

Szymon HOFFMAN, Rafał JASIŃSKI

Politechnika Częstochowska, Katedra Chemii, Technologii Wody i Ścieków
ul. Dąbrowskiego 69, 42-200 Częstochowa

Klasyfikacja luk pomiarowych w danych rejestrowanych na stacjach monitoringu powietrza

Rejestrowane na stacjach monitoringu powietrza zbiory danych nigdy nie są kompletne. W skali roku liczba odnotowywanych braków jest zmienna. Ocena jakości powietrza na podstawie niepełnych pomiarów jest utrudniona. Obowiązujące przepisy prawne dopuszczają możliwość wykorzystania modelowania w celu uzupełnienia brakujących danych. Rozpoznanie typowych struktur obszarów z brakującymi danymi umożliwia ich klasyfikację, a następnie rekomendację odpowiednich metod modelowania dla wyszczególnionych klas.

Celem badań było wytypowanie charakterystycznych struktur luk pomiarowych w zbiorach danych i określenie częstości ich występowania. Klasyfikację przypadków z brakującymi danymi zaproponowano na podstawie przeglądu wieloletnich danych, pochodzących z kilku różnych stacji pomiarowych automatycznego monitoringu powietrza. Analizowano serie czasowe chwilowych stężeń podstawowych zanieczyszczeń powietrza (O_3 , NO_2 , NO , PM_{10} , SO_2 , CO), zarejestrowanych w latach 2004-2008 na stacjach monitoringu powietrza Warszawa-Ursynów, Radom, Łódź-Widzew, Piotrków Trybunalski.

Na podstawie wyników przeprowadzonej analizy można stwierdzić, że brakujące dane występują powszechnie w zbiorach danych pochodzących z monitoringu powietrza. Częstość ich występowania w rocznych seriach pomiarowych może wynosić od kilku do nawet kilkadziesiąt procent. Większość luk pomiarowych jest krótka - stanowią je głównie pojedyncze przypadki. Zdecydowanie rzadziej występują bloki brakujących danych, przekraczające 3-4 przypadki (dłuższe od 3-4 godzin). Największą częstość występowania przypadków z niezarejestrowanymi wynikami odnotowano dla luk najdłuższych, obejmujących więcej niż 24 przypadki (>24 godziny).

Słowa kluczowe: zanieczyszczenia powietrza, monitoring automatyczny, dane, stężenia chwilowe, brakujące dane, luki pomiarowe, klasyfikacja

Wprowadzenie

Dane zbierane w sposób ciągły na stacjach monitoringu powietrza nigdy nie są kompletne [1]. W skali roku liczba odnotowywanych braków jest zmienna. Brakujące dane stanowią zwykle kilka procent wszystkich możliwych do zarejestrowania danych w rocznych seriach czasowych, ale dla niektórych mierzonych stężeń ta niekompletność może osiągnąć kilkanaście, a nawet kilkadziesiąt procent [2]. Duża ilość niezarejestrowanych stężeń zanieczyszczeń utrudnia, a nawet wyklucza ocenę czystości powietrza wymaganą przez regulacje prawne, w szczególności przez normy imisyjne [3]. Gdy kompletność zbioru jest zbyt mała, obowiązujące przepisy prawne dopuszczają możliwość wykorzystania modelowania w celu uzupełnienia brakujących danych. Zastosowane metody modelowania powinny zapewniać możliwie największą dokładność, aby wygenerowane dane pozwalały na ocenę

najbardziej zbliżoną do rzeczywistości. Brakujące dane mogą tworzyć w macierzy wyników ugrupowania o różnych kształtach i zasięgu. Luki pomiarowe mogą obejmować pojedyncze dane, serie danych odnoszących się do jednego zanieczyszczenia, pojedyncze przypadki w serii czasowej i całe bloki danych (o różnej długości) itp. Rozpoznanie typowych struktur obszarów z brakującymi danymi umożliwia ich klasyfikację, a następnie rekomendację odpowiednich metod modelowania dla wyszczególnionych klas.

Celem badań było wytypowanie charakterystycznych struktur luk pomiarowych w zbiorach danych i określenie częstości ich występowania. Klasyfikację przypadków z brakującymi danymi zaproponowano na podstawie przeglądu danych, pochodzących z kilku różnych stacji pomiarowych automatycznego monitoringu powietrza. Analizę przeprowadzono, wykorzystując wieloletnie zbiory danych chwilowych (średnich 1-godzinnych), zarejestrowane w systemach monitoringu powietrza działających w różnych miejscowościach w Polsce.

1. Metodyka badań

W analizie wykorzystano serie pomiarowe pochodzące z sieci monitoringu powietrza obejmujących swym zasięgiem województwa łódzkie i mazowieckie. Analizowano serie czasowe stężeń chwilowych podstawowych zanieczyszczeń powietrza (O_3 , NO_2 , NO , PM_{10} , SO_2 , CO), zarejestrowanych w latach 2004-2008 na stacjach monitoringu powietrza Warszawa-Ursynów, Radom, Łódź-Widzew, Piotrków Trybunalski. Lokalizację stacji pomiarowych przedstawiono na rysunku 1.



