

Politechnika Wrocławska
Instytut Inżynierii Ochrony Środowiska

Neuronowe modele prognostyczne i neuronowa metoda interpolacji stężeń imisyjnych

mgr inż. Monika Maciejewska

Promotor: prof. dr hab. inż. Jerzy Zwoździak

Wrocław, 2001

dedykuję Ojcu

Spis treści

Cel i zakres pracy	i
1. Wstęp	1
1.1. Wykorzystanie metod obliczeniowych do opisu imisji	3
1.2. Metody neuronowe w prognostyce stężeń zanieczyszczeń	5
2. Przegląd metod obliczeniowych stosowanych w opisie imisji	12
2.1. Metody analityczne	12
2.1.1. Podstawy metodologii interpolacji	12
2.1.2. Interpolacja metodą średniej ważonej	13
2.1.3. Interpolacja funkcyjna	14
2.2. Metody statystyczne	16
2.2.1. Analiza czynnikowa	16
2.2.2. Regresja	19
2.2.3. Modele stochastyczne ARIMA	22
2.2.4. Modele stochastyczne Funkcji Transferu	24
2.2.5. Model przestrzeni stanów z filtrem Kalmana	24
2.3. Metody neuronowe	27
2.3.1. Liniowy filtr adaptacyjny, sieć z liniową funkcją transferu i sieć z nieliniową funkcją transferu	30
2.3.2. Sieci z radialną funkcją transferu	31
2.3.3. Sieci samoorganizujące	33
3. Badania własne	35
3.1. Charakterystyka obszaru źródłowego danych	35
3.1.1. Zakładowa Sieć Monitoringu Powietrza Atmosferycznego Elektrowni Turów	37
3.2. Kryteria oceny modelu i wyników modelowania	39
3.3. Prognoza stężeń imisyjnych	42
3.3.1. Analiza przygotowawcza	42
3.3.2. Prognoza z wykorzystaniem neuronowego modelu deterministycznego	53
3.3.2.1. Propozycja rozwiązania problemu	53
3.3.2.2. Dyskusja wyników neuronowego prognozowania deterministycznego	58
3.3.3. Prognoza z wykorzystaniem modelu niedeterministycznego	72
3.3.3.1. Propozycja rozwiązania problemu	72
3.3.3.2. Dyskusja wyników neuronowego prognozowania niedeterministycznego	74
3.4. Interpolacja przestrzenna stężeń imisyjnych	83
3.4.1. Analiza przygotowawcza	83
3.4.2. Propozycja rozwiązania	90
3.4.3. Dyskusja wyników neuronowej interpolacji stężeń	93
4. Podsumowanie i wnioski	109
Spis literatury	112

Cel i zakres pracy

Zmienność stężeń imisyjnych w czasie i przestrzeni można opisywać na podstawie danych z sieci ciągłego monitoringu zanieczyszczeń powietrza stosując odpowiednie metody obliczeniowe. Do tradycyjnych należą metody analityczne i statystyczne. Możliwe są też rozwiązania oparte na nowej, uniwersalnej i bardzo efektywnej metodzie obróbki i analizy danych pomiarowych, którą stanowią sieci neuronowe.

Szereg czynników wpływających na czasową zmienność stężeń zanieczyszczeń jest wykorzystywanych do jej opisu. Należą tu np. warunki meteorologiczne, emisja czy wartości stężeń występujących wcześniej. Mimo posługiwania się tymi parametrami, ocena wielkości i charakteru ich oddziaływania na stężenia zanieczyszczeń, za pomocą metod tradycyjnych, jest utrudniona. Powodem są problemy metodologiczne.

Liniowy bądź nieliniowy charakter zależności stężeń zanieczyszczeń od warunków meteorologicznych i stężeń występujących wcześniej można wykazać posługując się neuronowymi modelami prognostycznymi stężeń zanieczyszczeń. Znajdują tu zastosowanie sieci neuronowe z liniowymi i nieliniowymi funkcjami transferu. Modele te, oparte na podejściu deterministycznym i niedeterministycznym umożliwiają również ocenę stopnia uzależnienia stężeń od warunków meteorologicznych oraz stężeń występujących wcześniej w serii czasowej.

Innym zagadnieniem jest identyfikacja źródeł emisji zanieczyszczeń. W warunkach równoległej rejestracji stężeń co najmniej dwóch zanieczyszczeń w punkcie pomiarowym, można do tego celu wykorzystać neuronowy model deterministyczny.

Stężenia zanieczyszczeń wykazują również zmienność przestrzenną. W związku z tym, że poprawnie wygenerowany przestrzenny rozkład stężeń zanieczyszczeń niesie bardzo dużo informacji o stanie zanieczyszczenia powietrza, istotnym jest stosowanie właściwej metody interpolacji.

Wybór spośród tradycyjnych metod interpolacji wymaga a priori znajomości cech pola stężeń, co w ogólnym przypadku jest niemożliwe. Na identyfikację własności pola stężeń pozwala neuronowa metoda interpolacji opracowana w oparciu o własności sieci radialnej. Dzięki temu metoda prowadzi do uzyskania rozkładu przestrzennego stężeń możliwie najbliższego rzeczywistości.

Zależnie od skali oddziaływania źródeł zanieczyszczeń, pole stężeń zanieczyszczeń może cechować jednorodność lub prawidłowości o lokalnym zasięgu. Neuronowa metoda interpolacji pozwala wskazać czynniki o lokalnej lub regionalnej skali oddziaływania na przestrzenny rozkład stężeń zanieczyszczeń.

Celem pracy było opracowanie neuronowych modeli prognostycznych i neuronowej metody interpolacji przestrzennej stężeń imisyjnych dla pozyskania informacji o uwarunkowaniach czasowej i przestrzennej zmienności stężeń zanieczyszczeń w powietrzu.

W pracy wykorzystano dane pomiarowe pochodzące z Zakładowej Sieci Monitoringu Zanieczyszczeń Powietrza Elektrowni "Turów" S.A. Dla osiągnięcia sformułowanego celu pracy wykonano następujące prace.

- Przeprowadzono wstępną analizę statystyczną:
 - szeregów czasowych stężeń, z wykorzystaniem współczynników autokorelacji i autokorelacji cząstkowej,
 - współzależności stężeń zanieczyszczeń i parametrów meteorologicznych, stosując współczynniki korelacji wzajemnej,
 - współzależności stężeń rejestrowanych równolegle w różnych punktach pomiarowych, za pośrednictwem współczynnika korelacji, analizy czynników głównych (PCA) i wskaźników zbudowanych przez autora.
- Opracowano deterministyczny model prognostyczny na bazie radialnej sieci neuronowej, która odwzorowuje zbliżone zestawy warunków meteorologicznych w zbliżone wartości stężeń zanieczyszczeń z uwzględnieniem nieliniowości odwzorowania.
- Zbudowano niedeterministyczne modele prognostyczne stosując sieci neuronowe z liniowymi i nieliniowymi funkcjami przejścia. Sprawdzono przydatność dla celów prognozy zarówno liniowych jak nieliniowych prawidłowości ujawniających się w seriach czasowych stężeń zanieczyszczeń.
- Przeprowadzono ocenę jakości neuronowych modeli prognostycznych i prognoz oraz sformułowano wnioski na bazie wyników obu podejść prognostycznych deterministycznego i niedeterministycznego.
- Opracowano metodę interpolacji przestrzennej wykorzystującą radialną sieć neuronową, która przekształca współrzędne punktów na wartości stężeń zanieczyszczeń w tych punktach.
- Przeprowadzono identyfikację własności pola stężeń zanieczyszczeń w Worku Żytawskim wskazując czynniki wpływające na przestrzenną zmienność stężeń.
- Podsumowano rezultaty całości badań wskazując możliwy kierunek dalszych prac.

W pracy zaproponowano opis imisji z wykorzystaniem metodyki neuronowej. Podstawowe zalety tego podejścia stanowią uniwersalność, efektywność i gwarancja poprawności metodologicznej modeli.

Należy jednak podkreślić, że skupiono uwagę na problemach czasowej i przestrzennej zmienności stężeń imisyjnych oraz możliwościach ich rozwiązań z wykorzystaniem sieci neuronowych. Demonstracja możliwości narzędzia została im podporządkowana i ma charakter drugoplanowy.

1. Wstęp

Podstawowym zadaniem badaczy zajmujących się ochroną atmosfery jest pozyskiwanie informacji o stanie zanieczyszczenia powietrza atmosferycznego. Istnieją trzy grupy metod, które to umożliwiają:

- metody sensoryczne,
- metody pomiarowe,
- metody obliczeniowe.

Najbardziej pierwotny charakter mają metody sensoryczne, czyli za pośrednictwem zmysłów. Przy pomocy zmysłów człowiek jest w stanie stwierdzić obecność niektórych zanieczyszczeń w powietrzu, a czasem nawet ocenić ich ilość, zwykle w skali proporcjonalnej. Dotyczy to np. związków zapachowych (zmysł węchu), zapylenia (zmysł wzroku) czy hałasu (zmysł słuchu).

Ze względu na ograniczoną ilość bodźców rejestrowanych drogą sensoryczną, problem poziomu czułości zmysłów oraz ograniczone możliwości wykonywania ocen ilościowych z ich pomocą, człowiek korzysta z drogi pośredniej - pomiaru. Posługuje się w tym celu różnorodnymi metodami i technikami pomiarowymi, ciągle je doskonaląc i opracowując nowe. Pomiaru są obecnie podstawowym sposobem pozyskiwania informacji o jakości powietrza atmosferycznego, mimo kosztów związanych z ich wykonywaniem.

Pozycję komplementarną w stosunku do metod sensorycznych i pomiarowych zajmują metody obliczeniowe. Korzystanie z nich umożliwia pozyskanie informacji niedostępnej drogą pomiarową. Dotyczy to np. stężeń zanieczyszczeń w punktach, gdzie nie można zainstalować urządzeń pomiarowych. W takich sytuacjach odpowiednie metody obliczeniowe umożliwiają interpolację, a nawet ekstrapolację wyników pochodzących z funkcjonujących punktów pomiarowych na obszar pozbawiony pomiaru.

Niedostępna na drodze pomiaru jest również informacja o przyszłych wartościach stężeń zanieczyszczeń. Jej pozyskiwanie stanowi przedmiot prognostyki, zdominowanej przez metody obliczeniowe. Prognostyka to obszar niezwykle istotny z punktu widzenia zarządzania jakością powietrza.

Posługiwanie się metodami obliczeniowymi umożliwia też poważne oszczędności nakładów ponoszonych na utrzymanie sieci punktów pomiarowych, zwłaszcza sieci monitoringu automatycznego. Istnieją metody obliczeniowe umożliwiające ocenę podobieństwa informacji pochodzącej z różnych punktów pomiarowych. Z powodzeniem wspomagają one rozwiązywanie

problemów redukcji liczby stacji pomiarowych i optymalizacji ich rozmieszczenia.

Metody obliczeniowe muszą bazować na informacji wejściowej. Najczęściej pochodzi ona z danych pomiarowych. Jej źródło mogą też stanowić wielkości wyliczone na podstawie technicznych parametrów źródła np. emisja zanieczyszczeń - jak w przypadku modelu rozprzestrzeniania zanieczyszczeń.

W sposób najbardziej bezpośredni stan zanieczyszczenia powietrza charakteryzują pomiary imisyjne, czyli pomiary stężeń zanieczyszczeń w powietrzu zewnętrznym. Do obróbki i analizy tych danych wykorzystuje się najwięcej różnorodnych metod obliczeniowych. Dominujące źródło danych imisyjnych stanowią sieci monitoringu powietrza atmosferycznego. Tego typu dane stanowią podstawę opracowanych i prezentowanych w pracy neuronowych modeli prognostycznych i neuronowej metody interpolacji.

Wyniki pomiaru imisji są danymi liczbowymi. Konsekwencją tego jest możliwość wykorzystywania wszystkich istniejących metod analizy i obróbki danych pomiarowych do pozyskiwania informacji zawartej w wynikach pomiaru stężeń zanieczyszczeń. Paleta dostępnych metod jest bardzo bogata i ciągle się powiększa, a wybór właściwej metody zależy od rodzaju poszukiwanej informacji. Wynika stąd wprost, że obliczeniowe metody pozyskiwania informacji o stanie zanieczyszczenia powietrza atmosferycznego bazujące na informacji pomiarowej są celowymi adaptacjami metod obróbki i analizy danych pomiarowych. Przegląd literatury poświęconej opisywaniu i prognozowaniu jakości powietrza na podstawie pomiarów imisyjnych ujawnia różnorodność stosowanych tu metod oraz nadążanie za rozwojem w dziedzinie metod obróbki i analizy danych liczbowych. Zaproponowano następującą klasyfikację metod obliczeniowych na podstawie kryterium stanu wiedzy o rozwiązywanym problemie:

- metody analityczne,
- metody statystyczne,
- metody neuronowe.

Modelowanie analityczne wiąże się z opisem zmiennych za pomocą równań matematycznych. Do tego celu konieczna jest znajomość dokładnej postaci zależności zmiennej opisywanej od zmiennych ją opisujących. Postać związku wynika z wiedzy o mechanizmach odpowiedzialnych za zachowanie zmiennej opisywanej.

Modelowanie statystyczne nie wymaga znajomości formy związku zmiennej zależnej i zmiennych niezależnych. Umożliwia natomiast stawianie i testowanie hipotez dotyczących istnienia tych związków oraz ich ogólnej postaci. Statystycznie istotne zależności są wykorzystywane do konstruowania modeli statystycznych.

Jeżeli wiedza o uwarunkowaniach zmiennej zależnej jest niedostępna lub narzędzia statystyczne są zbyt słabe by zidentyfikować postać tych uwarunkowań warto zaimplementować podejście neuronowe. Nie prowadzi ono do uzyskania żadnej czytelnej formy opisu zmiennej, lecz niejawnej reprezentacji tego opisu w strukturze sieci neuronowej. Ten rodzaj modelowania cechuje bardzo wysoka efektywność przy ograniczonych możliwościach rozpoznania mechanizmu rządzącego zachowaniem zmiennej opisywanej.

Propozycje analitycznych rozwiązań w zakresie opisu imisji istnieją w zasadzie wyłącznie w odniesieniu do zagadnień interpolacji przestrzennej. Rzadkość rozwiązań analitycznych w sferze prognozy stężeń zanieczyszczeń wynika z braku ogólnej teorii zmienności czasowej stężeń imisyjnych dającej się sformalizować w postaci analitycznego zapisu matematycznego. Brak teorii wymusza stawianie fragmentarycznych hipotez dotyczących charakteru zmienności stężeń i wpływu różnych czynników na tę zmienność. Są to cechy metodologii statystycznej. Właśnie metody statystyczne są najczęściej wykorzystywane do analizy danych pochodzących ze stacji monitoringowych. W dziedzinie analiz czasowych na czoło wysuwają się regresja wieloczynnikowa, zintegrowane modele autoregresji i średniej ruchomej (ARIMA) oraz modelowanie w przestrzeni stanów (state-space models). W zakresie analiz przestrzennych na uwagę zasługują metody analizy czynnikowej (PCA) i kriging.

1.1. Wykorzystanie metod obliczeniowych do opisu imisji

Literatura z zakresu obliczeniowych metod analizy danych imisyjnych pokazuje, że głównym obszarem zainteresowania badaczy, jest jakość powietrza na terenach zurbanizowanych. Wysiłek opracowywania metod progностycznych, jak również interpolacyjnych w odniesieniu do stężeń zanieczyszczeń w powietrzu atmosferycznym dotyczy przede wszystkim powietrza miejskiego. Jest to uzasadnione. Ze względu na koncentrację źródeł zanieczyszczeń oraz dużą gęstość zaludnienia, w aglomeracjach miejskich występuje problem podwyższonych stężeń zanieczyszczeń i dużej ekspozycji ludności na zanieczyszczenia powietrza. Wywołuje to konieczność kontroli stężeń zanieczyszczeń oraz profilaktyki, zwłaszcza w stosunku do epizodów stężeń. Najwięcej uwagi poświęca się prognozowaniu stężeń ozonu, tlenu i dwutlenku azotu w miastach oraz masach powietrza napływających do miast i opuszczających miasta. Konstruowaniu modeli przyświeca uzyskiwanie jak najdokładniejszych prognoz krótkoterminowych stężeń o różnym stopniu uśredniania: 30-minutowym, godzinnym, dobowym. Budowie modeli progностycznych towarzyszą próby interpretacji zależności stężeń zanieczyszczeń

od czynników wykorzystywanych w modelach jako zmienne objaśniające. W ostatnich latach obok rozwiązań bazujących na metodach statystycznych [Chen,1998; Chock,1974; Jorquera,2000; Kartal,1998; Milionis,1994; Unal,2000; Ping Shi,1997; Ziomas,1995] literatura fachowa dostarcza przykładów stosowania modeli neuronowych [Comrie,1997; Gardner,1998; Gardner,2000; Guardani,1999; Baran,1999; Jorquera,1998; Kao,2000; Kolehmainen,2001; Melas,2000; Ośródk,1995; Perez,2000; Soja,2000; Yu,2000] często prezentowanych równolegle z rozwiązaniami tradycyjnymi w celach porównawczych.

