjno - odtwarzający	Czasowo - częstotliwościowa
F	401 - 4000	Interpolacyjno - odtwarzający	Czasowo - częstotliwościowa

Tablica 1: Podział proponowany [Opracowanie własne]

1.1. Zakres metod

Jako podstawy do dalszych prac przyjęto następujące metody:

Kategoria	Grupa metod	metody interpolacji	Oznaczenie
A	Podstawowa I	średnia ogólna	A_1
	Podstawowa I	średnia z sąsiednich punktów	A_2
	Podstawowa I	średnia ważona (wagami są odległości od sąsiednich punktów)	A_3
	Podstawowa II	średnia k ucięta	A_4
	Podstawowa II	średnia p ucięta	A_5
	Podstawowa II	Średnie typu Winsora	A_6
	Podstawowa I	regresja prosta (MNK) pomiędzy sąsiednimi punktami	A_20
	Podstawowa I	regresja prosta (MNK) interpolacja trendem liniowym	A_21
	Podstawowa I	regresja prosta (MNK) interpolacja trendem liniowym ze wskaźnikami cyklicznymi	A_22
	Podstawowa II	wyrównanie wykładnicze (klasyfikacja Pegela; wybór)	A_40
	Podstawowa II	regresja harmoniczna	A_60
	Podstawowa II	funkcje sklepane	A_80
	Podstawowa I	metody odporne (np. Mediana)	A_100
B	Podstawowa I	regresyjne (wskaźnikowe)	B_1
	Podstawowa II	metoda wag harmonicznch	B_20
	Podstawowa II	funkcje sklepane	B_40
C	Podstawowa I	z wykorzystaniem informacji przestrzennej (wielowymiarowe złożenia liniowe)	C_1
	Rozszerzona I	z wykorzystaniem informacji przestrzennej (wielowymiarowe złożenia meteorologiczne)	C_5
	Rozszerzona I	wielowymiarowe procesy stochastyczne (informacje meteorologiczne)	C_10
	Rozszerzona II	wielowymiarowe procesy spektralne (informacje meteorologiczne)	C_15
	Rozszerzona I	z wykorzystaniem informacji meteorologicznych	C_20

Tablica 2: Zakres metod w podziale na grupy, będące etapami prac [Opracowanie własne]

1.2. Mierniki oceny jakości metod

1. RMSE

Ważnym, z punktu widzenia oceny jakości modelu, jest ocena jego błędów wyrażona w postaci błędu średniokwadratowego.

gdzie:

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2$$

$e_t = x_t - \hat{x}_t$ - reszty modelu.

W praktyce zamiast błędu średniokwadratowego zastosowanie znalazł jego pierwiastek, odchylenie standardowe składnika resztowego z powodu prostoty interpretacji.

$$RMSE = \sqrt{MSE}$$

2.1. Procedura badawcza

Celem badania przyjęto interpolację braków w danych wyników pomiarów sieci automatycznej stężenia dwutlenku siarki na stacji AM1. Aby można było ocenić jakość poszczególnych metod uprzednio przeprowadzono symulację braków w danych:

1. Wybrano dwa miesiące pomiarów dla stacji (AM1) bez braków danych, odpowiednio dla sezonu letniego i grzewczego, dzięki czemu uzyskano dwa zestawy testowe oznaczone: Z1 i Z2. Dodatkowy zestaw Z3 wyników pomiarów całego sezonu, np. letniego dla długich ciągów BD, przeznaczono dla oceny modeli o charakterze odtwarzającym bardzo długie serie braków (powyżej 71 obserwacji).
2. Wybrano w sposób losowy dla zestawów Z1 i Z2 co najwyżej po 10 serii z kategorii A i B oraz co najwyżej 5 serii z kategorii C, w wyniku czego uzyskano zestawy: Z1A, Z1B, Z1C oraz Z2A, Z2B, Z2C. Maksymalna długość ciągu braków w tych zestawach to 70 obserwacji, co oddaje rzeczywiste warunki.
3. Zestawy: Z1E i Z1F oraz Z2E i Z2F zawierały pojedyncze serie braków zgodnie z: Tablica 1.
4. Dla zestawów Z1 i Z2 zastosowanie metod zgodnie z Tablica 2.
5. Przeprowadzono wstępną ocenę jakości w oparciu o RMSE.
6. Na podstawie oceny wstępnej zbudowano plany interpolacyjne uwzględniające okres wystąpienia braków i długość serii wystąpień braków w danych.

Liczbę powtórzeń uzależniono od długości serii bez braków, i tak w przypadkach stosunkowo krótkich serii bez braków liczba powtórzeń była odpowiednio mniejsza.