Rys. 1. Lokalizacja stacji pomiarowych

Wstępnie dla badanych serii czasowych wymienionych zanieczyszczeń obliczono podstawowe parametry statystyczne, takie jak: kompletność serii, wartość

maksymalną, wartość minimalną, średnią arytmetyczną i odchylenie standardowe. W dalszym etapie analizy oszacowano liczbę luk pomiarowych o różnych długościach, oddzielnie dla każdego z zanieczyszczeń. Obliczono także częstość występowania przypadków w lukach o różnej długości. Wyniki uzyskane dla różnych stacji monitoringu powietrza zestawiono i porównano.

2. Wyniki i ich omówienie

2.1. Wstępna ocena danych

Podstawowe parametry statystyczne analizowanych serii czasowych przedstawiono w tabelach 1-6.

Tabela 1
Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń O_3 , zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	96,4	94,3	84,8	97,1
Wartość minimalna	$\mu\text{g}/\text{m}^3$	0,0	0,0	0,1	0,9
Wartość maksymalna	$\mu\text{g}/\text{m}^3$	215,3	184,8	189,1	197,9
Średnia	$\mu\text{g}/\text{m}^3$	50,6	44,7	47,6	58,1
Odchylenie standardowe	$\mu\text{g}/\text{m}^3$	31,6	27,9	30,9	31,2

Tabela 2
Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń NO , zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	58,0	89,5	95,1	72,1
Wartość minimalna	$\mu\text{g}/\text{m}^3$	0,0	0,0	0,0	0,0
Wartość maksymalna	$\mu\text{g}/\text{m}^3$	520,6	557,6	419,7	323,3
Średnia	$\mu\text{g}/\text{m}^3$	6,5	6,7	5,9	3,3
Odchylenie standardowe	$\mu\text{g}/\text{m}^3$	16,8	21,0	18,4	8,4

Tabela 3
Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń NO_2 , zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	76,2	87,4	94,3	91,4
Wartość minimalna	$\mu\text{g}/\text{m}^3$	0,0	0,6	0,4	0,0
Wartość maksymalna	$\mu\text{g}/\text{m}^3$	140,7	182,5	204,1	170,1
Średnia	$\mu\text{g}/\text{m}^3$	18,9	19,8	24,3	18,1
Odchylenie standardowe	$\mu\text{g}/\text{m}^3$	14,3	14,9	17,5	13,2

Tabela 4

Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń CO₂, zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	60,1	92,6	–	97,5
Wartość minimalna	µg/m ³	3,0	36,0	–	20,7
Wartość maksymalna	µg/m ³	8620,1	12562,0	–	5794,8
Średnia	µg/m ³	606,7	511,9	–	446,8
Odchylenie standardowe	µg/m ³	415,7	460,4	–	240,9

Tabela 5

Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń SO₂, zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	21,4	84,2	93,3	98,1
Wartość minimalna	µg/m ³	0,1	0,4	0,8	0,0
Wartość maksymalna	µg/m ³	248,5	195,3	157,8	471,2
Średnia	µg/m ³	24,3	12,1	9,8	14,8
Odchylenie standardowe	µg/m ³	24,4	11,6	8,5	14,2

Tabela 6

Podstawowe parametry statystyczne serii czasowych 1-godzinnych stężeń PM₁₀, zarejestrowanych na różnych stacjach monitoringu powietrza w latach 2004-2008

Parametr	Jednostka	Piotrków Trybunalski	Radom	Warszawa -Ursynów	Łódź -Widzew
Kompletność serii	%	91,1	93,0	92,7	93,7
Wartość minimalna	µg/m ³	0,0	0,8	0,7	0,0
Wartość maksymalna	µg/m ³	470,0	860,7	414,9	264,0
Średnia	µg/m ³	25,4	40,7	32,3	22,7
Odchylenie standardowe	µg/m ³	21,8	37,2	24,7	16,3

Kompletność serii pomiarowych jest bardzo zmienna i waha się w dość szerokim zakresie, w zależności od rodzaju zanieczyszczenia i lokalizacji stacji pomiarowej. Częstość występowania brakujących danych chwilowych w 5-letnich seriach pomiarowych zmienia się w następujących zakresach:

- 2,9÷15,2% dla O₃,
- 4,9÷42,0% dla NO,
- 5,7÷23,8% dla NO₂,
- 2,5÷39,9% dla CO,
- 1,9÷78,6% dla SO₂,
- 6,3÷8,9% dla PM₁₀.

Z podanych zakresów wynika, że braki pomiarów monitoringowych występują powszechnie i nawet dla najbardziej kompletnych serii pomiarowych częstość ich

występowania wynosi kilka procent. Braki na poziomie kilkunastu czy kilkudziesięciu procent są związane z wielomiesięcznymi, a nawet wieloletnimi przerwami w wykonywaniu pomiarów.

2.2. Ocena częstości występowania brakujących pomiarów w różnych kategoriach luk pomiarowych

Dla każdego z zanieczyszczeń oddzielnie oszacowano liczbę luk pomiarowych o różnych długościach. Obliczono także częstość pojawiania się przypadków niezarejestrowanych stężeń w lukach o różnej długości. Otrzymane wyniki przedstawiono w tabelach 7-12.