Część analiz poświęconych jakości powietrza w aglomeracjach miejskich, choć nie tylko, dotyczy poszukiwania metod interpolacji przestrzennej stężeń. Wynika ono z potrzeby optymalizacji liczby i rozmieszczenia stacji monitoringowych. Problem polega na tym, jak uzyskać faktyczny obraz jakości powietrza na obszarze rozmieszczając na nim jak najmniej pomiarów punktowych. Bazę dla opracowywania i testowania nowych metod interpolacji przestrzennej stanowią najczęściej dane pochodzące z sieci monitoringu powietrza atmosferycznego [Chu,1998; Egmond,1980; Elsom,1978; Fedorov,1989; Hopkins,1999; Langstaff,1986; Nakamori,1978; Nakamori,1984; Noll,1977]. W zakresie metod interpolacji przestrzennej brak doniesień literaturowych o propozycjach rozwiązań neuronowych.

Drugi główny nurt opisu imisji dotyczy problemów rozpatrywanych w wielkiej skali tj. w skali regionu, państwa, a nawet kontynentu. Przedmiotem zainteresowania są głównie ozon, dwutlenek siarki i pył zawieszony. Opis zmienności stężeń tych zanieczyszczeń ma najczęściej charakter deterministyczny. Celem opisu jest identyfikacja oraz analiza zależności stężeń zanieczyszczeń od wielkoskalowych prawidłowości meteorologicznych i przestrzennych wzorców emisji. W badaniach wykorzystywane są głównie dane z sieci monitoringu regionalnego oraz podstawowego, choć część opracowań dotyczy danych pochodzących z okresowych kampanii pomiarowych, przeprowadzonych dla zweryfikowania hipotez badawczych [Christakos,1998; Kamiński,1999; Myrup,1983; Statheropoulos,1998; van der Wal,2000].

Zwraca uwagę bardzo mała liczba publikacji poświęconych opisowi zmienności stężeń zanieczyszczeń wokół konkretnych źródeł emisji np. dużych emitorów punktowych, wyłącznie na podstawie pomiarów z sieci monitorujących ich oddziaływanie na jakość powietrza. Dotyczy to w równym stopniu zagadnień prognostycznych, co interpolacji przestrzennej. Stacje monitoringowe rozmieszczone wokół dużych źródeł emisji dostarczają wartościowego materiału badawczego, który umożliwia analizę uwarunkowań czasowej i przestrzennej zmienności stężeń zanieczyszczeń powietrza wokół źródła. Analizy te mogą prowadzić do nieoczekiwanych i interesujących wniosków dotyczących wpływu źródła emisji i innych czynników na stężenia zanieczyszczeń wokół niego.

Wiele różnych metod obliczeniowych jest wykorzystywanych do opisu imisji. Zarówno w dziedzinie prognozy jak interpolacji, duży potencjał poprawy rezultatów uzyskiwanych w wyniku stosowania metod tradycyjnych tzn. analitycznych i statystycznych wiąże się z wykorzystaniem sieci neuronowych. Praca prezentuje propozycje neuronowych rozwiązań w zakresie prognozy i interpolacji stężeń imisyjnych.

1.2. Metody neuronowe w prognozie stężeń zanieczyszczeń

Sieci neuronowe są nową metodą obliczeniową. Stosuje się je bardzo często do rozwiązywania problemów, dla których zaproponowano już rozwiązania z wykorzystaniem metod opracowanych wcześniej np. analitycznych i statystycznych. Szereg nowych, powstających zagadnień badawczych podlega nadal analizie za pomocą tych metod. Dotyczy to również problemów, których rozwiązania neuronowe przedstawiono w tej pracy. Zastanowienia wymaga więc celowość poświęcania uwagi sieciom neuronowym i rozwijania tej grupy metod obliczeniowych. Poniżej przedstawiono argumenty przemawiające za stosowaniem sieci neuronowych na tle problematyki modelowania stężeń zanieczyszczeń w powietrzu.

1. Możliwość poprawy rezultatów w stosunku do uzyskanych z wykorzystaniem innych metod

Elementem rozstrzygającym jest w tym przypadku zdolność sieci do modelowania zależności nieliniowych o nieznanej formie nieliniowości. Większość zmiennych występujących w przyrodzie zależy od innych zmiennych w sposób nieliniowy, a niezwykle rzadkie są przypadki, kiedy postać nieliniowości jest znana, co umożliwia jej przybliżanie za pomocą metod analitycznych lub numerycznych. W niektórych przypadkach udaje się zidentyfikować postać nieliniowości za pośrednictwem metod statystycznych.

Charakter nieliniowości, jak też jej natężenie mogą być bardzo różne. Zależności o słabo nieliniowym charakterze są aproksymowalne liniowo. Służą do tego narzędzia statystyczne np. metoda regresji liniowej. Wykonanie przybliżenia liniowego wiąże się często z koniecznością rozbicia zakresu zmienności zmiennej na fragmenty, w których nieliniowość jest zaniedbywalna. Jednakże opis zmiennej w całym zakresie jej wartości wymaga narzędzia zdolnego modelować zależności nieliniowe o nieznanej, często złożonej postaci nieliniowości. Do tego zadania bardzo dobrze nadają się sieci neuronowe. Pokazano, że sieć neuronowa o dwóch warstwach ukrytych jest zdolna do przybliżania dowolnej funkcji nieliniowej, również funkcji charakteryzującej się nieciągłościami. Warunkiem jest różnowartościowość odwzorowania dziedziny funkcji w jej

przeciwdziedzinę [Masters,1996]. Oczywiście, im słabsza nieliniowość zależności, tym mniejsza przewaga sieci nad innymi metodami obliczeniowymi, a efektywność podejścia neuronowego zbliża się do efektywności rozwiązań liniowych.

Omawiana właściwość sieci neuronowych ujawnia się w ich zastosowaniu do prognostryki stężeń zanieczyszczeń w oparciu o podejście deterministyczne. Powszechnie uznawana jest zależność stężeń zanieczyszczeń od parametrów meteorologicznych, emisji zanieczyszczeń oraz warunków topograficznych. Ten fakt jest wykorzystywany do budowy modeli regresyjnych, gdzie w roli zmiennych objaśniających występują parametry meteorologiczne oraz wielkość emisji zanieczyszczeń, często w postaci oszacowania. Warunki topograficzne zwykle pomija się ze względu na ich stabilność w czasie. Jeżeli zmiennych objaśniających ani objaśnianych nie poddano żadnym przekształceniom, regresja jest narzędziem modelowania liniowego. Takie rozwiązanie zastosowano w przytoczonych dalej pracach. Gardner i Dorling,1999 pokazali, że w zakresie modelowania średniogodzinnych stężeń tlenków azotu NO i NO_x w powietrzu w Londynie na podstawie bieżących wartości parametrów meteorologicznych perceptron wielowarstwowy umożliwia uzyskanie lepszych prognoz stężeń niż równanie regresji. Znajduje to swój wyraz we wskaźnikach jakości modelu jak średni błąd względny (MAE), średni błąd kwadratowy (RMSE), i współczynnik determinacji, które są wyraźnie słabsze dla równań regresji. Analizą objęto okres dwóch lat, z których jeden rok posłużył do nauki sieci, a drugi do jej testowania.

Również Comrie,1997 wykazał przewagę sieci neuronowych nad modelem regresji wykorzystywanym do prognozowania maksymalnych średniogodzinnych stężeń ozonu w oparciu o parametry meteorologiczne dla miast w Stanach Zjednoczonych. Swoje badania oparł na danych pomiarowych z pięciu lat 1991-95 dla miesięcy od maja do września. Wyniki prognozowania z zastosowaniem sieci nie są dramatycznie lepsze od wyników uzyskanych za pomocą równania regresji, natomiast lepsza jest jakość modeli. Sieć poprawia odtwarzanie wysokich i niskich stężeń ozonu w stosunku do modelu regresji, choć nie eliminuje zjawiska zaniżania wysokich stężeń i zawyżania stężeń niskich, typowego dla wszystkich modeli statystycznych.

Także Baran,1999 zaprezentował wyniki analiz potwierdzających przewagę sieci neuronowej nad liniowym modelem regresyjnym w modelowaniu średniogodzinnych stężeń ozonu w Krakowie na podstawie temperatury, prędkości wiatru i stężeń NO_x . Wysokie współczynniki korelacji między wynikami modelu a pomiarami, w porównaniu z modelami innych autorów, sugerują, że skorzystano z selektywnego doboru danych pochodzących z godzin, które charakteryzuje silna współzależność stężeń ozonu i tlenków azotu. Do analiz wybrano dane

pomiarowe z godzin nocnych 21:00-3:00 w 1994 roku. Nie zmienia to jednak faktu, że wykazano przewagę podejścia neuronowego nad regresją liniową. Wyniki wskazują na występowanie nieliniowej zależności stężeń godzinnych NO_2 , NO_x jak też maksymalnych godzinnych stężeń ozonu od warunków meteorologicznych. Sugerują ponadto istnienie nieliniowego związku godzinnych stężeń ozonu nocą z warunkami meteorologicznymi i stężeniami NO_x . Identyfikacja nieliniowości i wprowadzenie jej do modelu opisującego zmienną stało się możliwe dzięki wykorzystaniu sieci neuronowej.

Gardner i Dorling, 2000 opracowali modele służące do prognozy godzinowych stężeń ozonu na podstawie parametrów meteorologicznych dla wybranych miast Wielkiej Brytanii. Uzyskane wyniki wykazują przewagę modeli neuronowych nad modelami regresji. Wyższość modeli neuronowych jest uwarunkowana ich zdolnością do modelowania zależności nieliniowych. Skomplikowane przemiany, jakim ozon podlega w powietrzu wykazują taki charakter. Autorzy zaproponowali też alternatywny model prognostyczny w postaci drzewa regresji. Jakość tego modelu jest słabsza niż sieci neuronowej, lecz sam model umożliwia pogłębioną analizę wpływu czynników meteorologicznych na kształtowanie stężenia ozonu.

Inny przykład poprawy wyników modelowania dzięki zastosowaniu sieci neuronowych podała Boznar, 1993. Uzyskała polepszenie dokładności prognoz w stosunku do modelu dyfuzyjnego. Boznar zajmowała się prognozowaniem stężeń SO_2 wokół elektrowni Sostanje w Słowenii z wyprzedzeniem od 1 do 2 godzin, pod kątem możliwości sterowania emisją. Rezultaty uzyskane w wyniku zastosowania modelu dyfuzji Pasquille'a okazały się niezadowolające. Model dobrze przybliżał stężenia rzeczywiste wyłącznie w dużej odległości od elektrowni i w warunkach dużej niestabilności atmosfery. Natomiast prognoza stężeń wykonana przez model neuronowy uwzględniający wielkość emisji, zestaw parametrów meteorologicznych w kilku punktach wokół elektrowni jak też wcześniej rejestrowane wartości stężeń zanieczyszczeń wypadła pozwoliła uzyskać oczekiwaną dokładność. Przedstawione przebiegi stężeń rzeczywistych i prognozowanych informują, że dotyczy to zwłaszcza stacji Veliki, gdzie autorzy modelu uniknęli odtwarzania przez sieć wcześniejszych wartości stężeń prezentowanych na wejściu sieci.

Ciekawe porównanie metod przeprowadził Kolehmainen, 2001. Zestawił wyniki prognozowania stężeń NO_2 na terenie Sztokholmu za pomocą modelu regresji, gdzie zmienne objaśniające stanowiły składowe periodyczne szeregu czasowego stężeń, sieci neuronowych wykorzystujących surowe stężenia (samoorganizującej i perceptronu) oraz sieci modelujących zakłócenia pozostałe po usunięciu składowych periodycznych z serii czasowej stężeń (samoorganizującej i perceptronu). Badania doprowadziły do wniosku, że wszystkie metody

pozwołyły uzyskać dobre rezultaty, jednak najlepiej funkcjonował perceptron wielowarstwowy, uczony stężeń nie poddawanych wcześniejszej obróbce. Jakość modelu była w tym przypadku najlepsza, błąd prognozy stężeń najmniejszy, a rozkład reszt modelu zbliżony do normalnego.

W literaturze przedmiotu można również znaleźć przykład modelu neuronowego realizującego niedeterministyczne podejście do modelowania stężeń w atmosferze i jego porównanie z modelem ARIMA. Przykład dotyczy godzinnych stężeń SO_2 i ozonu rejestrowanych przez trzy stacje monitoringu powietrza na Tajwanie [Kao,2000]. Dwie z nich monitorowały poziom tła, a jedna była zlokalizowana na terenie uprzemysłowionym. Jakość obu modeli porównano dla prognoz wykonywanych z wyprzedzeniem 1 godziny i doby. Wykazano, że prognozy jednogodzinne wykonane siecią neuronową były obciążone mniejszym średnim błędem kwadratowym niż prognozy wykonane za pomocą równania regresji, aczkolwiek ta różnica nie była znacząca. Również trafność większości prognoz 24-godzinnych, wykonanych z zastosowaniem sieci była lepsza niż prognoz sformułowanych na bazie modelu regresji. Wynika stąd, że sieć jest narzędziem modelowania szeregów czasowych co najmniej tak dobrym, jak klasyczna metoda statystyczna. Wyraźnie gorsza jakość prognoz z wyprzedzeniem 24 godzinnym w stosunku do 1 godzinnego świadczy, że tendencja będąca podstawą prognoz krótkoterminowych, nie sięgała okresu doby. W tym horyzoncie czasowym mogły się ujawnić czynniki zewnętrzne niewidoczne dla modelu, a decydujące o zmianach jakościowych w szeregu czasowym stężeń zanieczyszczeń. Informacja o ich istnieniu nie była wykorzystywana do formułowania prognozy. Mimo to, sieć wydawała się również, lepiej niż model regresji, identyfikować prawidłowości obowiązujące w dłuższej perspektywie. Ujawniło się to przewagą sieci nad regresją liniową w zakresie prognoz średniogodzinnych SO_2 i ozonu z wyprzedzeniem dobowym.

2. Możliwość automatyzacji procesów modelowania

Ten argument może się wydawać nieistotny dla badaczy, którzy zajmują się rozwiązywaniem problemów obliczeniowych korzystając z modelowania i posiadają dużą wiedzę zarówno z zakresu problemów będących przedmiotem ich uwagi, jak też różnych obliczeniowych metod i technik modelowania. Zainteresuje jednak zapewne badaczy początkujących, jak też potencjalnych użytkowników rozwiązań, którzy nie będąc zaznajomieni z dziedziną modelowania chcieliby korzystać z jego efektów.

Stosowanie modeli klasycznych wymaga zaproponowania postaci modelu (identyfikacja), wyliczenia wartości współczynników (estymacja parametrów), określenia jakości statystycznej modelu tzn. statystycznej rangi współczynników i poprawności struktury modelu (weryfikacja),

a dopiero na zakończenie oceny efektów modelowania w odniesieniu do rzeczywistych danych liczbowych. Rozwiązanie każdego problemu badawczego jest związane z koniecznością niezależnego pokonania wszystkich przedstawionych etapów procesu modelowania. Sieć neuronowa jest bardzo pomocna, jeżeli badacz nie potrafi zaproponować matematycznego ani statystycznego opisu zjawiska, którym się zajmuje lub proponowany opis nie daje zadowalających rezultatów. Stosowanie podejścia neuronowego wymaga jedynie danych o wartościach zmiennych objaśniających i objaśnianych. Adekwatny model powstaje sam w formie układu powiązań neuronów w sieci neuronowej. Proces identyfikacji modelu przebiega więc automatycznie. Jest to rozwiązanie bardzo korzystne, zwłaszcza gdy całą procedurę identyfikacji modelu trzeba wielokrotnie powtarzać.

Soja, 1999 rozważał prognozowanie wskaźników ekspozycji na ozon normowanych w Austrii tj. siedmiodzinnych stężeń (9:00-16:00) oraz dawki zakumulowanej AOT40, na podstawie dwóch parametrów meteorologicznych: temperatury maksymalnej w ciągu doby i długości dnia. Wykazał, że modele regresyjne spełniają to zadanie równie dobrze, jak modele neuronowe. Przygotowanie modeli regresyjnych poprzedziła jednak pracochłonna procedura identyfikacji ich postaci dla kolejnych miesięcy oraz lat. Najlepszych rozwiązań poszukiwano testując równania regresji, w których występują różne wersje nieliniowych przekształceń zmiennych opisujących. Wykorzystanie modelu neuronowego umożliwia pominięcie całej tej procedury i znacznie ułatwia budowę modelu.