7. Wnioski.

BDi AM1	ni	BDi AM2	ni	BDi AM3	ni	BDi AM4	ni	BDi AM5	ni	BDi AM8	ni
1	7	1	21	1	8	1	4	1	6	1	4
3	1	3	2	8	1	2	1	2	5	3	3
5	1	9	1	12	1	4	1	3	2	12	1
8	1	18	1	19	1	34	1	4	3	24	1
42	1	66	1	54	1	1%	7	5	1	29	1
65	1							8	1	44	1
1021	1	1%	26	1%	12			13	1	69	1
3425	1							14	1	332	1
52%	14							16	1	6%	13
								18	2		
								19	3		
								21	1		
								25	1		
								51	1		
								3%	29		

Tablica 3: Rozkłady braków danych w pomiarach SO₂ w roku 2003
[Opracowanie własne]

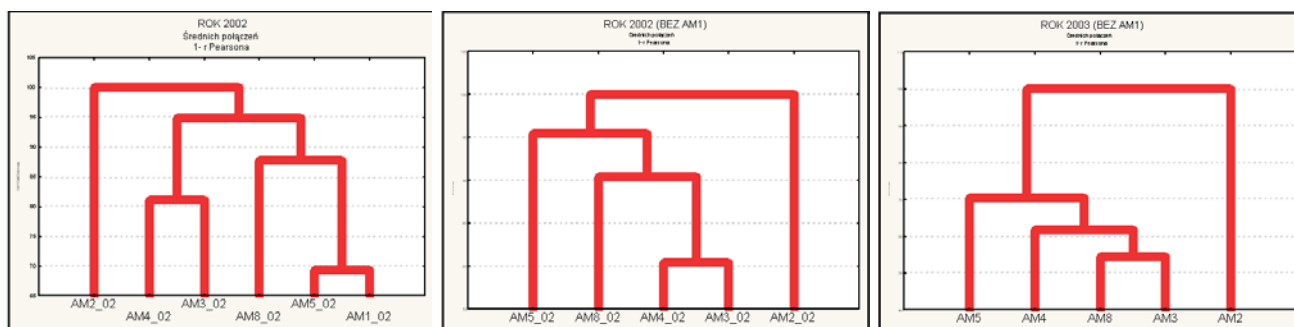
Ni BD AM1	ni	Ni BD AM2	ni	Ni BD AM3	ni	Ni BD AM4	ni	Ni BD AM5	ni	Ni BD AM8	ni
1	7	1	7	1	10	1	11	1	6	1	5
2	1	3	2	2	1	2	1	2	7	2	1
13	1	4	1	3	1	6	1	4	2	3	2
15	1	5	1	4	1	41	1	5	2	13	1
64	1	0.3%	26	11	1	0.7%	14	6	2	15	1
379	1			12	1			7	2	17	1
5.5%	12			15	1			10	1	18	2
				0.7%	16			11	2	30	1
								13	3	107	1
								14	1	163	1
								20	1	4.5%	16
								21	2		
								22	1		
								38	1		
								51	1		
								62	1		
								1178	1		
								17.8%	36		

Tablica 4: Rozkłady braków danych w pomiarach SO₂ w roku 2002
[Opracowanie własne]

3. Przestrzenna analiza podobieństw rozkładów emisji

Wyniki analiz podobieństw, zarówno w dziedzinie częstotliwości jak i w czasowo - częstotliwościowej wskazują, że emisja dwutlenku siarki na stacji AM1 jest najmniej podobna do emisji na stacji AM2, czego należało się spodziewać. Stacja AM2 leży w znacznym oddaleniu, w przeciwieństwie od pozostałych stacji, od skupisk miejskich a co za tym idzie oddziaływanie różnych rodzajów emisji ma innych charakter niż na pozostałych stacjach.

1. Wyniki analizy podobieństw stężeń dwutlenku siarki - analiza skupień:



2. Wyniki analizy podobieństw stężeń dwutlenku siarki - analiza wariancji ANOVA:

Tukey HSD (nierówne N); zmienna $SO_2_0_2$ (SO_2_{2003})
 Grupy jednorodne, $\alpha = 05000$
 Błąd: MS międzygrupowe = 84.653, $df = 49979$

	STACJA	$SO_2_0_2$	1	2	3	4	5
1	AM1	5.92138	****				
2	AM5	6.56853		****			
3	AM4	9.26308			****		
4	AM8	10.92984				****	
5	AM3	10.97105				****	
6	AM2	12.10888					****

Tablica 5: Rok 2002 [Opracowanie własne]