Tabela 7

Liczba luk pomiarowych w zależności od ich długości dla stężeń O₃ oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	89	214	174	75	0,230	0,488	0,397	0,171
2	28	3	2	2	0,128	0,137	0,912	0,912
3	3	11	13	7	0,250	0,753	0,889	0,479
4	4	1	4	5	0,365	0,912	0,365	0,456
5	6	4	2	2	0,684	0,456	0,228	0,228
6	1	3	0	2	0,137	0,411	0	0,274
7	1	3	2	0	0,160	0,479	0,319	0
8	3	7	1	2	0,547	0,128	0,182	0,365
9	1	2	1	1	0,250	0,411	0,250	0,250
10	2	3	1	1	0,456	0,684	0,228	0,228
11	1	2	1	2	0,251	0,520	0,251	0,520
12	0	3	0	1	0	0,821	0	0,274
13	0	1	0	3	0	0,296	0	0,889
14	1	0	1	2	0,319	0	0,319	0,639
15	1	3	0	1	0,342	0,126	0	0,342
16	0	0	1	0	0	0	0,365	0
17	1	0	0	1	0,388	0	0	0,388
18	2	0	0	0	0,821	0	0	0
19	0	0	1	1	0	0	0,433	0,433
20	0	0	1	0	0	0	0,456	0
21	0	1	1	1	0	0,479	0,479	0,479
22	0	0	0	1	0	0	0	0,520
23	0	0	0	0	0	0	0	0
24	1	0	0	1	0,547	0	0	0,547
>24	1	7	1	7	2,755	4,187	14,245	1,896

Tabela 8

Liczba luk pomiarowych w zależności od ich długości dla stężeń NO oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	19	215	216	71	0,249	0,493	0,493	0,162
2	28	25	41	21	0,128	0,114	0,187	0,958
3	9	11	9	7	0,616	0,753	0,616	0,479
4	13	8	2	1	0,119	0,730	0,182	0,912
5	4	4	2	2	0,456	0,456	0,228	0,228
6	4	6	0	1	0,547	0,821	0	0,137
7	1	3	2	1	0,160	0,479	0,319	0,160
8	2	3	2	0	0,365	0,547	0,365	0
9	2	0	1	2	0,411	0	0,250	0,411
10	0	2	0	1	0	0,456	0	0,228
11	0	1	0	0	0	0,251	0	0
12	0	1	1	0	0	0,274	0,274	0
13	1	0	0	0	0,296	0	0	0
14	0	1	1	0	0	0,319	0,319	0
15	0	0	0	0	0	0	0	0
16	2	0	0	0	0,730	0	0	0
17	0	0	0	0	0	0	0	0
18	1	0	0	0	0,411	0	0	0
19	0	0	1	0	0	0	0,433	0
20	0	0	1	0	0	0	0,456	0
21	2	1	1	1	0,958	0,479	0,479	0,479
22	0	0	0	1	0	0	0	0,520
23	2	1	0	0	0,149	0,525	0	0
24	3	0	0	0	0,164	0	0	0
>24	16	3	5	7	4,775	9,284	3,813	27,338

Tabela 9

Liczba luk pomiarowych w zależności od ich długości dla stężeń NO₂ oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	145	193	181	8	0,337	0,442	0,413	0,182
2	32	38	39	21	0,146	0,173	0,178	0,958
3	11	22	1	1	0,753	0,155	0,684	0,684
4	18	19	2	12	0,164	0,173	0,182	0,195
5	3	8	8	3	0,342	0,912	0,912	0,342
6	4	13	4	3	0,547	0,178	0,547	0,411
7	1	11	5	0	0,160	0,176	0,798	0
8	3	5	1	0	0,547	0,912	0,182	0
9	2	2	2	3	0,411	0,411	0,411	0,616
10	1	2	2	1	0,228	0,456	0,456	0,228
11	0	2	2	1	0	0,520	0,520	0,251
12	0	2	1	0	0	0,547	0,274	0
13	1	1	3	0	0,296	0,296	0,889	0
14	0	1	1	0	0	0,319	0,319	0
15	0	2	2	0	0	0,684	0,684	0
16	2	2	2	0	0,730	0,730	0,730	0
17	0	2	2	0	0	0,775	0,775	0
18	2	1	0	1	0,821	0,411	0	0,411
19	1	1	2	0	0,433	0,433	0,867	0
20	0	0	2	0	0	0	0,912	0
21	4	1	2	2	0,192	0,479	0,958	0,958
22	0	0	0	0	0	0	0	0
23	2	1	0	0	0,149	0,525	0	0
24	4	0	0	1	0,219	0	0	0,547
>24	24	6	7	8	22,717	1,473	3,973	7,724

Tabela 10

Liczba luk pomiarowych w zależności od ich długości dla stężeń CO oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	97	175	-	1	0,221	0,399	-	0,228
2	26	25	-	11	0,119	0,114	-	0,520
3	2	13	-	14	0,137	0,889	-	0,958
4	1	5	-	3	0,910	0,456	-	0,274
5	3	6	-	2	0,342	0,684	-	0,228
6	1	6	-	1	0,137	0,821	-	0,137
7	1	1	-	2	0,160	0,160	-	0,319
8	1	1	-	0	0,182	0,182	-	0
9	2	0	-	0	0,411	0	-	0
10	2	1	-	0	0,456	0,228	-	0
11	1	1	-	0	0,251	0,251	-	0
12	1	0	-	0	0,274	0	-	0
13	0	1	-	0	0	0,296	-	0
14	1	1	-	1	0,319	0,319	-	0,319
15	1	1	-	0	0,342	0,342	-	0
16	2	1	-	0	0,730	0,365	-	0
17	1	0	-	0	0,388	0	-	0
18	1	0	-	0	0,411	0	-	0
19	1	0	-	0	0,433	0	-	0
20	0	0	-	0	0	0	-	0
21	0	1	-	0	0	0,479	-	0
22	0	1	-	1	0	0,520	-	0,520
23	1	1	-	0	0,525	0,525	-	0
24	0	1	-	0	0	0,547	-	0
>24	18	17	-	8	38,966	6,196	-	1,959