Klasycznym modelem stosowanym dla prognoz krótkoterminowych jest model serii czasowych Boxa-Jenkinsa, zarówno w wersji ARIMA jak też z funkcją transferu. Jak wynika z prac np. [Kao, 2000; Milionis, 1994; Hernandez, 1992] umożliwia on uzyskiwanie dobrych prognoz krótkoterminowych stężeń zanieczyszczeń tzn. z wyprzedzeniem jednego kroku czasowego. Jednak poważnym czynnikiem ograniczającym stosowanie tych modeli jest pracochłonna procedura identyfikacji modelu wymagająca pewnego doświadczenia statystycznego. Ponadto stosowanie modelu w formie właściwej dla wybranego fragmentu serii czasowej jest nieuprawnione dla innej serii czasowej stężeń tego samego zanieczyszczenia, a nawet innego fragmentu tej samej serii. W zasadzie, przed wykonaniem każdej prognozy należy weryfikować postać modelu i zmieniać ją, jeśli istnieje taka potrzeba. W praktyce prognostycznej wywołuje to niemożność przyjęcia modelu o stałej strukturze i poprzestania na ewentualnym uaktualnianiu współczynników, co daje się zautomatyzować, lecz wymaga ciągłej manualnej rewizji poprawności strukturalnej modelu. Wiąże się to z koniecznością kosztownej obsługi procesu prognostycznego przez ludzi wykwalifikowanych i skutecznie zapobiega upowszechnieniu tych modeli.

W przeciwieństwie do modelu serii czasowych, sieć neuronowa automatycznie modyfikuje swoją strukturę dostosowując się do nowych danych, które jej dostarczono. Model prognostyczny w formie sieci neuronowej minimalizuje błąd związany z udziałem w procesie formułowania prognozy czynnika ludzkiego, a wynikający z pomyłek lub błędnego osądu człowieka. W ten sposób niezawodność systemu prognostycznego ulega poprawie. Podejście neuronowe jest więc obiecującą techniką otwierającą drogę do budowy automatycznych systemów prognozowania jakości powietrza.

3. Gwarancja poprawności metodologicznej modelu

Warto poruszyć temat poprawności metodologicznej modeli statystycznych proponowanych jako narzędzia modelowania czasowej zmienności stężeń zanieczyszczeń w powietrzu atmosferycznym. Traktowanie tej kwestii z uwagą jest zaletą bardzo niewielu publikacji poświęconych problemom modelowania. Spośród wszystkich dostępnych rozwiązań najpowszechniej proponowane są modele regresji. Sporadycznie jednak prezentowana analiza błędu obejmuje analizę reszt modelu, która stanowi podstawę oceny efektywności estymatorów parametrów modelu, czyli jego optymalności. Wiadomo jednocześnie, że reszty deterministycznych modeli regresyjnych, bazujących na opisie stężeń zanieczyszczeń wyłącznie w oparciu o takie zmienne, jak parametry meteorologiczne i emisja zanieczyszczeń, charakteryzuje bardzo wysoka autokorelacja. Oznacza to, że estymatory parametrów modelu wyznaczone metodą najmniejszych kwadratów nie są efektywne, a otrzymany model nie jest optymalnym rozwiązaniem problemu. Uzyskanie najlepszego modelu regresyjnego wymaga zastosowania nieliniowych metod estymacji parametrów równania regresji [Unal,2000; Kartal 1998].

Znaną praktyką jest włączanie do zestawu zmiennych objaśniających opóźnienia zmiennej objaśnianej. Umożliwia ono poprawę rezultatów prognozowania. W takich okolicznościach jednakże testy weryfikujące poprawność modelu regresji oprogramowane standardowo w pakietach statystycznych tracą ważność i ocena poprawności modelu na ich podstawie prowadzi do fałszywych wniosków. Autentyczna weryfikacja modelu wymaga posłużenia się specjalistycznymi testami statystycznymi. Z doświadczenia autora wynika, że reszty modelu regresyjnego opisującego stężenia półgodzinne w omawiany sposób są obciążone autokorelacją. Świadczy to o nieoptymalności modelu w sensie statystycznym. Autor nie znalazł w literaturze przedmiotu przykładów pełnej oceny poprawności modeli regresyjnych zawierających opóźnienie zmiennej objaśnianej po stronie zmiennych objaśniających. Brak informacji o rzetelnej weryfikacji prezentowanych modeli regresyjnych wywołuje wątpliwości dotyczące ich poprawności metodologicznej [Milionis,1994].

Znacznie poważniej traktowane są założenia dotyczące modeli serii czasowych, gdzie istnieje wymóg stacjonarności modelowanej serii czasowej i rozpoznania wzorców autokorelacji oraz korelacji cząstkowej charakteryzujących tę serię. Wynika to prawdopodobnie z faktu, że tym narzędziem posługują się zasadniczo badacze związani profesjonalnie ze statystyczną obróbką danych, a od specjalistów z zakresu zagadnień zanieczyszczenia powietrza atmosferycznego pochodzi tylko wsparcie interpretacyjne.

Podjęcie neuronowe uwalnia badacza od zajmowania się zagadnieniem poprawności metodologicznej modelu. Każda sieć neuronowa jest poprawna metodologicznie pod warunkiem, że na jej wejścia podawane są dane z dziedziny odwzorowania, które sieć ma zrealizować, a na wyjście, dane z przeciwdziedziny tego odwzorowania, najlepiej znormalizowane w zakresie $[-1, 1]$. Jedynym kryterium oceny sieci neuronowej jest skuteczność realizacji zadania, które przed nią postawiono. Tę ocenę wykonuje się na podstawie analizy reszt modelu. Raz przeprowadzona i zakończona sukcesem identyfikacja typu i wielkości sieci rozwiązującej konkretny problem badawczy jest w zasadzie ostateczna i nie musi być powtarzana. Jest to sytuacja zupełnie przeciwna do zasad obowiązujących przy posługiwaniu się statystycznymi modelami serii czasowych. W przypadku modeli neuronowych nie występuje zagrożenie doboru niewłaściwej metody estymacji parametrów modelu, które dotyczy modeli regresyjnych. Metoda wyliczania wag sieci jest jedna, a jej modyfikacje wpływają w zasadzie tylko na tempo uczenia sieci i objętość pamięci, wymaganej do przeprowadzenia obliczeń.

Zamykając wypowiedź o argumentach za stosowaniem sieci neuronowych należy również uświadomić czytelnikowi podstawowy słaby punkt tej metody obliczeniowej. Jest nim niemożność przedstawienia zależności zidentyfikowanej przez sieć w postaci analitycznej. Jak wiadomo zapis w postaci formuły umożliwia wnioskowanie o oddziaływaniu jednych zmiennych na inne, o sile i kierunku tego oddziaływania. Sieć nie pozwala na uzyskanie formalnego opisu zależności zmiennych od siebie, jednak możliwości, które stwarza w znacznym stopniu rekompensują tę wadę.

2. Przegląd metod obliczeniowych stosowanych w opisie imisji

Jak stwierdzono we wstępie, pozyskiwanie informacji o bieżącej i spodziewanej jakości powietrza atmosferycznego na podstawie pomiarów imisyjnych stężeń zanieczyszczeń wiąże się ze stosowaniem właściwych metod obliczeniowych. Dla potrzeb tej pracy przyjęto podział metod na analityczne, statystyczne i neuronowe. Wszystkie są stosowane do rozwiązywania zagadnień analizy czasowej i przestrzennej stężeń zanieczyszczeń stanowiących przedmiot zainteresowania autora. Krótka charakterystyka wymienionych grup metod przedstawiona w tym rozdziale ma wprowadzić czytelnika w aparat pojęciowy metod analitycznych, statystycznych i neuronowych, przedstawić ich zalety oraz ograniczenia. Autor wskaże zagadnienia z zakresu oceny jakości powietrza atmosferycznego, do których rozwiązywania poszczególne metody najbardziej się nadają.

2.1. Metody analityczne

W zakresie pozyskiwania informacji z wyników monitoringu zanieczyszczeń powietrza metody analityczne mogą być wykorzystywane przede wszystkim do celu interpolacji przestrzennej. Propozycje rozwiązań analitycznych w sferze analiz czasowych nie będą tu rozważane. Poniżej przedstawiono zasadnicze elementy ogólnej metodologii interpolacji oraz scharakteryzowano analityczne metody interpolacji. Stanowią one bazę koncepcyjną dla opracowanej przez autora i prezentowanej w pracy neuronowej metody interpolacji. Z tego powodu omówiono je szerzej.

2.1.1. Podstawy metodologii interpolacji

Zgodnie z teorią interpolacji, każdemu punktowi pomiarowemu odpowiada jego otoczenie nazywane inaczej zasięgiem oddziaływania punktu. Jest to strefa, w której istnieje związek między wartością parametru mierzoną w punkcie pomiarowym a dowolnym innym punkcie. W miarę oddalania od punktu pomiarowego aż do granicy zasięgu oddziaływania, zależność stopniowo zanika. Według definicji zasięgu oddziaływania wartości przyjmowane poza jego granicą przez wielkość monitorowaną nie wykazują związku z wartościami zmierzonymi w punkcie pomiarowym. W większości przypadków nie wiadomo jaki jest faktyczny zasięg oddziaływania punktu pomiarowego oraz jego kształt. W związku z tym poszukuje się sposobu na uzależnienie wpływu punktu pomiarowego na punkty położone w jego otoczeniu od odległości między nimi. Spotykane w literaturze rozwiązania stały się wyróżnikiem odrębnych metod interpolacji i opierają się na

specyficznych dla siebie przesłankach np. na założeniu że:

- Wpływ punktu pomiarowego na jego otoczenie jest odwrotnie proporcjonalny do odległości od punktu jak w metodzie odwrotności odległości (*inverse distance*), lub do kwadratu tej odległości, jak w metodzie kwadratu odwrotności odległości (*inverse distance to a power*) [Hopkins,1999].
- W całym otoczeniu punktu pomiarowego parametr mierzony ma taką samą wartość jak w punkcie pomiarowym, a otoczenie kończy się w połowie odległości między sąsiednimi punktami pomiarowymi. To cecha metody najbliższego sąsiada (*nearest neighbour*).
- Związek między wartością zmierzoną w punkcie pomiarowym i w dowolnym innym punkcie nie jest wyłącznie funkcją odległości między punktami, lecz zależy również od charakteru zmienności przestrzennej wielkości mierzonej. Takie podejście wykorzystuje metoda pod nazwą *kriging* [Hopkins,1999].
- Wpływy poszczególnych punktów pomiarowych nie są definiowane explicite, natomiast pośrednio odzwierciedleniem ich istnienia jest powierzchnia interpolacji należąca do pewnej kategorii funkcji dwóch zmiennych (współrzędnych (x,y)). Postać tej funkcji powinna być typowa dla rozkładu przestrzennego wielkości interpolowanej. Taki przypadek dotyczy metod funkcyjnych.

Oszacowanie wartości mierzonego parametru między punktami pomiarowymi wymaga podania zasady składania zasięgów oddziaływania punktów pomiarowych w przestrzeni. Wdrożenie procedury składania otoczeń punktów pomiarowych prowadzi do uzyskania ciągłego przestrzennego obrazu mierzonej wielkości na podstawie informacji o charakterze dyskretnym. Sposób składania zasięgów oddziaływań punktów pomiarowych jest kluczowym elementem każdej metody interpolacji. Stanowi on drugie kryterium rozróżniające metody interpolacji, obok metody określania kształtu oddziaływania punktów pomiarowych, w obrębie ich zasięgu.

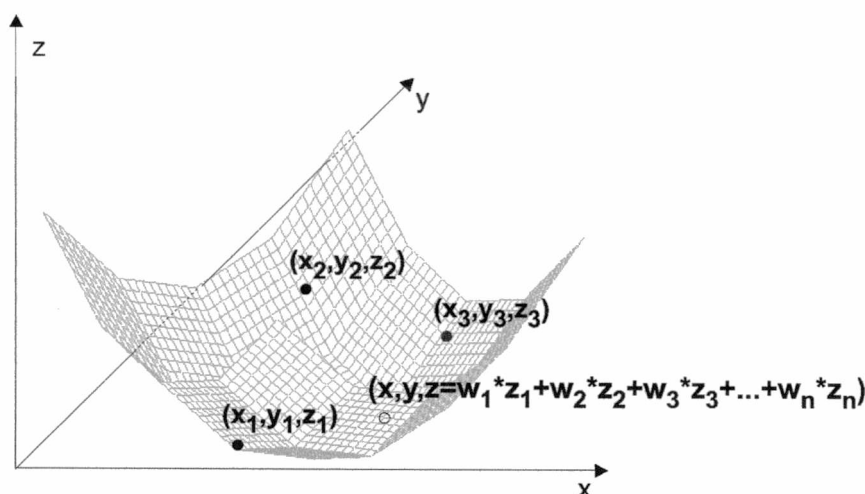
Ze względu na kryterium składania zasięgów oddziaływań punktów pomiarowych wyróżnia się dwie podstawowe grupy metod interpolacji. Noszą one nazwy interpolacji metodą średniej ważonej oraz interpolacji przez dopasowanie funkcji. [Watson,1992]

2.1.2. Interpolacja metodą średniej ważonej

Zagadnienia określania postaci oddziaływania punktów pomiarowych w obrębie ich otoczeń oraz składania otoczeń występują w tej grupie metod w sposób jawny. Wartość zmiennej interpolowanej w zadanym punkcie jest wyznaczana poprzez sumowanie w nim wpływów punktów pomiarowych, do których otoczenia należy ten punkt. Miarą wpływu punktu

pomiarowego na punkt, gdzie wykonuje się interpolację jest waga, będąca zwykle funkcją odległości między punktami. Im większa odległość tym mniejsza waga. Zestaw wag od wszystkich punktów pomiarowych jest wyliczany dla każdego punktu, w którym wielkość mierzona wymaga interpolacji. Zwykle wagi sumują się do jedności.

Zasadniczą zaletą interpolacji metodą średnich ważonych jest uwypuklenie cech lokalnych zmiennej mierzonej. Lokalne trendy w danych mogą znaleźć odzwierciedlenie w powierzchni interpolowanej w stopniu niemożliwym do uzyskania przy zastosowaniu metod funkcyjnych, jeśli liczba parametrów funkcji ma być rozsądna. Uzyskana w metodzie średnich ważonych powierzchnia interpolacji jest zdominowana przez wpływy lokalne (rys.2.1). Proporcje nasilenia wpływu punktów bliższych i dalszych można regulować metodą doboru wag. Jest ona zasadniczym elementem różnicującym poszczególne metody w grupie metod średniej ważonej [Watson,1992].



Rysunek 2.1. Przykład interpolacji metodą średniej ważonej. Wartość zmiennej interpolowanej z w dowolnym punkcie o współrzędnych (x,y) jest wyliczana jako suma iloczynów wag $(w_1, w_2, w_3, \dots, w_n)$ od wszystkich punktów pomiarowych i zmierzonych tam wartości zmiennej $(z_1, z_2, z_3, \dots, z_n)$. Na rysunku pokazano wyniki pomiaru w trzech przykładowych punktach pomiarowych.

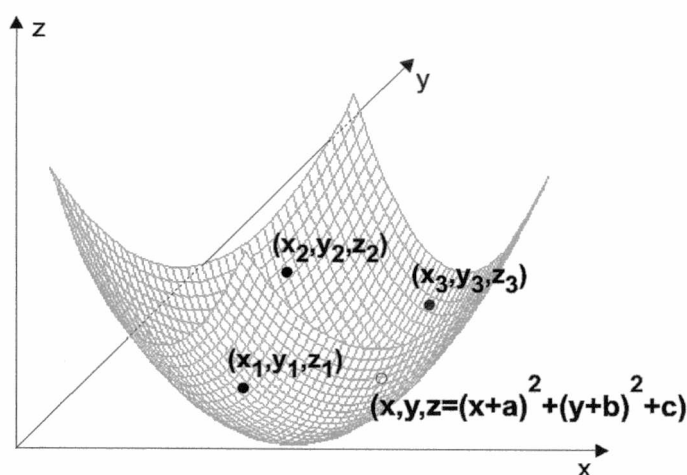
Ze względu na podane cechy interpolacja metodami z grupy średniej ważonej jest preferowana w rozwiązywaniu zagadnień interpolacji przestrzennej stężeń zanieczyszczeń. Należą tu wspomniane już wcześniej metody *inverse distance*, *nearest neighbour*, *kriging* oraz szereg innych metod.

2.1.3. Interpolacja funkcyjna

Podjęcie alternatywne do metod średniej ważonej reprezentuje interpolacja funkcyjna. W tej grupie metod nie analizuje się i nie składa zasięgów oddziaływań punktów pomiarowych.