Tabela 11

Liczba luk pomiarowych w zależności od ich długości dla stężeń SO₂ oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	37	39	25	19	0,844	0,747	0,468	0,249
2	12	52	25	29	0,547	0,237	0,114	0,132
3	4	13	1	6	0,274	0,889	0,684	0,411
4	5	18	3	6	0,456	0,164	0,274	0,547
5	5	7	1	2	0,570	0,798	0,114	0,228
6	1	6	1	1	0,137	0,821	0,137	0,137
7	2	0	1	0	0,319	0	0,160	0
8	0	4	2	1	0	0,730	0,365	0,182
9	0	0	1	0	0	0	0,250	0
10	0	2	1	0	0	0,456	0,228	0
11	2	0	0	0	0,520	0	0	0
12	0	1	0	0	0	0,274	0	0
13	0	1	0	1	0	0,296	0	0,296
14	0	0	1	0	0	0	0,319	0
15	0	1	0	0	0	0,342	0	0
16	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	1	0	0	1	0,411	0	0	0,411
19	0	0	1	0	0	0	0,433	0
20	0	0	1	0	0	0	0,456	0
21	0	1	1	1	0	0,479	0,479	0,479
22	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	2	0	0	0	0,195	0	0
>24	5	11	4	5	78,172	14,417	5,777	1,242

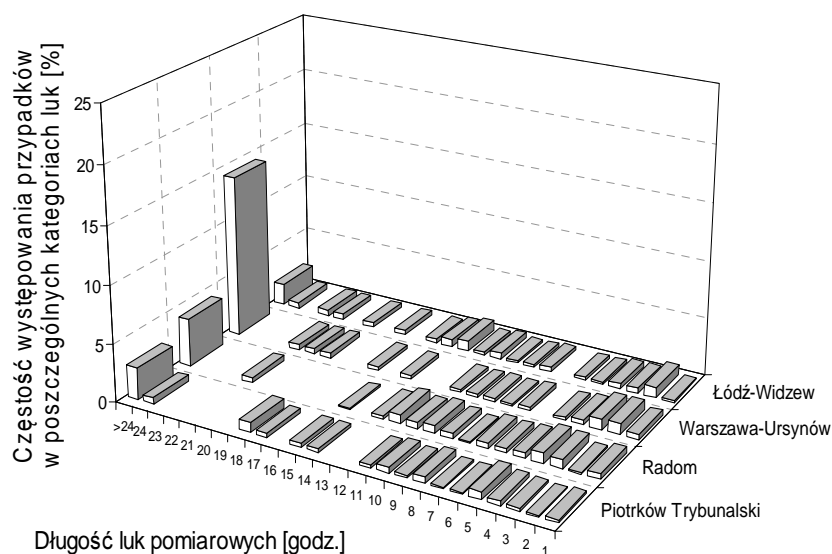
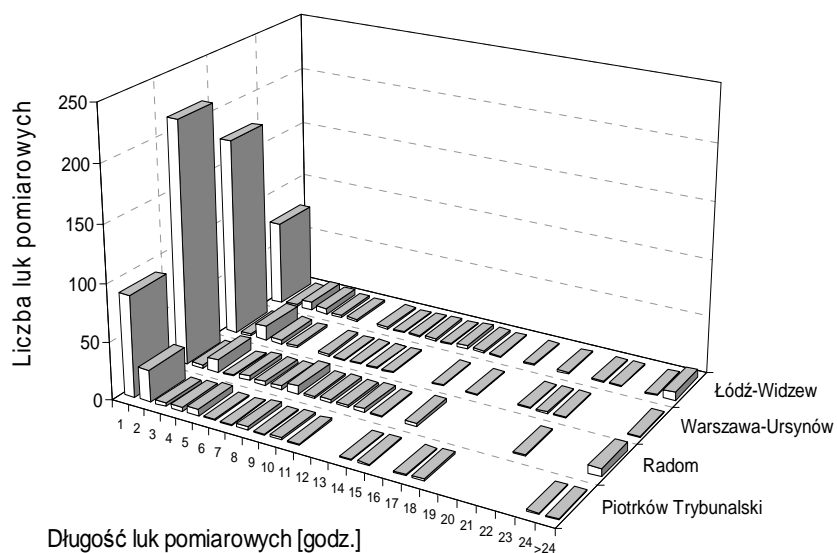
Tabela 12

Liczba luk pomiarowych w zależności od ich długości dla stężeń PM10 oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