Interpolacja przez dopasowanie funkcji polega na przybliżeniu powierzchni interpolowanej funkcją analityczną o zadanej postaci. Najczęściej wielomianem odpowiedniego stopnia. Funkcja ma reprezentować regionalne lub/i lokalne związki między interpolowanymi danymi. Charakter związków odzwierciedlają: liczba i wartości parametrów w równaniu funkcji. Zadanie interpolacji polega więc na zaproponowaniu postaci funkcji interpolującej, wyznaczeniu wartości parametrów, a następnie wyliczeniu wartości funkcji we wszystkich punktach, gdzie dokonywana jest interpolacja. Podstawą doboru funkcji jest pewna wiedza a priori dotycząca charakteru rozkładu przestrzennego zmiennej interpolowanej.

Zastosowanie metod funkcyjnych pozwala uzyskać bardzo duże prawdopodobieństwo, że rzeczywisty i powstały w wyniku interpolacji rozkład przestrzenny zmiennej wyglądają tak samo. Wynika to z wykorzystania wiedzy na temat faktycznej postaci rozkładu przestrzennego zmiennej interpolowanej.



Rysunek 2.2. Interpolacja funkcyjna. Wielkość mierzona przybliżono funkcją o równaniu $z=(x+a)^2+(y+b)^2+c$. Postać funkcji zależy od charakterystyki przestrzennej zmiennej z . Parametry funkcji: a , b i c są wyliczane na podstawie informacji o wartościach z_1 , z_2 i z_3 , ... z_n zebranych w punktach pomiarowych o współrzędnych (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , ... (x_n, y_n) . Na rysunku pokazano wyniki pomiaru w trzech przykładowych punktach pomiarowych.

Inna zaleta podejścia funkcyjnego polega na uzyskaniu gładkiej powierzchni interpolującej. Określona analitycznie funkcja nadaje powierzchni interpolowanej ogólny kształt, co umożliwia wyeliminowanie wpływu danych wyraźnie odstających od zadanej formy powierzchni. Przejawia się to ścinaniem pików i wyrównywaniem nadmiernych zagłębień (rys.2.2). Cechą metody interpolacji przez dopasowanie funkcji jest również dominacja informacji globalnej nad informacją lokalną [Watson, 1992].

Istnieje podstawowy warunek, którego niespełnienie uniemożliwia interpolację metodą dopasowania funkcji. Jest nim znajomość postaci analitycznej funkcji, która ma opisywać

powierzchnię interpolowaną. Z tego powodu metody interpolacji funkcyjnej praktycznie nie są stosowane do rozwiązywania zagadnień interpolacji przestrzennej stężeń zanieczyszczeń.

Dla uściślenia zagadnienia warto zaznaczyć, że generowanie rozkładów przestrzennych wielkości mierzonej, często wymaga wykonania zarówno interpolacji jak i ekstrapolacji. Sytuacja taka zachodzi wówczas, gdy rozkład przestrzenny obejmuje pewien obszar leżący poza granicami wyznaczonymi przez zewnętrzne punkty pomiarowe sieci. Punkty znajdujące się w tym obszarze nie leżą między żadnymi dwoma punktami pomiarowymi. Nie można tam więc wykonać interpolacji, konieczna jest zaś ekstrapolacja. W ograniczonym zakresie, przy niewielkim przekraczaniu granic układu punktów pomiarowych, pewne metody interpolacji można jednak wykorzystać do rozwiązania problemu ekstrapolacji.

2.2. Metody statystyczne

Największą grupę obliczeniowych metod analizy danych pomiarowych stanowią metody statystyczne. Umożliwiają one stawianie hipotez dotyczących charakteru przestrzenno-czasowej zmienności stężeń zanieczyszczeń oraz uwarunkowań tej zmienności. Szereg metod statystycznych opracowano pod kątem prognostycznym. Poniżej przedstawiono charakterystykę kilku metod wybranych ze względu na możliwości, które otwierają w zakresie analizy danych.

2.2.1. Analiza czynnikowa

Analiza metodą czynników głównych (PCA) stanowi statystyczne narzędzie analizy danych, które poszukuje w nich dominującego wzorca zmienności. Deterministyczny fragment pierwotnych serii czasowych jest wyrażany w formie czynników głównych. U podstaw analizy leży założenie, że zbiór danych składa się ze zmiennych powiązanych ze sobą. Dzięki temu istnieje możliwość przekształcenia m wymiarowej przestrzeni zmiennych pierwotnych w k wymiarową przestrzeń czynników głównych, gdzie $k < m$. Kolejne czynniki, będące nowymi zmiennymi, tłumaczą kolejne części wariancji zbioru zmiennych pierwotnych. Pierwszy czynnik tłumaczy największą część wariancji, drugi mniejszą i tak w kolejności malejącej, aż do ostatniego czynnika. Formalnie, m zmiennych pierwotnych podlega przekształceniu w m czynników. Jednakże procent wariancji wyjaśnianej przez ostatnie czynniki jest zanedbywalny, stąd zwykle pomija się je skupiając uwagę na pierwszych k czynnikach. Redukując wymiar zbioru zmiennych poszukuje się takiego k , żeby pozostawić jak najwięcej zmienności zawartej w danych pierwotnych i jednocześnie by k było jak najmniejsze. Konieczny jest więc kompromis między liczbą nowych zmiennych a zmiennością, która zostaje w nich zachowana.

Analiza czynnikowa odfiltrowuje część serii czasowych, która jest zaszumiona skupiając się na części wspólnej uchwyconej w osiach głównych. Jeżeli D jest macierzą n na m , gdzie każda z m kolumn zawiera jedną zmienną, a n jest liczbą realizacji tej zmiennej, to czynniki główne p_i są odpowiednimi kombinacjami liniowymi zmiennych pierwotnych wg równania:

$$p_i = De_i \quad \text{dla } i=1, \dots, k \text{ gdzie } k < m \quad (2.1)$$

Wektory e_i to znormalizowane wektory współczynników czynnikowych (eigenvectors) macierzy korelacji (kowariancji) $C=D^TD$ z próby. Wektory współczynników są rozwiązaniami równania:

$$Ce_i = \lambda e_i \quad (2.2)$$

Gdzie λ_i to wartości własne (eigenvalues) odpowiadające kolejnym wektorom współczynników. Są zdefiniowane tak, by spełniać związek $e_i e_j^T = \delta_{ij} \lambda_i$. Jeżeli macierz C jest macierzą korelacji, δ_{ij} wynosi jeden dla $i=j$ oraz zero dla $i \neq j$. Wartości własne reprezentują wariancję związaną z odpowiednimi wektorami współczynników, co jest tożsame z wariancją związaną z odpowiednimi czynnikami. Są zwykle indeksowane w kierunku wartości malejących tak, że pierwsza wartość własna reprezentuje wariancję zgromadzoną w pierwszej składowej i jest to wariancja największa, druga wartość własna odpowiada drugiej największej wariancji itd. Najwyżej indeksowana wartość własna wyjaśnia najmniej wspólnej wariancji. Czynniki odpowiadające poszczególnym wartościom własnym odzwierciedlają czasową zmienność zmiennych pierwotnych poprzez projekcję tych zmiennych na te czynniki (2.1). Czynniki główne są ortogonalne tzn. liniowo niezależne, w przeciwieństwie do zmiennych pierwotnych.

Stopień powiązania zmiennych pierwotnych z wyróżnionymi czynnikami oddają dwie miary: ładunek zmiennej oraz zasób zmienności wspólnej (communalty). Ładunek czynnikowy zmiennej jest współczynnikiem korelacji między zmienną a czynnikiem. Zasób zmienności wspólnej oznacza procent wariancji zmiennej, który wyjaśnia czynnik. Wartości własne czynników, ładunki czynnikowe zmiennych oraz zasoby zmienności wspólnej, wyliczone dla k składowych głównych stanowią podstawę interpretacji wyników analizy czynnikowej.

W swej podstawowej wersji analiza czynnikowa porządkuje czynniki według wyjaśnianej przez nie części wariancji zmiennych pierwotnych. Jeżeli w wyniku zastosowania analizy czynnikowej w podstawowej wersji powstają problemy interpretacyjne, pomocnym może się okazać obrót czynników. Pierwszy zestaw czynników nie jest jedynym możliwym. Jego liniowe przekształcenie lub obrót prowadzi do uzyskania równoważnej reprezentacji danych. Po obrocie

jednakże każdy kolejny czynnik oprócz pierwszego nie będzie wyjaśniał maksimum pozostającej wariancji. Uzyskany układ składowych może natomiast ułatwić interpretację.

Analiza czynnikowa spełnia szereg zadań związanych z opisem imisji. Najpowszechniej jest stosowana do określania przestrzennych wzorców stężeń zanieczyszczeń nad obszarami objętymi monitoringiem. Wyróżnione wzorce są następnie kojarzone z czynnikami, które mogły wywołać ich powstanie. Należą do nich zarówno źródła emisji, jak też specyficzne warunki meteorologiczne [Zwoździak, 1986].

Wykorzystując analizę czynnikową You, 1999 podzielił cały obszar Tajwanu na cztery strefy o odrębnej charakterystyce pod względem średniodobowych stężeń ozonu. Przyczynę odrębności stanowiły wzorce warunków meteorologicznych typowe dla wyróżnionych obszarów. Eder, 1993 wydzielił za pomocą analizy czynnikowej sześć regionów w Stanach Zjednoczonych, gdzie maksymalne dobowe stężenia ozonu układały się w specyficzny sposób. Uzyskany podział dobrze korespondował z trasami wędrówki oraz częstością występowania wyżów atmosferycznych. Badania Ashbough, 1984 dotyczyły pyłu zawierającego siarkę występującego w powietrzu na terenie zachodnich Stanów Zjednoczonych. Po rotacji czynników głównych, uzyskanych w wyniku analizy czynnikowej zidentyfikowano pięć regionów, gdzie pył w powietrzu był wspólnego pochodzenia. Źródło pyłu w pierwszym regionie stanowiła huta miedzi, drugi odpowiadający Wielkim Równinom znajdował się pod okresowym wpływem mas zanieczyszczonego powietrza ze wschodu, a za stężenia pyłu w trzech pozostałych regionach odpowiadały czynniki lokalne. Tą samą metodykę zastosował Liu, 1996 do oceny przestrzennych wzorców pyłu zawierającego siarkę i pyłu zawierającego ołów na terenie New South Wales pod kątem czynników wpływających na wykształcenie się wyróżnionych wzorców stężeń.

PCA przeprowadzona w inny sposób umożliwia grupowanie zanieczyszczeń według ich źródeł. Takie badania przeprowadził Statheropoluous, 1998 w Atenach. Analizie czynnikowej poddano stężenia CO, NO, NO₂, O₃, SO₂ i dymu rejestrowane przez pięć lat, poszukując czynników odpowiadających za ich emisję. Główne wydzielone czynniki stanowiły: spalanie benzyny, spalanie oleju napędowego oraz interakcje związków chemicznych z ozonem.

Bardzo ciekawym przykładem wykorzystania analizy czynnikowej są badania Poissant, 1996, który przeanalizował współzależność stężeń ozonu, szeregu związków organicznych i temperatury na obszarach wiejskich z Southern Ontario w Kanadzie, dla różnych sektorów wiatru w porze nocnej i dziennej. W tej grupie parametrów poszukiwał wspólnych wzorców zmienności w celu wyjaśnienia prawideł rządzących zachowaniem ozonu w powietrzu.

2.2.2. Regresja

Przykładem metody statystycznej, która realizuje modelowanie deterministyczne jest regresja. Model regresji umożliwia statystyczny opis zmiennej objaśnianej przy pomocy różnych zmiennych objaśniających. Choć za ostateczny cel analizy regresji uważana jest prognozyka, modele regresji mają też inną właściwość, wynikającą z ich struktury. Poprawnie zbudowany model regresji umożliwia ocenę wpływu zmiennych objaśniających na zmienną objaśnianą. Ten aspekt diagnostyczny nie występuje w innych modelach statystycznych, których celem jest prognozyka.

Budowa modelu regresji dla zmiennej wynika najczęściej z braku wiedzy wystarczającej do jej funkcyjnego opisu. Stąd konieczność stosowania opisu statystycznego. Wymaga on postawienia hipotezy o przyczynach zmienności zmiennej objaśnianej, co jest tożsame z wyborem zestawu zmiennych objaśniających oraz założenia postaci zależności między zmienną objaśnianą a zmiennymi objaśniającymi. Na przykład Ping Shi, 1997 uzależnił stężenia NO_2 i NO_x od prędkości wiatru, temperatury, wysokości warstwy mieszania, klasy stabilności atmosfery, wilgotności względnej i promieniowania, proponując do opisu tej zależności równanie regresji liniowej.

Najczęściej pod pojęciem regresji rozumiana jest regresja liniowa, czyli modelowanie związków o charakterze liniowym między zmiennymi objaśniającymi a zmienną objaśnianą. Narzędzie statystyczne, którym jest regresja umożliwia jednak również modelowanie zależności nieliniowych. Warunkiem jest możliwość przekształcenia nieliniowej postaci zależności zmiennych w liniową zależność funkcji tych zmiennych. Powodem są procedury estymacji parametrów modelu regresji, które opracowano w szczególności dla przypadku regresji liniowej. Kartal, 1998 zaproponował opis przekształconych logarytmicznie dobowych stężeń SO_2 i dymu za pomocą logarytmów parametrów meteorologicznych i stężeń z dnia poprzedniego. Faktycznie, oznacza to opis stężeń SO_2 i dymu iloczynem zmiennych niezależnych podniesionych do potęg odpowiadających współczynnikom, które stoją przy ich logarytmach w równaniu regresji.

Ogólna postać modelu regresji jest następująca:

$$y = X\beta + \varepsilon \quad (2.3)$$

Gdzie:

- y- wektor obserwacji zmiennej Y, o wymiarach $(n \times 1)$
- X- macierz obserwacji na zmiennych niezależnych o wymiarach $(n \times (k+1))$
- β - wektor współczynników (parametrów) o wymiarach $((k+1) \times 1)$
- ε - wektor składników losowych o wymiarach $(n \times 1)$

Najczęściej stosowaną metodą estymacji współczynników równania regresji β jest metoda najmniejszych kwadratów (MNK). Kryterium optymalizacyjnym w tej metodzie jest minimalizacja sumy średnich kwadratów reszt modelu, czyli różnic między obserwacjami zmiennej objaśnianej a wartościami wyliczonymi z modelu:

$$S = \sum \varepsilon^2 = \sum (y - X\beta)^2 \rightarrow \min \quad (2.4)$$

Metoda zapewnia uzyskanie estymatorów nieobciążonych (wartość oczekiwana estymatora parametru jest równa wartości parametru), zgodnych (estymator parametru jest stochastycznie zbieżny do tego parametru) i efektywnych (estymator parametru ma najmniejszą wariancję ze wszystkich estymatorów nieobciążonych) współczynników równania regresji pod warunkiem spełnienia następujących założeń dotyczących reszt modelu:

1. $E\varepsilon_i = 0$
2. $D^2(\varepsilon_i) = E\varepsilon_i^2 = s^2$
3. $Cov(\varepsilon_i, \varepsilon_j) = E\varepsilon_i \varepsilon_j = 0$ dla $i \neq j$

Warunek 1 oznacza, że wartość oczekiwana reszt modelu wynosi zero.

Warunek 2 oznacza, że wariancja reszt modelu jest homogeniczna.

Warunek 3 oznacza, że nie występuje zjawisko autokorelacji reszt modelu.

Jeżeli nie jest spełniony pierwszy warunek, estymatory modelu regresji są obciążone. Jeżeli nie są spełnione założenia dotyczące wariancji (warunek2) lub kowariancji (warunek3) składników losowych modelu, estymatory uzyskane MNK przestają być efektywne. W każdym przypadku nie mają one własności estymatorów optymalnych [Luszniewicz, 1997].

Prawdziwość założeń można sprawdzić dopiero po zbudowaniu modelu i wyliczeniu współczynników metodą najmniejszych kwadratów. Istnieją testy statystyczne umożliwiające taką ocenę. W odniesieniu do wartości oczekiwanej zakłóceń, najprostszą metodą jest test serii. Silnym testem homogeniczności wariancji jest test Levene'a. Występowanie autokorelacji składników losowych można natomiast sprawdzić testem Durбина-Watsona, który jest stosowany najczęściej, ale dotyczy tylko autokorelacji pierwszego rzędu. Inne wzorce autokorelacji wykrywa test Brusch'a-Godfrey'a [Luszniewicz, 1997].

Niespełnienie założeń modelu regresji wywołuje konieczność zmiany postaci modelu lub przeszacowania parametrów modelu za pomocą innej metody estymacji. Jeżeli niespełnione jest założenie o zerowej wartości oczekiwanej reszt, model wymaga przeformułowania poprzez poszerzenie grupy zmiennych objaśniających lub uwzględnienie nieliniowości. Niespełnienie któregoś z pozostałych dwóch założeń można próbować neutralizować za pomocą metody estymacji w postaci uogólnionej metody najmniejszych kwadratów.

Często wykorzystywanym sposobem usuwania autokorelacji reszt modelu jest włączanie do zestawu zmiennych objaśniających przesunięcia czasowego zmiennej objaśnianej. Takie rozwiązanie zastosował Ping Shi, 1997 dla polepszenia wyników prognozy stężeń średniogodzinnych NO_2 i NO_x . Na etapie sprawdzania założeń często okazuje się jednak, że autokorelacja nadal występuje w resztach modelu, współczynniki regresji wyliczone metodą NMK są obciążone, nie są zgodne, a wyliczone reszty modelu zupełnie nie oddają charakteru rzeczywistych zakłóceń. W takich okolicznościach do sprawdzenia obecności autokorelacji nie wystarcza już stosowany standardowo test Durбина-Watsona. Optymalne estymatory współczynników modelu można próbować uzyskać metodą największej wiarygodności lub metodą zmiennych instrumentalnych [Milionis, 1994].

Dopiero wyniki pochodzące z poprawnego metodologicznie modelu regresji można analizować pod kątem zgodności z danymi empirycznymi. Wartości i znaki współczynników poprawnego modelu regresji stanowią podstawę wnioskowania o wielkości i kierunku wpływu zmiennych objaśniających na zmienną objaśnianą. Unal, 2000 opisał średniodobowe stężenia SO_2 za pomocą sześciu czynników głównych wyróżnionych w wyniku analizy czynnikowej, której poddano zestaw warunków meteorologicznych. Jego badania doprowadziły do wniosku, że na najważniejszy czynnik kształtujący stężenia SO_2 w Istambule składają się: ciśnienie atmosferyczne oraz wiatry strefowe i południkowe.

Na koniec, pozostaje zbadanie właściwości prognostycznych modelu regresji. Wiąże się to z oceną różnic wyników modelu i wartości obserwowanych zmiennej objaśnianej dla zestawów wartości zmiennych objaśniających nie wykorzystanych na etapie budowy modelu regresji. Testowe wartości zmiennych objaśniających nie powinny jednak wykraczać poza zakres wykorzystany dla wyliczenia parametrów modelu regresji.

Metody estymacji parametrów równań regresji inne niż MNK nie są oprogramowane w powszechnie stosowanych pakietach statystycznych jak Statistica czy Statgraph. Również w literaturze dotyczącej modelowania stężeń zanieczyszczeń, rzadko można znaleźć wyczerpującą analizę poprawności proponowanych modeli regresji uzyskanych z zastosowaniem metody najmniejszych kwadratów. Powoduje to możliwość podważenia prawomocności wniosków o uwarunkowaniach zmienności stężeń zanieczyszczeń wyprowadzanych na tej podstawie. Również wykonywanie prognoz z zastosowaniem modeli niepoprawnych metodologicznie jest nieuzasadnione. Ze względu na trudności związane z budową poprawnego modelu regresji często stosowane są inne rozwiązania.

Wiele publikacji dotyczy porównań rezultatów modeli regresyjnych z wynikami sieci neuronowych pracujących w układzie deterministycznym np. [Comrie, 1997; Gardner, 1999; Gardner, 2000; Soja, 1999]. Rezultaty uzyskane metodą regresji dla stężeń ozonu są zbliżone

do wyników modeli neuronowych w przypadku stężeń średniodobowych [Comrie,1997]. Jednak już prognozy godzinnych stężeń ozonu są lepsze przy wykorzystaniu sieci neuronowych [Gardner,2000]. To samo spostrzeżenie dotyczy średniogodzinnych stężeń NO_2 i NO_x w Londynie [Gardner,1999]. Zwykle modele regresji liniowej nie uwzględniają elementu nieliniowości, który występuje w zależności stężeń ozonu, NO_2 i NO_x od parametrów meteorologicznych.

2.2.3. Modele stochastyczne ARIMA

Przy braku wiedzy o uwarunkowaniach zmiennej opisywanej, pozostaje możliwość oceny przyszłych wartości zmiennej na podstawie wyłącznie jej zachowania w przeszłości. Jest to podejście niedeterministyczne. Odpowiedni model nazwa się często modelem Box'a-Jenkins'a, od nazwisk jego twórców i umożliwia modelowanie stacjonarnych serii czasowych. W zasadzie jest to model szumu białego [Milionis,1994].

$$Q_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t \quad (2.5)$$

Gdzie: Q_t - stacjonarny szereg czasowy

μ - średnia serii

$\phi_1 \dots \phi_n$ - współczynniki wielomianu modelującego proces autoregresyjny

$\theta_1 \dots \theta_n$ - współczynniki wielomianu modelującego proces średniej ruchomej

$y_{t-1} \dots y_t$ - opóźnione wartości zmiennej

a_t - impuls losowy

Inna obowiązująca nazwa ARIMA(p,d,q) (*Auto Regressive Integrated Moving Average*) niesie informację o sposobie funkcjonowania modelu. Wykorzystuje on bowiem procesy autoregresji i średniej ruchomej występujące w seriach czasowych, które są stacjonarne lub które przekształcono do postaci serii stacjonarnych w wyniku eliminacji trendu oraz sezonowości. Służą do tego odpowiednie operacje matematyczne np. różnicowanie. Krotności różnicowania serii czasowej odpowiada w opisie modelu litera d. Pojęcie stacjonarności oznacza, że średnia, wariancja i wewnętrzna korelacja szeregu czasowego powinny być w przybliżeniu stałe w czasie. Model ARIMA wymaga identyfikacji rzędów wielomianów wykorzystywanych do modelowania procesów autorgresyjnego (p) oraz średniej ruchomej (q) występujących w serii czasowej. Identyfikacja przebiega odpowiednio na podstawie wzorca autokorelacji i autokorelacji cząstkowej uzyskanych dla szeregu czasowego zmiennej. Prawidłowości występujące w serii czasowej mogą zostać zidentyfikowane na podstawie tej serii bez żadnych dodatkowych założeń. Wymagana jest tylko odpowiednia długość serii czasowej. Kwestię tę rozważał teoretycznie Milionis,1994. Również Foszcz,2000, który wykorzystał model ARIMA do prognozy średniodobowych stężeń SO_2 na terenie GOP zauważył, że modele ARIMA opracowane dla jednego sezonu grzewczego

mogą być nieaktualne w innym sezonie grzewczym ze względu na zmianę warunków meteorologicznych.

Dobór struktury modelu ARIMA dla każdej zmiennej jest kwestią indywidualną i wynika ze wzorca autokorelacji i autokorelacji cząstkowej szeregu czasowego tej zmiennej. Fakt, że wybraną serię czasową stężeń można skutecznie modelować z zastosowaniem modelu o określonej strukturze np. (1,1,1) nie oznacza jednocześnie, że model ten jest odpowiedni również dla innej, na pozór podobnej serii. Ponadto, istnieje więcej niż jeden model ARIMA, za pomocą którego można z powodzeniem modelować jedną i tę samą serię czasową. Spośród nich wybierany jest jednak model o najmniejszej liczbie współczynników.

Do weryfikacji poprawności metodologicznej modelu ARIMA konieczna jest analiza reszt modelu. Nie powinny one wykazywać istotnej statystycznie autokorelacji ani autokorelacji cząstkowej.

Modelowane szeregu czasowego zmiennej za pomocą modelu ARIMA ma sens tylko wówczas, gdy zmiany wszystkich czynników zewnętrznych wpływających na zmienną modelowaną nie odbiegają od ustalonego kierunku. Z drugiej strony prognoza jest wykonywana w zasadzie na podstawie kilku ostatnich obserwacji nałożonych na warunki początkowe. Wobec tego, w warunkach rzeczywistych prognozowanie z wykorzystaniem modeli Boxa-Jankinsa może być stosowane wyłącznie do prognoz krótkookresowych [Cieślak,1997]. Chock,1975, który wykorzystał model ARIMA do prognozowania średnich stężeń tygodniowych i maksymalnych dobowych stężeń sumy związków utleniających, CO oraz NO₂ podkreślał, że zdolności predykcyjne modeli opartych na analizie jednoczynnikowej zależą od stałości czynników zewnętrznych oddziałujących na poziom stężeń. W dłuższych okresach czasu czynniki te ulegają zmianom a prognoza wymaga podejścia wieloczynnikowego.

Modele ARIMA są często wykorzystywane jako porównanie dla rozwiązań opartych na modelowaniu serii czasowych, lecz stworzonych w oparciu o inne narzędzia modelowania np. sieci neuronowe [Kao,2000], czy logikę rozmytą [Jorquera,1998].

Wyniki prognozowania za pomocą modelu ARIMA będą słabe, jeżeli seria czasowa zmiennej opisywanej nie ma charakteru stochastycznego, lecz zmienna jest uwarunkowana czynnikami zewnętrznymi. Takı przykład opisał Hernandez,1992, który dowiódł, że wyniki prognozy średniodobowych stężeń aerozolu w powietrzu ulegają wyraźnej poprawie, jeśli w opisie stężeń uwzględnić prędkość wiatru, przechodząc od modelu ARIMA do modelu funkcji transferu.

2.2.4. Modele stochastyczne Funkcji Transferu

Przykładem innych modeli stochastycznych, stosowanych w analizie danych imisyjnych, są modele funkcji przejścia. Stosuje się je w przypadkach, gdy między dwoma seriami czasowymi Y_t i X_t istnieje dynamiczny związek dający się przybliżyć równaniem postaci:

$$Y_t = u_0 X_t + u_1 X_{t-1} + \dots = u(B)X_t \quad (2.6)$$

W tym równaniu $u(B)$ jest funkcją transferu, czy też funkcją przejścia. Jak widać wartość zmiennej Y_t jest wyliczana nie w oparciu o charakterystykę jej szeregu czasowego jak w przypadku modeli ARIMA, lecz na podstawie szeregu czasowego innej zmiennej. Podstawą identyfikacji funkcji transferu są wzorce korelacji wzajemnej zmiennych. Wymogiem stosowania modelu jest jednak stacjonarność zmiennych. Stąd konieczność wybielania surowych serii czasowych stężeń przed przystąpieniem do oceny ich współzależności. Zaletą podejścia, które korzysta ze wspólnego wzorca zmienności ujawniającego się w dwóch seriach czasowych jest możliwość znalezienia faktycznego związku przyczynowego między tymi seriami.

Chock, 1995 zwrócił uwagę, że model z funkcją transferu stanowi alternatywę dla modelu regresji, którego nie można zastosować w przypadkach wysokiego poziomu autokorelacji w seriach czasowych stężeń. Spowoduje to bowiem uzyskanie nieefektywnych estymatorów parametrów modelu. W tych sytuacjach ocena realnego wpływu wybranego czynnika zewnętrznego na zmienną modelowaną jest możliwa za pomocą modelu z funkcją transferu. Przykładem takiego zastosowania są badania Milionis, 1994, który stosując model z funkcją transferu dowiódł, że średniodobowe stężenie dymu silniej zależy od prędkości wiatru niż od grubości warstwy inwersyjnej, a rzeczywisty wpływ obu tych czynników na stężenie pyłu jest bardzo niski.

2.2.5. Model przestrzeni stanów z filtrem Kalmana

Na powietrze atmosferyczne można patrzeć jako na dynamiczny system, który da się opisać za pomocą stanów, przyjmowanych przez niego w kolejnych chwilach czasu. Stany systemu charakteryzuje pewien zestaw parametrów. Natomiast źródłem informacji o stanach systemu jest pomiar wybranej wielkości charakteryzującej system. Tak rozumiany system daje się opisać dwoma równaniami: stanu i pomiaru. Postać równań jak i zestaw zmiennych, które w nich występują zależą od aspektu systemu, który interesuje badacza. Równanie stanu wiąże przyszły (prognozowany) stan systemu ze stanem bieżącym, a równanie pomiaru umożliwia porównanie stanu wyliczonego ze stanem zmierzonym. Na podstawie rozbieżności dokonywana

jest korekta parametrów modelu, która powoduje jego adaptację do poziomu ostatniej wartości wielkości mierzonej. Skorygowany model stanu systemu jest podstawą prognozy stanu systemu w następnej chwili czasu. Korekta parametrów modelu w oparciu o ostatni pomiar przebiega za pośrednictwem techniki statystycznej znanej jako filtr Kalmana. Jest to technika modelowania adaptacyjnego, bardzo szeroko stosowania w nauce i technice. Umożliwia przetwarzanie pomiarów dotyczących systemu, w celu uzyskania oceny jego stanu obciążonej najmniejszym błędem, zgodnie z zadaniem kryterium optymalizacyjnym. Filtr Kalmana jest przykładem mieszanego podejścia do modelowania. Deterministyczny model prognostyczny podlega uaktualnianiu w oparciu o wykonywany w czasie rzeczywistym pomiar zmiennej charakteryzującej system. Uaktualnianie parametrów modelu na podstawie ostatniego pomiaru wielkości opisywanej jest zaś cechą modeli niedeterministycznych.

Wielkością, która stanowi przedmiot prognozy dotyczącej przyszłego stanu zanieczyszczenia powietrza atmosferycznego jest stężenie zanieczyszczenia. Uznajmy je również za wielkość mierzoną, charakteryzującą system oraz wskaźnik stanu modelu. Uzyskujemy dzięki temu najprostsze z rozwiązań zagadnienia prognozy stężeń zanieczyszczeń wykorzystujące filtr Kalmana.

Przyjęcie rzeczywistego stężenia zanieczyszczenia za wskaźnik stanu powietrza atmosferycznego powoduje, że równania stanu systemu (2.7) oraz pomiaru (2.8) przyjmują następującą postać:

$$X_{t+1} = H_t X_t + A_t S_t + G_t e_t \quad (2.7)$$

$$C_t = F_t X_t + r_t \quad (2.8)$$

Gdzie:

- X_t - wektor stanu systemu, zawiera informację o wszystkich przeszłych oraz bieżącym stężeniu zanieczyszczenia
- S_t - reprezentuje oddziaływania zewnętrzne (jak zmienne meteorologiczne oraz emisję i pochłanianie zanieczyszczeń)
- e_t - błąd modelowania (zakłada się, że jest szumem białym)
- H_t, A_t, G_t - macierze przejścia
- C_t - mierzone stężenie zanieczyszczenia
- F_t - macierz przejścia między rzeczywistymi i obserwowanymi wartościami zmiennej
- r_t - błąd pomiaru (zakłada się, że jest szumem białym)

Jak wynika z równania (2.7) na przyszły stan systemu, którym jest stężenie zanieczyszczenia, wpływają bieżąca i poprzednie wartości stężeń zanieczyszczeń, oddziaływania zewnętrzne jak parametry meteorologiczne, emisja i pochłanianie zanieczyszczeń oraz zaburzenie o charakterze szumu białego. Prognoza przyszłego stanu systemu X_{t+1} przebiega na podstawie skorygowanego o pomiar C_t stanu bieżącego X_t (2.8) i wartości bieżących oddziaływań

zewnętrznych S_t . Zakładając, że rozważany system jest dyskretnym systemem liniowym, który można opisać układem równań (2.7 i 2.8), filtr Kalmana umożliwia wyliczenie prognozy wektora stanu systemu X_{t+1} , o minimalnej wartości wariancji błędu prognozy.

Modelowanie w przestrzeni stanów z zastosowaniem tak rozumianego filtra Kalmana zrealizował Hernandez, 1992 dla średniodobowych stężeń pyłu. Uzyskane wyniki wyraźnie przewyższają rezultaty modelu ARIMA. Hernandez zaproponował modelowanie w przestrzeni stanów z zastosowaniem filtra Kalmana jako metodę modelowania stężeń pyłu w powietrzu.