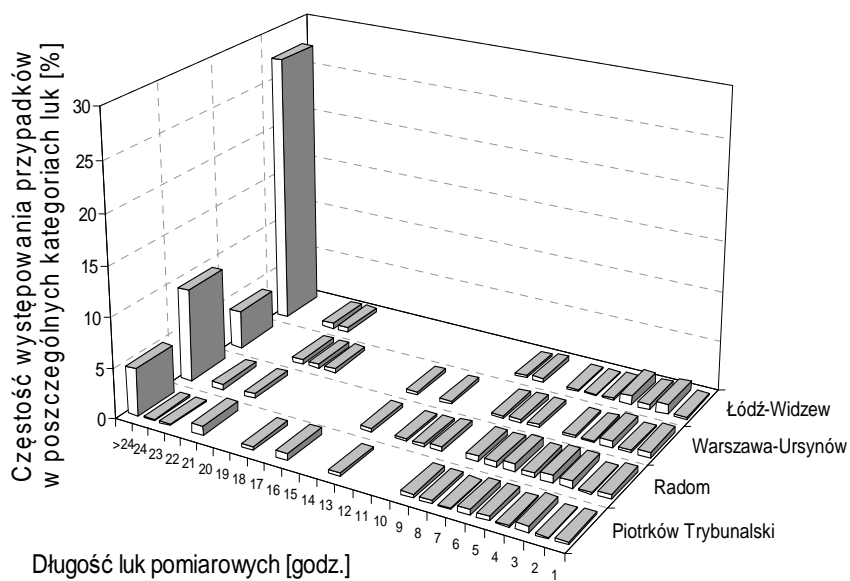
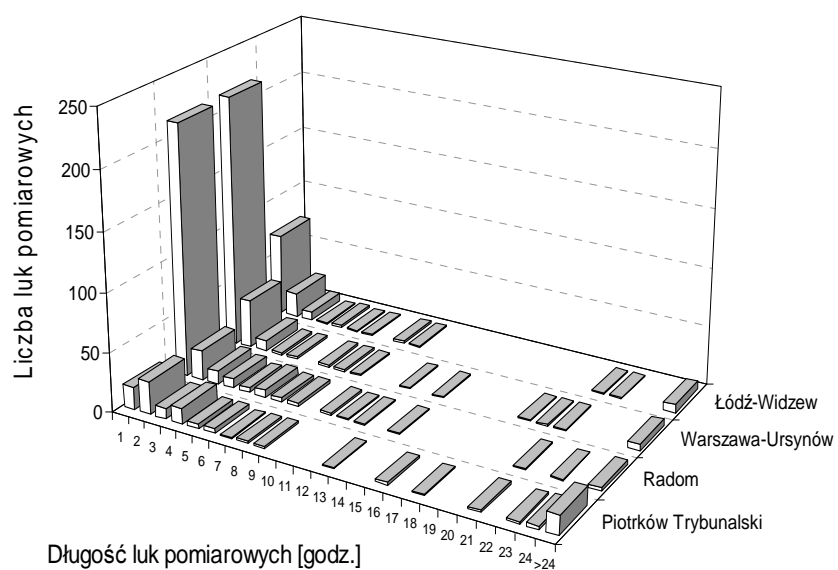
Długość luk pomiarowych	Liczba luk pomiarowych				Procent przypadków występujących w lukach pomiarowych			
	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew	Piotrków Trybunalski	Radom	Warszawa - Ursynów	Łódź - Widzew
godz.	-				%			
1	497	183	711	342	1,134	0,417	1,622	0,780
2	135	38	132	64	0,616	0,173	0,621	0,292
3	39	2	44	2	0,267	0,137	0,310	0,137
4	21	1	15	7	0,192	0,912	0,137	0,639
5	11	4	5	4	0,125	0,456	0,570	0,456
6	5	4	4	1	0,684	0,547	0,547	0,137
7	7	5	0	5	0,112	0,798	0	0,798
8	4	2	1	0	0,730	0,365	0,182	0
9	1	4	3	0	0,250	0,821	0,616	0
10	2	8	2	0	0,456	0,182	0,456	0
11	1	9	0	0	0,251	0,226	0	0
12	1	6	0	0	0,274	0,164	0	0
13	1	6	0	0	0,296	0,178	0	0
14	0	2	0	0	0	0,639	0	0
15	0	1	2	0	0	0,342	0,684	0
16	0	3	0	0	0	0,195	0	0
17	0	4	0	0	0	0,155	0	0
18	0	1	1	0	0	0,411	0,411	0
19	1	0	0	0	0,433	0	0	0
20	0	2	0	0	0	0,912	0	0
21	0	1	0	0	0	0,479	0	0
22	0	2	1	0	0	0,130	0,520	0
23	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0
>24	9	1	5	6	6,133	4,486	4,277	4,849

Wykresy ilustrujące liczebność luk pomiarowych w zależności od ich długości przedstawiono na rysunkach 2-7, oddzielnie dla każdego z zanieczyszczeń. Dla wszystkich zanieczyszczeń najczęściej występującymi lukami pomiarowymi są pojedyncze przypadki. Liczba luk bardzo szybko maleje z rosnącą ich długością (czasem trwania). Bloki brakujących danych o długościach przekraczających 3-4 przypadki (3-4 godziny) występują stosunkowo rzadko.