Proszę zwrócić uwagę na możliwość zastosowania innego podejścia do zagadnienia prognozy stężeń zanieczyszczeń z wykorzystaniem filtra Kalmana. Jak łatwo zauważyć, w rozwiązaniu proponowanym powyżej, wpływ oddziaływań zewnętrznych na stężenie zanieczyszczenia ma charakter ustalony i wyraża się niezmienną macierzą przejścia A_t . Rezultaty analiz wskazują, że w rzeczywistości jest on zmienny, a uwzględnienie tego faktu umożliwia poprawę jakości prognozy stężenia [Wall, 2000]. Według modelu zaproponowanego przez van der Wall, 2000, poziom oddziaływania czynników zewnętrznych na stężenia zanieczyszczeń stanowi miernik stanu systemu. Równanie stanu (2.9) przedstawia związek między bieżącymi a przyszłymi wartościami poziomu oddziaływania tych czynników. Natomiast równanie pomiaru (2.10) pokazuje jak poziom stężeń odchyła się od poziomu zadanego trendem i wahaniami sezonowymi wskutek oddziaływania czynników zewnętrznych.

$$a_i(t) = T \cdot a_i(t-1) + \eta(t) \quad (2.9)$$

$$T(t) = \mu(t) + \gamma(t) + a_1(t)x_1(t) + a_2(t)x_2(t) + a_3(t)x_3(t) + \dots + \xi(t) \quad (2.10)$$

Gdzie:

$a_i(t)$ - współczynniki przy kolejnych zmiennych objaśniających

$T(t)$ - zmienna objaśniana

$x(t)$ - zmienne objaśniające

$\mu(t)$ - trend

$\gamma(t)$ - sezonowość

Sprzężenie równań odbywa się w ten sposób, że wartość wielkości mierzonej wyliczana na podstawie parametrów stanu na daną chwilę czasu jest porównywana z rzeczywistą wartością mierzoną. W oparciu o błąd oceny wielkości mierzonej, dokonywana jest korekta parametrów stanu. Skorygowane parametry stanu służą do prognozy wartości mierzonej w następnym kroku czasowym. Dzięki uwzględnieniu w modelu zmienności parametrów stanu van der Wall uzyskał poprawę wyjaśniania przez model wariancji średniodobowych stężeń pyłu z 25% na 45%, w stosunku do zwykłego modelu regresji.

W rozwiązaniu zaproponowanym powyżej parametry stanu to współczynniki przy zmiennych objaśniających (parametrach meteorologicznych) w równaniu regresji opisującym

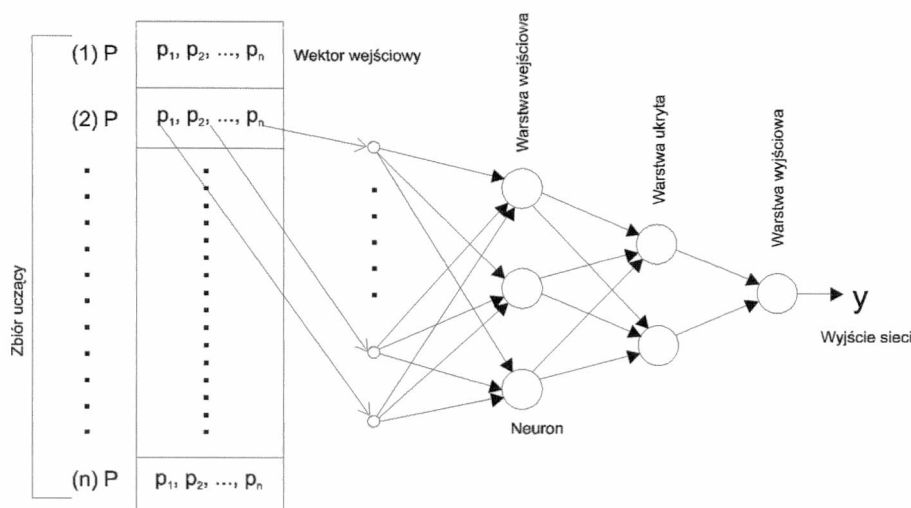
stężenie zanieczyszczenia, którym pozwolono zmieniać się w czasie i adaptować do zmian zmiennej objaśnianej (stężenia zanieczyszczenia). Tak zastosowany filtr Kalmana można postrzegać jako narzędzie uelastycznienia modelowania stężeń zanieczyszczeń z wykorzystaniem równania regresji. Sam w sobie filtr Kalmana jest jednak techniką stochastyczną.

2.3. Metody neuronowe

Sieci neuronowe, nazywane również neuronalnymi stanowią grupę stosunkowo nowych obliczeniowych metod pozyskiwania informacji z danych liczbowych. Początki teorii sieci neuronowych sięgają lat 40-tych. Bardzo intensywny rozwój tych metod trwa jednak od początku lat 80-tych, kiedy opracowano metodę uczenia sieci wielowarstwowej określaną jako metoda propagacji wstecznej błędu. Ponadto dzięki rozwojowi maszyn liczących nastąpił wyraźny wzrost dostępnych mocy obliczeniowych, umożliwiającą realizację sztucznych sieci neuronowych.

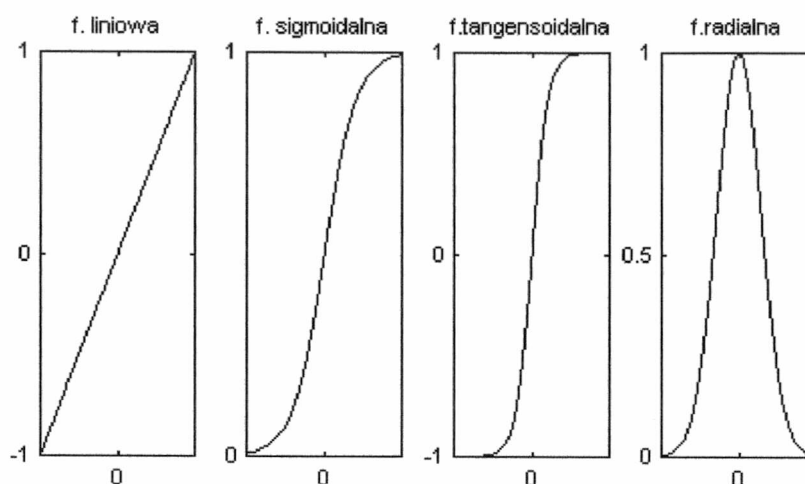
Sieć neuronowa może być wykorzystywana do rozwiązywania bardzo wielu problemów z zakresu obróbki danych. Jednak ze względu na zakres poruszanych zagadnień w polu zainteresowania tej pracy znajdują się wyłącznie sieci w roli aproksymatorów odwzorowań i klasyfikatorów obiektów. Problem aproksymacji funkcji jest bardzo powszechny w nauce i technice. Jeśli postać funkcji jest znana, zagadnienie to rozwiązuje się zwykle za pomocą metod analitycznych lub numerycznych, za wyjątkiem przypadku, gdy postać funkcji jest niezwykle skomplikowana. Domniemania na temat istnienia i postaci związku pozwalają posłużyć się metodami statystycznymi. Natomiast brak wiedzy o możliwej postaci zależności, który nie pozwala na skorzystanie z wymienionych metod, otwiera drogę zastosowaniu sieci neuronowych. Teoretycznie można dobrać sieć neuronową, która z dowolną dokładnością odzwierciedli w swojej strukturze przekształcenie jednego zestawu danych w inny zestaw, pod warunkiem, że to przekształcenie jest różnowartościowe. Ta własność sieci jest niezwykle cenna, ponieważ do aproksymacji odwzorowania potrzebują one wyłącznie danych z dziedziny funkcji i odpowiadających im wartości z jej przeciwdziedziny. Sieć jest w stanie nauczyć się zależności występującej między tymi dwoma zbiorami danych, a następnie posłużyć do znajdowania wartości funkcji odpowiadających dowolnym argumentom z jej dziedziny. Jednakże analityczna postać zależności pozostaje niejawna i obecnie nie istnieją metody jej identyfikacji na podstawie struktury sieci neuronowej. Na tym polega niższość sieci neuronowych w stosunku do metod analitycznych, numerycznych czy statystycznych. Będąc metodą bardzo skuteczną, sieci neuronowe stanowią jednocześnie "czarną skrzynkę". Sytuacja zmienia się w przypadku, gdy odwzorowanie nie jest różnowartościowe. Wówczas osiągnięcie przekształcenia o dowolnej dokładności nie jest możliwe, nie zapewni tego zresztą żadna metoda. W ogólnym przypadku, rozwiązanie podane przez sieć

będzie minimalizowało średni błąd kwadratowy przekształcenia dziedziny odwzorowania w jego przeciwdziedzinę. Jest to uwarunkowane kryterium optymalizacji sieci.



Rysunek 2.3. Schemat sieci neuronowej

Model neuronowy nie ma postaci równania lecz raczej jest strukturą. Składa się z elementów obliczeniowych nazywanych neuronami (rys.2.3). Każdy neuron pobiera informację liczbową przez wejścia. Każde z wejść neuronu ma swoją wagę. Neuron może mieć wiele wejść, lecz tylko jedno wyjście, wyposażone w tzw. funkcję przejścia (transferu). Neuron wykonuje operacje liczbowe mnożenia liczb, które pojawiły się na jego wejściach przez wagi wejść, sumowania tak otrzymanych iloczynów i wyznaczenia wartości funkcji transferu dla tej sumy. Jeżeli funkcja transferu jest liniowa neuron potrafi wykonywać wyłącznie przekształcenia liniowe. Jeżeli funkcja transferu ma charakter nieliniowy, neuron może odwzorowywać zależności nieliniowe. Najczęściej stosowane nieliniowe funkcje transferu to funkcja sigmoidalna i tangensoidalna. Szczególnym przypadkiem nieliniowej funkcji przejścia jest funkcja radialna (rys.2.4).



Rysunek 2.4. Przebieg zmienności różnych funkcji transferu w zależności od wartości ich argumentu

Sieć neuronowa składa się z wielu neuronów (rys.2.3). Są one zorganizowane w warstwy. Od jednej do wielu. Zasadniczo neurony jednej warstwy nie komunikują się między sobą poza przypadkami sprzężenia zwrotnego. Natomiast kontakt neuronów sąsiednich warstw jest istotą rozwiązania neuronowego. Wyjścia neuronów poprzedniej warstwy podają wartości liczbowe na wejścia neuronów kolejnej warstwy. Pierwsza warstwa neuronów sieci to warstwa wejściowa. Ta warstwa przyjmuje dane z dziedziny funkcji aproksymowanej siecią. Liczba wejść neuronów w warstwie wejściowej odpowiada liczbie zmiennych wejściowych, po jednej zmiennej na każde wejście. Liczba wyjść z sieci neuronowej, czyli liczba neuronów w warstwie wyjściowej sieci jest równa liczbie zmiennych należących do przeciwdziedziny funkcji. Najczęściej wykorzystywane są sieci z jednym wyjściem. Nie jest to jednak reguła. Między warstwami wejściową i wyjściową znajdują się tzw. ukryte warstwy neuronów. One nie mają kontaktu z danymi pochodzącymi z zewnątrz, a jedynie z wyjściami i wejściami neuronów należących do innych warstw [Gardner,1998].

Warunkowana swoją strukturą sieć działa w następujący sposób. Elementy warstwy wejściowej pobierają informację z zewnątrz, równolegle wykonują obliczenia i przekazują wyniki elementom kolejnej warstwy. Te, również równolegle, przeliczają dane według swoich wag i funkcji transferu, a rezultaty komunikują z następną warstwą. Informacja liczbową jest w ten sposób propagowana przez kolejne warstwy sieci aż do warstwy wyjściowej, gdzie pojawia się ostateczny rezultat, odpowiedź całej sieci na przedstawiony jej zestaw danych wejściowych.

Zasadniczym etapem budowy modelu neuronowego jest proces uczenia sieci neuronowej, który polega na dostosowaniu wag neuronów do danych zawartych w zbiorze uczącym. Odpowiednikiem tego procesu w przypadku metod statystycznych jest wyliczanie parametrów modelu. Zmierzają one do uzyskania takiego stanu sieci, w którym jej odpowiedź na dowolny zestaw danych z dziedziny funkcji nie różni się od wartości funkcji aproksymowanej bardziej, niż o zadany błąd. Wymaga to korekty wag sieci zmierzającej do minimalizacji tego błędu. Miarą różnicy między odpowiedzią sieci a wartością rzeczywistą jest najczęściej średni błąd kwadratowy. Procedura korekty wag sieci neuronowej prowadząca do spadku błędu odtwarzania wartości modelowanych siecią neuronową nosi nazwę wstecznej propagacji błędu. Wsteczna propagacja błędu rozkłada odpowiedzialność za niego między poszczególne neurony korygując ich wagi tak, że odpowiedź sieci wyliczona na podstawie nowych wag jest bliższa wartościom rzeczywistym. Jeżeli sieć uczona jest w tzw. trybie "batch", korekta wag następuje po przedstawieniu sieci całego zbioru uczącego. Podczas uczenia adaptacyjnego wagi ulegają

korekcie po prezentacji sieci każdego zestawu danych uczących.

Bardzo ważną kwestią związaną z uczeniem sieci neuronowej jest zagadnienie końca procesu uczenia. Okazuje się, że rozwiązanie odpowiadające zminimalizowanemu błędowi uczenia sieci nie jest optymalnym rozwiązaniem. W takich warunkach sieć zwykle osiąga nadmierne dopasowanie do zbioru uczącego i traci zdolności generalizowania, a właśnie zdolność sieci do generalizacji jest najbardziej interesująca. Sieć, która osiągnęła zdolność generalizowania poprawnie reaguje na zestawy danych pojawiające się w zbiorze testowym, choć nie były jej prezentowane wcześniej. Dzięki własności generalizacji sieć przekształca w sposób ciągły dziedzinę odwzorowania w jego przeciwdziedzinę mimo to, że zapoznano ją wyłącznie z elementami należącymi do obu tych zbiorów [Ossowski,1996; Masters,1996; Tadeusiewicz,1993].

Poniżej, zostanie przedstawiona krótka charakterystyka typów sieci neuronowych wykorzystanych w niniejszej pracy.

2.3.1. Liniowy filtr adaptacyjny, sieć z liniową funkcją transferu i sieć z nieliniową funkcją transferu

Liniowy filtr adaptacyjny jest zbudowany z neuronu o liniowej funkcji transferu. Wykorzystuje się go do filtracji sygnału, a więc obróbki czasowej. Na wejście filtra podawanych jest kilka zarejestrowanych po sobie wartości zmiennej, a na wyjście wartość następująca zaraz po nich. Długość okna wejściowego, które przesuwają się po szeregu czasowym stanowi parametr zadawany przez użytkownika filtra. Im dłuższe okno, tym dłuższa historia zmiennej jest uwzględniana jako mająca wpływ na jej bieżącą wartość. Korekta wag sieci następuje po prezentacji każdej pary wektor wejściowy-wektor wyjściowy i minimalizuje odległość między rzeczywistą wartością zmiennej a wartością wyliczoną przez filtr. Filtr adaptuje się więc do ostatniego prezentowanego mu zestawu danych uczących. Stąd zależność między ostatnim wektorem wejściowym i przypisaną mu wartością wyjścia ma największy wpływ na kształt wag jak też na wynik prognozy.

Sieć z liniową funkcją transferu w odróżnieniu od neuronowego filtra liniowego może odwzorowywać w siebie wartości należące do szeregu czasowego zmiennej, a rozsunięte w czasie bardziej niż o jeden krok czasowy. Za pomocą tej sieci można realizować przekształcenia liniowe wielu zmiennych opisujących w jedną zmienną opisywaną lub wielu zmiennych opisujących w wiele zmiennych opisywanych. Dopuszczalna jest nauka sieci w trybie adaptacyjnym, jak w przypadku neuronowego filtra liniowego. Najczęściej stosowany jest jednak tryb batch.

W tym trybie nauki korekta wag sieci następuje po przedstawieniu jej całego zbioru uczącego w oparciu o sumaryczny błąd wyliczony dla całego tego zbioru. Kolejne wektory uczące nie mają więc rosnącego wpływu na wartość wag sieci. Wszystkie są równouprawnione. Sieć neuronowa z liniowymi funkcjami transferu jest zawsze jednowarstwowa i składa się z tylu neuronów, ile wyjść potrzeba, czyli ile jest zmiennych modelowanych.

Do rozwiązywania problemów nieliniowych służy sieć zbudowana z neuronów wyposażonych w nieliniowe funkcje transferu np. sigmoidalną lub tangensoidalną. Trening sieci przebiega według metody propagacji wstecznej. Możliwe jest uczenie w trybie adaptacyjnym, aczkolwiek najczęściej stosowane są sieci z neuronami nieliniowymi uczone w trybie batch. Kryterium optymalizacji sieci jest średni błąd kwadratowy popełniany przy wyliczaniu siecią wartości zmiennej modelowanej podczas kolejnych prezentacji zbioru uczącego. Sieć może się składać z wielu warstw. Dwie warstwy ukryte o odpowiednio dobranych liczbach neuronów powinny wystarczyć do rozwiązania dowolnego problemu nieliniowego.

Zarówno sieci z liniowymi, jak nieliniowymi funkcjami przejścia są powszechnie stosowane w prognostyce stężeń zanieczyszczeń [Comrie,1997; Gardner,1998; Gardner,2000; Guardani,1999; Baran1999; Jorquera,1998; Kao2000; Kolehmainen,2001; Melas2000; Ośródk,1995; Perez,2000; Soja,200; Yu2000]. Wśród cytowanych rozwiązań znaleźć można przykłady zarówno niedeterministycznego jak deterministycznego podejścia do prognozy stężeń. Przedmiotem opisu są najczęściej O_3 , NO_2 , NO_x ze względu na ich nieliniowe interakcje z warunkami meteorologicznymi [Gardner,1998; Gardner,2000; Comrie,2000; Guardani,1999; Melas,2000; Jorquera,1998]. Prognozą stężeń 30-minutowych SO_2 wokół dużego punkтового źródła emisji zajmowała się natomiast Boznar,1993,1995,1996. Przyczyną były niewystarczające, do sterowania pracą elektrowni rezultaty modelu dyfuzyjnego. Modelowanie stężeń SO_2 w układzie niedeterministycznym prowadził również Kao,2000. Jego celem było porównanie rezultatów sieci neuronowej z wynikami modelu ARIMA. Interesujące rozważania dotyczące neuronowej prognozy stężeń pyłu $PM_{2,5}$ przedstawił Perez,2000.

Liniowy filtr adaptacyjny, sieć neuronową o liniowej funkcji przejścia i sieć zbudowaną z neuronów o nieliniowych funkcjach transferu wykorzystano w pracy do modelowania szeregu czasowego stężeń zanieczyszczeń.

2.3.2. Sieci z radialną funkcją transferu

Sieci radialne, są to sieci o ściśle określonej strukturze. Składają się z dwóch warstw: warstwy wejściowej, z neuronami o radialnych funkcjach przejścia, oraz warstwy wyjściowej,

zbudowanej z neuronów liniowych. Podstawowym czynnikiem wyróżniającym te sieci jest radialna funkcja transferu. Najczęściej stosowaną funkcją radialną jest funkcja Gaussa (rys.2.4) gdzie $1/2\delta$ jest parametrem decydującym o promieniu otoczenia wektora wag neuronu, na które ten neuron odpowiada pozytywnie. Jest to tzw. spread. Sieci radialne realizują inne podejście do wektora danych niż funkcje liniowe, sigmoidalne czy tangensoidalne. Podczas gdy te ostatnie operują na iloczynach skalarnych wag neuronów i wektorów wejściowych $W \cdot P$, sieci radialne posługują się odległością wektora wejściowego od wektora wag neuronu $|P-W|$. Dzięki temu neurony z radialną funkcją transferu dzielą przestrzeń danych kołowo, nie zaś liniowo jak neurony z funkcją liniową, sigmoidalną lub tangensoidalną. Przedstawiono to symbolicznie na rys.2.5 [Ossowski,1996]. Zgodnie z funkcją Gaussa występującą najczęściej w roli funkcji przejścia neuronów radialnych udzielają one pozytywnej odpowiedzi, gdy spełniony jest warunek podany na rys.2.5. Według funkcji sigmoidalnej, najbardziej popularnej spośród pozostałych rozwiązań, przestrzeń danych jest dzielona liniowo według kryterium podanego na rys.2.5.

$$\begin{array}{ccc}
 e^{-\left(\frac{W-P}{2\delta}\right)^2} & & \frac{1}{1+e^{-\left(W \cdot P\right)}} \\
 \text{0} & \text{0} & \text{0} \\
 \text{0} & \text{+} & \text{0} \\
 \text{0} & \text{0} & \text{0}
 \end{array}$$

Rysunek 2.5. Podział przestrzeni danych przez neuron z radialną funkcją przejścia i neuron z sigmoidalną funkcją przejścia.

Maksimum odpowiedzi neuronu z radialną funkcją przejścia wypada wówczas, gdy wektor wejściowy jest równy wektorowi wag neuronu. Im większa odległość między tymi wektorami tym mniejsza odpowiedź neuronu. Sieć radialna w naturalny sposób realizuje aproksymację lokalną w przeciwieństwie do sieci z sigmoidalną funkcją przejścia, której odpowiedź rozciąga się do nieskończoności i która lepiej nadaje się do aproksymacji globalnej. Sieć radialna jest zwykle większa, składa się z większej liczby neuronów niż sieć sigmoidalna. W przypadku granicznym, każdemu wektorowi wejściowemu może odpowiadać jeden neuron. Natomiast proces uczenia sieci radialnej przebiega szybko, co wynika stąd, że de facto w zbiorze uczącym znajduje się informacja dotycząca położenia wag neuronów. Dzięki temu proces uczenia startuje w dobrym punkcie i jest szybciej zbieżny niż w przypadku sieci z inną funkcją transferu. W praktyce dąży się do zmniejszenia liczby neuronów i takiego dobrania ich parametru spread, aby czynne obszary funkcji transferu wszystkich neuronów pokrywały również całą przestrzeń leżącą między

centrami, nadając sieci w ten sposób własności generalizujące. Wartość spread dobierana jest w ten sposób, by była mniejsza od odległości między skrajnymi wektorami w zbiorze uczącym, a większa niż odległość między najbliższymi wektorami sąsiadującymi [Ossowski, 1996].

W pracy podjęto próbę zastosowania sieci radialnych do prognozowania chwilowych stężeń zanieczyszczeń w oparciu o zestawy warunków meteorologicznych. Takie podejście uzasadniała interpretacja neuronu jako reprezentanta sytuacji meteorologicznej, której skutkiem jest określone stężenie zanieczyszczenia. Jeżeli sytuacja testowa jest bliska sytuacji reprezentowanej przez neuron, odpowiedź w postaci wartości stężenia również powinna być bliska rzeczywistej. Ze względu na zdolność sieci radialnych do aproksymacji lokalnej zaproponowano i przedstawiono w pracy propozycję rozwiązania zadania interpolacji przestrzennej chwilowych stężeń zanieczyszczeń z ich zastosowaniem. Jest to rozwiązanie nowatorskie nie opisywane dotychczas w literaturze.

2.3.3. Sieci samoorganizujące

Idea sieci samoorganizujących polega na tym, że sieć samodzielnie wyróżnia w zbiorze wejściowym grupy wektorów podobnych do siebie. Jest to możliwe dzięki metodzie uczenia konkurencyjnego, która polega na wzmacnianiu połączeń neuronu z tymi wejściami, które są zbliżone do wektora wag neuronu. W rezultacie neuron uczy się reakcji na określony typ wektorów wejściowych. Poszczególne neurony sieci, różniące się swoimi zestawami wag odpowiadają pozytywnie na różne typy wektorów wejściowych. Efektem działania sieci samoorganizującej jest podzielenie zbioru wektorów wejściowych na grupy, które nazwać można klasami lub wzorcami, czyli klasyfikacja zbioru wejściowego. Szczególnym przypadkiem sieci samoorganizującej jest mapa cech Kohonena SOM (Self Organizing Map).

Kolehmainen, 2000 zastosował SOM do prognozy stężeń godzinnych NO_2 dla porównania skuteczności tego rozwiązania z perceptronem wielowarstwowym. Drugie rozwiązanie okazało się lepsze.

Bardzo ciekawą propozycję wykorzystania SOM przedstawił Kamiński, 1999. Mapa samoorganizująca posłużyła do wytypowania lokalizacji stacji pomiarowych w ten sposób, by wykonywane pomiary umożliwiły dokładną reprezentację pola stężeń zanieczyszczeń na terenie Łodzi. Analizy wykonano dla SO_2 i pyłu. Stwierdzono, że rozmieszczenie punktów pomiarowych tych zanieczyszczeń powinno być nieco inne od występującego obecnie.

Przeprowadzenie klasyfikacji danych przez sieć neuronową stwarza następnie możliwość przeprowadzania kwalifikacji nowych zbiorów danych do wcześniej wyróżnionych klas. Jest to

tw. kwestia rozpoznawalności przynależności wektora do wzorca. Ten problem można rozwiązać z wykorzystaniem sieci neuronowej Kohonena typu LVQ (Learning Vector Quantization).

LVQ jest metodą "uczenia z nauczycielem" warstwy neuronów współzawodniczących. Sieć tą również wprowadził Kohonen. Uczy się ona klasyfikować wektory wejściowe na zasadzie ich euklidesowej bliskości do narzuconego wektora wyjściowego. Sieć realizuje odwzorowanie wektora liczbowego na wektor zerojedynkowy, który jest identyfikatorem klasy przynależności wektora wejściowego.

Według Malinowskiego, 1999 interesująco przedstawiają się wyniki klasyfikacji rozkładów przestrzennych 30-minutowych stężeń zanieczyszczeń na terenie Czarnego Trójkąta. Wzorce są rozpoznawalne siecią LVQ.

3. Badania własne

3.1. Charakterystyka obszaru źródłowego danych

Modele neuronowe przedstawione w pracy zbudowano dla danych o stężeniach imisyjnych zanieczyszczeń pochodzących z automatycznej sieci monitoringu zanieczyszczeń powietrza Elektrowni Turów, wykonanej według projektu Zakładu Ochrony Atmosfery Politechniki Wrocławskiej. Sieć pracuje na terenie Worka Żytawskiego i ma za zadanie kontrolować wpływ Elektrowni Turów na stan zanieczyszczenia powietrza wokół zakładu.

Topografia Worka Żytawskiego

Worek Żytawski znajduje się w południowo-zachodnim krańcu Polski, gdzie zbiegają się granice Polski, Czech i Niemiec (rys.3.1). Geograficznie, położony jest między 14°50"-15°E oraz 50°50"-51°N w Obniżeniu Żytawsko-Zgorzeleckim. Z trzech stron otaczają go góry, oraz wysokie wzniesienia. Są nimi Góry Izerskie od wschodu, Góry Łużyckie od południa i Pogórze Łużyckie od zachodu. Teren samego Worka Żytawskiego jest łagodnie pofałdowany z centralnie położonym wzniesieniem zwanym Górką Działoszyńską. W topografii terenu wyraźnie zaznacza się wyrobisko Kopalni Węgla Brunatnego Turów oraz zwałowisko zewnętrzne. Wyrobisko, z którego wydobywany jest węgiel brunatny, zajmuje powierzchnię ok. 8km² i sąsiaduje z Bogatynią od strony zachodniej. Zwałowisko rozciąga się na powierzchni ok. 10km² i jest usypywane na północ od Bogatyni z nadkładu kopalnianego oraz popiołu z elektrowni. Wspólnie zajmują one prawie połowę powierzchni Worka Żytawskiego.

Wzdłuż granicy posko-niemieckiej płynie Nysa Łużycka o rozmiarach górskiego potoku średniej wielkości. Poza nią, istotniejsze ciała wody stanowią zbiornik zaporowy Witka, położony u wylotu Worka Żytawskiego i będący rezerwuarem wody dla Elektrowni Turów, oraz zbiornik pośredni Elektrowni, znajdujący się w jej sąsiedztwie.

Obszar Worka Żytawskiego w równej mierze porastają drzewa, krzewy i trawy. Trudno wyróżnić dominujące zbiorowiska roślinne. Prawdopodobnie wraz z postępem rekultywacji zwałowiska stopniowo będzie się uwypuklała przewaga lasów.

Osadnictwo

Największym skupiskiem ludności na terenie Worka Żytawskiego jest miasto Bogatynia. Poza nim miejscowa ludność zamieszkuje niewielkie wsie i osady, liczące po kilka domostw,

jak Jasna Góra, Opolno Zdrój, Rybarzowice, Białopole, Kopaczów, Porajów, Sieniawka, Biedrzychowice Grn., Turoszów, Trzciniec, Działoszyn, Wyszaków, Wigancice, Wolanów, Posada, Bratków, Kostrzyna, Krzewina, Ręczyn, Spytków, Radomierzyce, Niedów, Wilka. Za wyjątkiem Bogatyni, Turoszowa i Działoszyna położonych centralnie, pozostałe jednostki osadnicze znajdują się głównie na południu i północy Worka Żytawskiego.

Warto zauważyć, że Work Żytawski charakteryzujący się skąpym osadnictwem o dużym rozproszeniu, znajduje się w otoczeniu trzech dużych aglomeracji miejskich. Są to Zittau po stronie niemieckiej (8km na zachód od Bogatyni), Liberec (12km na pd-wsch. od Bogatyni) i Frydlant (7km na wschód od Bogatyni), oba po stronie czeskiej. Ich wpływ na jakość powietrza w Worku Żytawskim jest bez wątpienia odczuwalny.

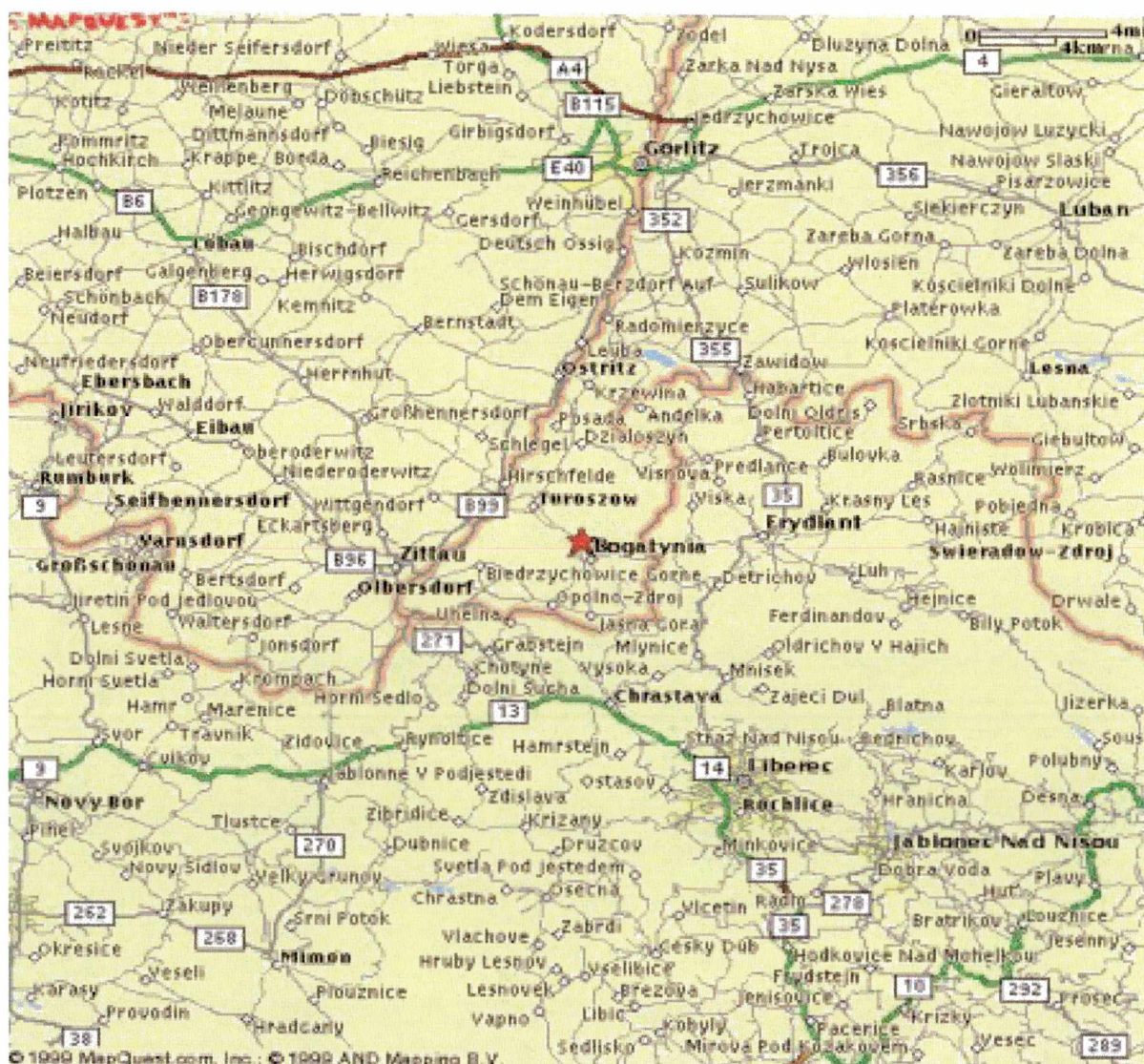
Przemysł

Aktywność przemysłowa na terenie Worka Żytawskiego jest zorientowana na wydobycie węgla brunatnego z lokalnego złoża i produkcję energii elektrycznej w wyniku spalania wydobytego węgla. Głównymi zakładami przemysłowymi są: Kopalnia Węgla Brunatnego Turów oraz Elektrownia Turów w Turoszowie.

Produkcja energii elektrycznej wiąże się z równoczesnym wytwarzaniem zanieczyszczeń powietrza. Powstają one na etapie wydobycia i transportu węgla, spalania węgla oraz transportu, a następnie składowania nadkładu i popiołu. Biorąc pod uwagę aspekt ilościowy podstawowymi zanieczyszczeniami są tu pył i popiół lotny oraz zanieczyszczenia gazowe jak SO_2 , NO_x i CO.