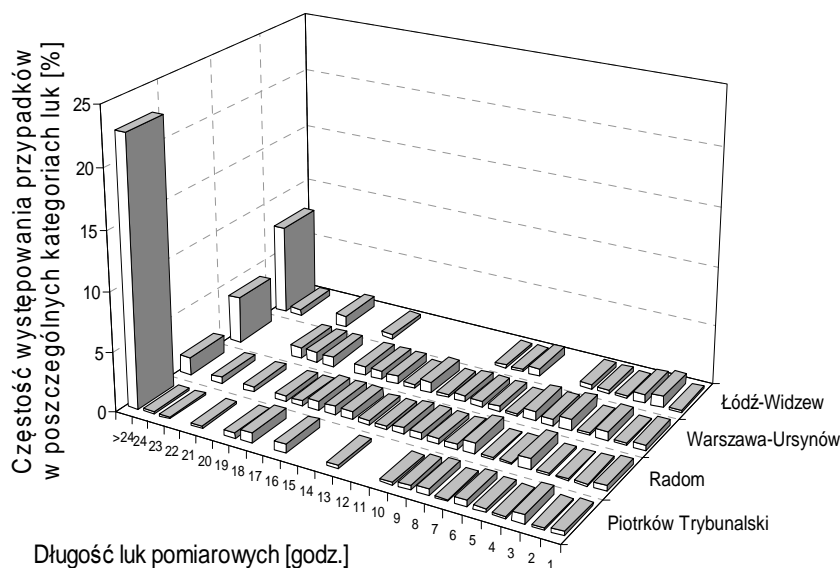
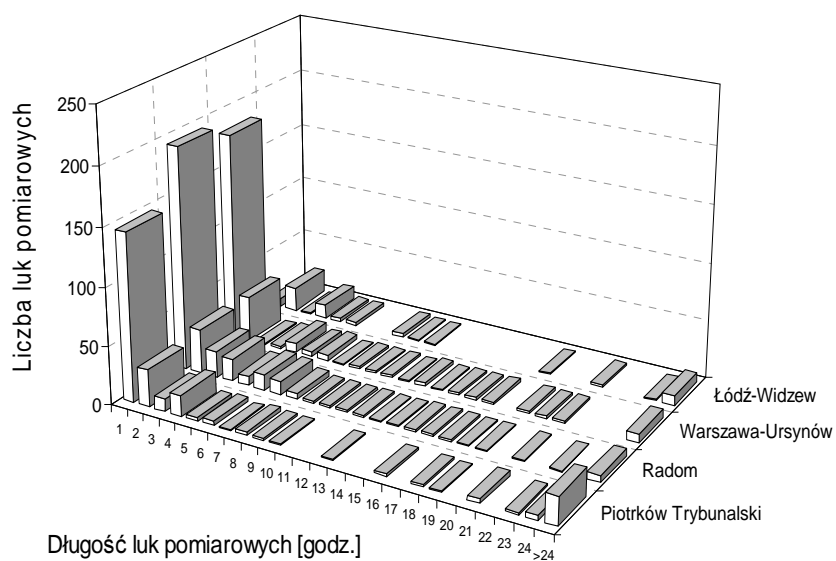
Na rysunkach 2-7 przedstawiono również dodatkowe wykresy prezentujące częstość występowania przypadków w poszczególnych typach luk pomiarowych. Dla wszystkich zanieczyszczeń najwyższą częstość odnotowano dla luk najdłuższych, obejmujących więcej niż 24 przypadki (>24 godziny).



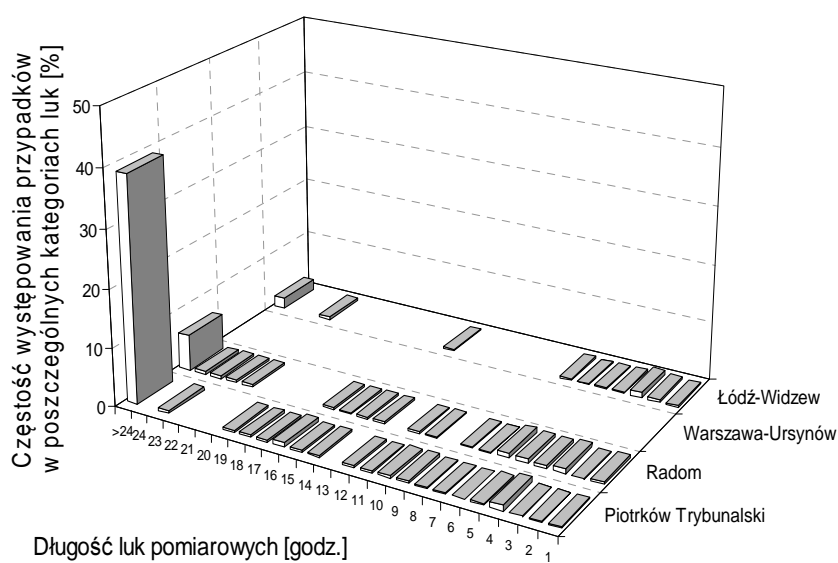
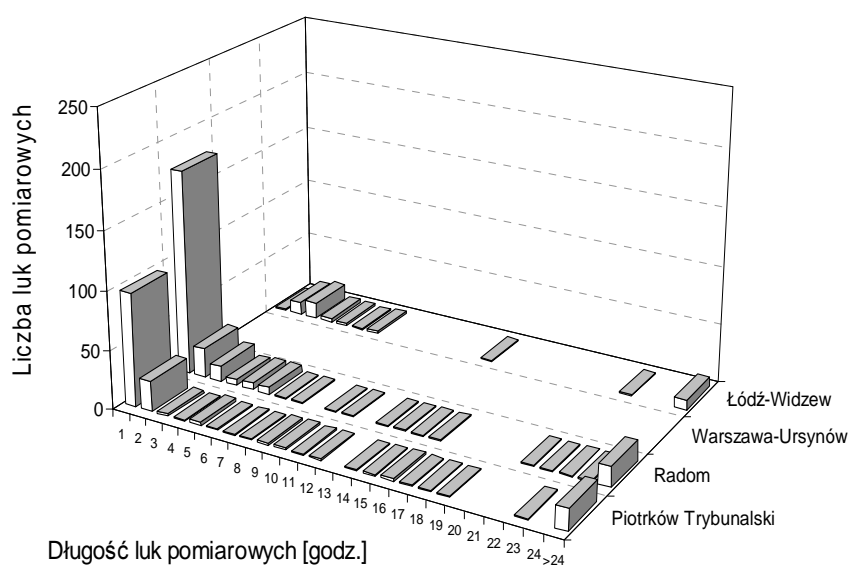
Rys. 2. Liczba luk pomiarowych w zależności od ich długości dla stężeń O_3 oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk



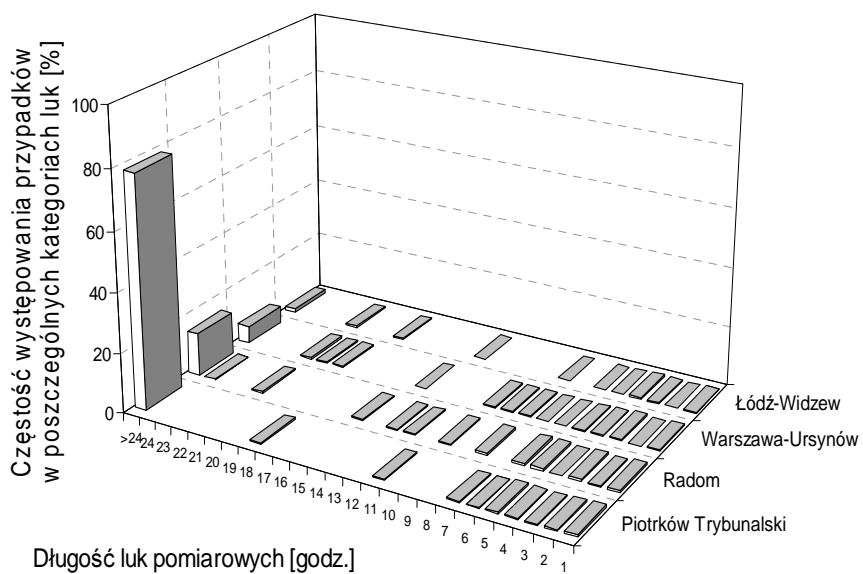
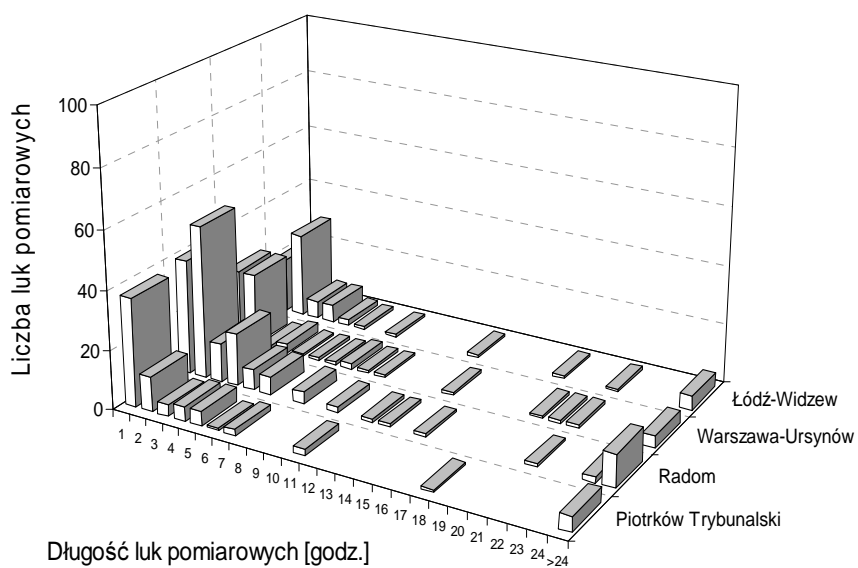
Rys. 3. Liczba luk pomiarowych w zależności od ich długości dla stężeń NO oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk



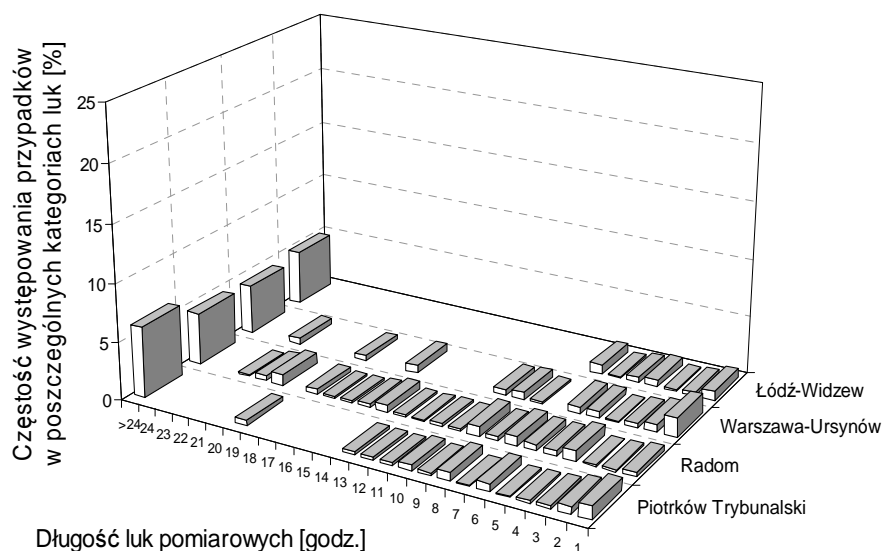
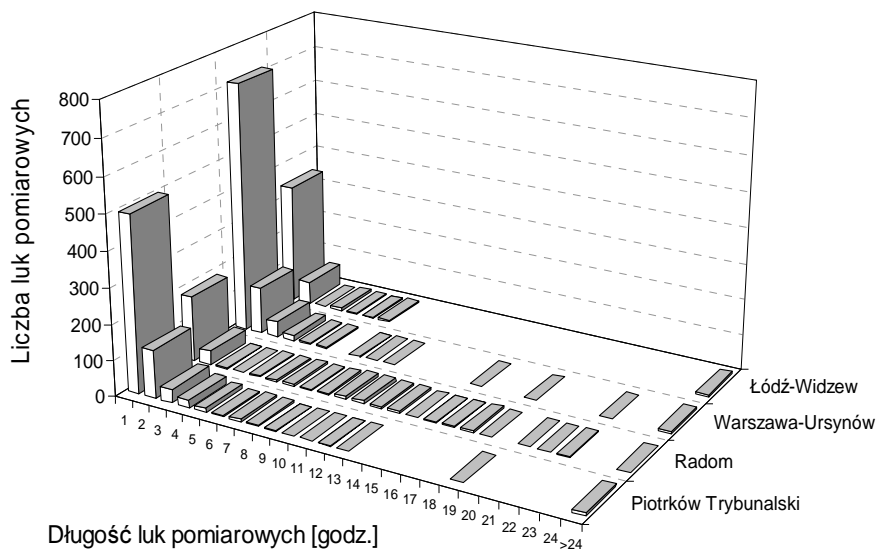
Rys. 4. Liczba luk pomiarowych w zależności od ich długości dla stężeń NO_2 oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk



Rys. 5. Liczba luk pomiarowych w zależności od ich długości dla stężeń CO oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk



Rys. 6. Liczba luk pomiarowych w zależności od ich długości dla stężeń SO₂ oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk



Rys. 7. Liczba luk pomiarowych w zależności od ich długości dla stężeń PM_{10} oraz częstość (w %) występowania przypadków w poszczególnych kategoriach luk

Podsumowanie

Na podstawie wyników przeprowadzonej analizy można stwierdzić, że:

1. Brakujące dane występują powszechnie w zbiorach danych pochodzących z monitoringu powietrza. Częstość ich występowania w rocznych seriach pomiarowych może wynosić od kilku do nawet kilkudziesięciu procent.

- Większość luk pomiarowych jest krótka – stanowią je głównie pojedyncze przypadki. Zdecydowanie rzadziej występują bloki brakujących danych, przekraczające 3-4 przypadki (dłuższe od 3-4 godzin).
- Największą częstość występowania przypadków z niezarejestrowanymi wynikami odnotowano dla luk najdłuższych, obejmujących więcej niż 24 przypadki (24 godziny).

Praca naukowa finansowana ze środków na naukę w latach 2006-2008 jako projekt badawczy nr 1 T09D 037 30.

Literatura

- [1] Hauck H., Kromp-Kolb H., Petz E., Requirements for the completeness of ambient air quality data sets with respect to derived parameters, *Atmos. Environ.* 1999, 33, 2059-2066.
- [2] Hoffman S., Ocena kompletności serii pomiarowych zarejestrowanych na wybranych stacjach monitoringu powietrza, *Materiały Konferencji Mikrozanieczyszczenia w środowisku człowieka, Częstochowa, 25-27 czerwca 2003*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2003, 493-495.
- [3] Rozporządzenie Ministra Środowiska z dnia 6 czerwca 2002 r. w sprawie oceny poziomów substancji w powietrzu, *DzU Nr 87, poz. 798*.

Classification of Air Monitoring Data Gaps

The data gathered continuously in the air monitoring systems are never entire. In the whole year, the number of missing records is changeable. The deficiency of data could result in uncertainty of a statistical assessment, required by the air quality standards, and cause the uselessness of monitoring measurements. Air quality standards permit to use modelling in order to recreate the missing data when the completeness of the monitoring set is not sufficient. Applied modelling methods should guarantee possibly the best precision to achieve the air quality assessment being closest to reality. Single, specified method does not assure the maximal accuracy because the missing data in data matrix may create gaps of various shapes and ranges. Recognition of typical structures of missing data fields should be the base of their classification. For the specified classes of gaps the optimum modelling methods may be recommended and assigned.

The main objective of the analysis was to select typical patterns of gaps in air monitoring data matrixes, and the assessment of their appearing. The missing data classification was suggested after long-term data survey. The analyzed data sets derived from 4 different air monitoring sites in the Central Poland (Warsaw-Ursynów, Radom, Lodz-Widzew, Piotrków Trybunalski). The data were gathered in the period 2004-2008. The examined time-series involved hourly concentrations of main air pollutants: O₃, NO₂, NO, PM₁₀, SO₂, CO.

The results allow coming to some general conclusions. Missing data commonly occur in sets of air monitoring records. Gaps may include up to several or even more per cent of all expected data in yearly measuring series. For all air pollutants, the most of the gaps in monitoring time series are very short. Single (1-hour) missing values dominate among gaps of different length. Gaps lengths exceeding 3-4 hours are observed occasionally. However, the greatest frequency of single without-data cases appearing is observed in the longest gaps (>24 hours), because of their lengths.

Keywords: air pollution, air monitoring, hourly concentrations, monitoring data, missing data, measure gaps, classification