W związku z działaniami proekologicznymi i polityką Elektrowni Turów oraz KWB Turów w tym zakresie faktyczna emisja powstających zanieczyszczeń do atmosfery jest silnie zredukowana. Warto podkreślić, że oba zakłady należą do wzorcowych pod względem dbałości o ochronę środowiska naturalnego przed skutkami swojej działalności. Szeroki zakres ich przedsięwzięć koncentruje się na ograniczaniu emisji zanieczyszczeń u źródła oraz ciągłej kontroli stanu zanieczyszczenia powietrza na terenie Worka Żytawskiego.

W bezpośrednim sąsiedztwie Worka Żytawskiego, po stronie niemieckiej również zlokalizowano elektrownie. Są to elektrownie Bogsberg i Schwartze Pumpe. Ze względu na, przeważający w tym obszarze, południowo-zachodni kierunek wiatru emisja zanieczyszczeń z tych elektrowni oraz innych położonych w głębi Dolnej Saksonii oraz na Słowacji oddziałuje na stan zanieczyszczenia powietrza w Worku Żytawskim. Napływ zanieczyszczeń emitowanych przez te źródła formuje tło adwekcyjne, na które nakłada się oddziaływanie Elektrowni Turów i KWB Turów.



Rysunek 3.1. Położenie Worka Żytawskiego u zbiegu granic Polski, Czech i Niemiec

3.1.1. Zakładowa Sieć Monitoringu Powietrza Atmosferycznego Elektrowni Turów

Jednym z wyrazów troski o rzetelną informację dotyczącą jakości powietrza atmosferycznego w Worku Turoszowskim jest Zakładowa Sieć Monitoringu Powietrza Elektrowni Turów. Wyniki pomiarów prowadzonych przez stacje sieci umożliwiają kompleksową ocenę jakości powietrza atmosferycznego na terenie Worka Żytawskiego. Stacji jest osiem, ich lokalizację pokazano na rys.3.2. W obrębie Worka Żytawskiego rozmieszczono stacje Sieniawka, Jasna Góra, Bogatynia, Działoszyn, Wyszków, Witka i Radomierzyce. Ostatnia stacja, pracująca w Zgorzelcu została pomyślana jako wskaźnik wpływu Elektrowni Turów na obszary znajdujące się poza Workiem Żytawskim w odległości kilkuset kilometrów od niego. Stacje monitoringowe są rozmieszczone w miarę równomiernie za wyjątkiem Witki i Radomierzyc, które znajdują się blisko siebie. Przeciętnie na jedną stację przypada więc obszar ok. 5km².



Rysunek 3.2. Rozmieszczenie stacji pomiarowych Zakładowej Sieci Monitoringu Powietrza Atmosferycznego Elektrowni Turów

Sieć rozpoczęła pomiary w połowie 1996 roku. Obecnie dostępna jest więc dokumentacja Worka Żytawskiego w zakresie stanu zanieczyszczenia powietrza atmosferycznego za okres ponad czterech lat. Wszystkie stacje wykonują pomiary stężeń zanieczyszczeń wytypowanych jako reprezentatywne dla obszaru na podstawie kryteriów:

- powszechność występowania,
- osiąganie odpowiednio wysokich wartości stężeń,
- wspólne pochodzenie [Rutkowski, 1989].

Do zanieczyszczeń monitorowanych należą pył zawieszony, SO_2 , NO_2 , NO oraz dodatkowo CO i O_3 na stacji Działoszyn. Pomiary odbywają się automatycznie w sposób ciągły za pomocą automatycznych analizatorów firmy HORIBA. Rejestracji podlegają wartości stężeń uśrednione dla okresu 30 minut w jednostkach $[\mu\text{g}/\text{m}^3]$. Oprócz stężeń zanieczyszczeń reprezentatywnych pomiarami objęte są również charakterystyczne parametry meteorologiczne. Są to: prędkość $[\text{m}/\text{s}]$ i kierunek wiatru $[\text{deg}]$, temperatura $[\text{°C}]$ i wilgotność powietrza $[\%]$ oraz ciśnienie atmosferyczne $[\text{hPa}]$. Rejestracji podlegają 30-minutowe wartości tych parametrów.

Dane pomiarowe zebrane ze stacji pomiarowych należących do Zakładowej Sieci Monitoringu Powietrza Atmosferycznego Elektrowni Turów wykorzystano w pracy do budowy prognostycznych modeli neuronowych i opracowania neuronowej metody interpolacji. Uzyskane rezultaty umożliwiły sformułowanie szeregu wniosków dotyczących uwarunkowań czasowej i przestrzennej zmienności stężeń zanieczyszczeń powietrza atmosferycznego nad obszarem Worka Turoszowskiego.

3.2. Kryteria oceny modelu i wyników modelowania

Ocena procesu modelowania wymaga przyjęcia kryteriów oceny modelu i wyników modelowania. Kryteria oceny wyników modelowania pokazują jak dalece wartości wyliczone odbiegają od wartości rzeczywistych. Natomiast kryteria oceny modelu mierzą, w jakim stopniu model jest zdeterminowany odtwarzać poprawnie rzeczywiste wartości zmiennej modelowanej.

W literaturze przedmiotu można znaleźć szereg miar poprawności rezultatów modelowania. Za miarodajny dla oceny wyników modelowania wykonanego w pracy przyjęto zestaw kryteriów oceniających wyniki modelowania:

- średni błąd kwadratowy,
- średni błąd względny,
- i zestaw kryteriów oceniających model:
- współczynnik determinacji,
- błąd modelu,
- różnica średniej z wartości rzeczywistych i średniej z wartości wyliczonych przez model.

Dla przypomnienia, poniżej zamieszczono podstawowe informacje na temat tych kryteriów.

1. Średni błąd kwadratowy

Średni błąd kwadratowy oblicza się według formuły określonej równaniem (3.1)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

RMSE jest liczbą wyrażoną w jednostkach zmiennej modelowanej. Pokazuje jak bardzo wyniki x_i modelowania odchylają się przeciętnie od rzeczywistych wartości zmiennej y_i . Średni błąd kwadratowy jest stały w całym zakresie zmienności zmiennej rzeczywistej. Oznacza to, że w stosunku do małych wartości zmiennej błąd średniokwadratowy jest duży, podczas gdy w stosunku do dużych wartości zmiennej błąd ten jest relatywnie mniejszy. Błąd średniokwadratowy jest bardzo dobrym i najczęściej stosowanym kryterium doboru wag w sieci neuronowej oraz kryterium wykorzystywanym do szacowania współczynników w równaniach regresji.

2. Średni błąd względny

Komplementarnym kryterium oceny wyników modelowania w stosunku do średniego błędu kwadratowego jest średni błąd względny. Zgodnie ze wzorem (3.2) określa on, jaką część rzeczywistej wartości zmiennej y_i stanowi średnio różnica między tą wartością a jej oszacowaniem x_i .

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{y_i} \right| \quad (3.2)$$

Każdy ułamek stanowiący element powyższej sumy jest błędem względnym popełnianym przy wyliczaniu konkretnej wartości zmiennej. Zwykle nie przyjmuje on wartości stałej w całym zakresie zmienności zmiennej rzeczywistej. Jest to możliwe jedynie w wyjątkowym przypadku. Warto zauważyć, że przy stałym liczniku wartość błędu względnego maleje ze wzrostem wartości zmiennej. Kryterium optymalizującym większość modeli statystycznych i neuronowych jest konstruowane w oparciu o błąd średniokwadratowy. Ze względu na to w wynikach modelowania pochodzących z tych modeli najczęściej obserwuje się większe wartości błędu względnego w zakresie małych wartości zmiennej rzeczywistej, niż w zakresie dużych wartości zmiennej.

3. Współczynnik determinacji

Podstawowym i najszerzej stosowanym wskaźnikiem statystycznym, który wykorzystuje się do oceny poprawności odtwarzania zmiennej zależnej przez model jest współczynnik determinacji. Wskaźnik został skonstruowany dla modelu regresji liniowej. Jest jednak powszechnie wykorzystywany także w przypadku innych modeli. Współczynnik determinacji informuje, jaka część wariancji zmiennej zależnej jest wyjaśniana przez zmienne niezależne uwzględnione w modelu. Pokazano, że wariancję całkowitą zmiennej zależnej można wyrazić jako sumę wariancji wyjaśnionej przez model opisujący tę zmienną i wariancji nie wyjaśnionej przez model.

$$\sum (y - \bar{y})^2 = \sum (x - \bar{y})^2 + \sum (y - x)^2 \quad (3.3)$$

wariancja	wariancja	wariancja
całkowita	wyjaśniona	niewyjaśniona

y - wartość rzeczywista zmiennej

x - wartość zmiennej z modelu

y - wartość średnia zmiennej rzeczywistej

Współczynnik determinacji jest stosunkiem wariancji zmiennej wyjaśnionej przez model do wariancji całkowitej zmiennej (równ.3.4).

$$R^2 = \frac{\sum (x - \bar{y})^2}{\sum (y - \bar{y})^2} \quad \text{lub} \quad R^2 = 1 - \frac{\sum (y - x)^2}{\sum (y - \bar{y})^2} \quad (3.4)$$

Formuła obliczeniowa na współczynnik determinacji ma postać opisaną równaniem (równ.3.5).

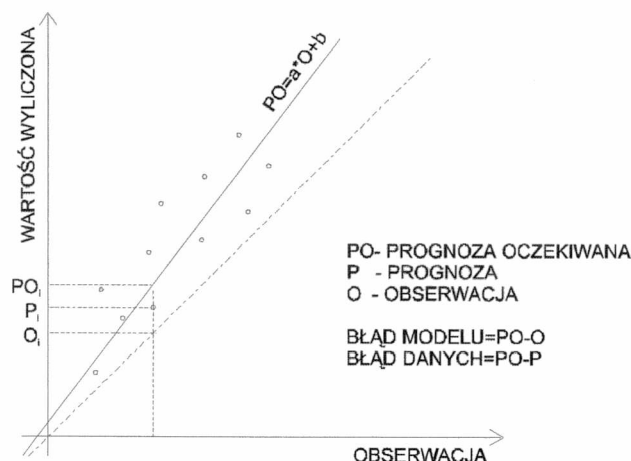
$$R^2 = \frac{(\sum xy - n\bar{x}\bar{y})^2}{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)} \quad (3.5)$$

Jeżeli zmienna wyjaśniana nie była, na etapie poprzedzającym modelowanie, przekształcana funkcyjnie, współczynnik determinacji jest równy podniesionemu do kwadratu współczynnikowi korelacji (R) między rzeczywistymi wartościami zmiennej zmiennej i wynikami modelu. W przeciwnym razie jest to kwadrat współczynnika korelacji między zmienną przekształconą funkcyjnie a otrzymaną z modelu.

Im większa jest wartość współczynnika determinacji, tym lepiej model oddaje wartości zmiennej zależnej. W niektórych zastosowaniach wymagane są wartości tego współczynnika na poziomie co najmniej 0.7÷0.8. Oznacza to, że model musi wyjaśniać co najmniej 70÷80% wariacji zmiennej modelowanej. Ideałem jest współczynnik determinacji równy jeden, wówczas pełny zakres, tj. 100% zmienności zmiennej zależnej jest wyjaśniany przez model.

4. Błąd modelu

Średni błąd kwadratowy składa się z dwóch części. Za jedną odpowiada model, a za drugą odpowiadają dane na podstawie których zbudowano model. Stosunki tych błędów częściowych do całego średniego błędu kwadratowego stanowią odpowiednio błąd modelu (ME) i błąd danych (DE). Na rys.3.3 przedstawiono poglądowo, jak należy rozumieć elementarny błąd związany z modelem i elementarny błąd związany z danymi.



Rysunek 3.3. Poglądowa ilustracja jednostkowego błędu modelu i jednostkowego błędu danych

5. Różnica średniej z wartości rzeczywistych i średniej z wartości wyliczonych przez model

Zbiór wartości rzeczywistych zmiennej i zbiór jej wartości wyliczonych z modelu mają swoje wartości średnie. Różnica między nimi pokazuje, czy model wykazuje systematyczną tendencję do zawyżania czy zaniżania wyników w stosunku do wartości rzeczywistych.

3.3. Prognoza stężeń imisyjnych

3.3.1. Analiza przygotowawcza

Rozprzestrzenianie zanieczyszczeń w powietrzu zewnętrznym przebiega w przestrzeni i czasie. Stężenie zanieczyszczenia w wybranym punkcie przestrzeni i zadanej chwili czasu jest skutkiem transportu i dyfuzji zanieczyszczeń wyemitowanych ze źródeł, w których zasięgu oddziaływania znalazł się ten punkt. Zagadnieniem prognostycznym jest, jak będzie się kształtowało stężenie w przyszłości. Jego rozwiązanie wymaga opisu procesu rozprzestrzeniania zanieczyszczeń, pewnego modelu tego procesu. Okazuje się jednak, że nie jest konieczny opis ściśle formalny. Nie musi to nawet być opis jawny. Można przyjąć, że model, który poprawnie opisuje skutki procesu rozprzestrzeniania, czyli wartości stężeń zanieczyszczeń stanowi reprezentację tego procesu. Jest to zgodne z regułą poprawności implikacji na gruncie logiki Arystotlesa.

Brak wymogu formalnej reprezentacji procesu rozprzestrzeniania, w celu opisu stężeń zanieczyszczeń w powietrzu, wykorzystuje prognostyka stężeń zanieczyszczeń oparta na analizie danych imisyjnych. Modele prognostyczne budowane na podstawie pomiarów stężeń zanieczyszczeń w powietrzu zewnętrznym nie wychodzą od formalnego opisu procesu rozprzestrzeniania. Prawidła rządzące tym procesem są jednak wbudowywane w ich strukturę, często w formie wykluczającej identyfikację tych zasad. Konstruowanie i posługiwanie się tego typu modelami jest konieczne, gdyż formalny opis procesu rozprzestrzeniania zanieczyszczeń z wielu źródeł zanieczyszczeń równocześnie nie został dotychczas opanowany z zadowalającą dokładnością.

Opracowanie modelu prognostycznego wiąże się zasadniczo z dwoma możliwymi podejściami: deterministycznym lub niedeterministycznym. Modelowanie deterministyczne wykorzystuje zależność zmiennej opisywanej od zmiennych opisujących. Modelowanie niedeterministyczne natomiast posługuje się prawidłowościami występującymi w szeregu czasowym zmiennej modelowanej. Przejawia się to opisem przyszłych wartości zmiennej wyłącznie przy pomocy wartości, które przyjmowała wcześniej. Możliwe jest również rozwiązanie pośrednie, gdzie do zbioru zmiennych opisujących należą zarówno zmienne zewnętrzne, jak i przesunięcie czasowe zmiennej opisywanej.

Proponowane w pracy neuronowe modele prognostyczne realizujące koncepcję deterministyczną i niedeterministyczną. Pominięte zostało rozwiązanie mieszane. Neuronowy model deterministyczny umożliwia oszacowanie faktycznego poziomu zależności stężeń

30-minutowych SO_2 , NO_2 i pyłu od parametrów meteorologicznych rejestrowanych standardowo na stacjach monitoringowych. Neuronowy model niedeterministyczny został opracowany z myślą o odpowiedzi na pytanie jak dokładną prognozę stężeń 30-minutowych można sformułować wyłącznie w oparciu o znajomość stężeń występujących wcześniej oraz jaki jest horyzont takiej prognozy. Z podejścia mieszanego zrezygnowano, gdyż w toku analiz stwierdzono, że nie wnosi ono nic nowego w sensie poznawczym. Jego efektywność zaś mieści się między modelem deterministycznym a niedeterministycznym.

Nie podlega dyskusji, że wymienione koncepcje modelowania są możliwe do zrealizowania za pomocą tradycyjnych rozwiązań statystycznych. Przykładem podejścia deterministycznego są modele regresji, niedeterministycznego - modele ARIMA, a pośredniego - modele z funkcją transferu lub modele przestrzeni stanów. Z doniesień literaturowych wynika jednak, że opracowanie statystycznych modeli prognostycznych stężeń zanieczyszczeń zgodnie z zasadami poprawności metodologicznej jest trudnym zadaniem [Milionis, 1994]. Powszechnemu korzystaniu z modeli statystycznych, zwłaszcza regresyjnych, w tej dziedzinie towarzyszy pomijanie rzetelnej informacji o jakości modeli, a charakter danych sugeruje, że prezentowane modele mogą być niepoprawne metodologicznie. Opis za pomocą równania regresji umożliwia modelowanie zależności liniowych lub zależności nieliniowych, jednak pod warunkiem znajomości ich postaci. Bardzo często wyklucza to stosowanie regresji nieliniowej ze względu na brak wiedzy o postaci zależności modelowanej. Modele serii czasowych Boxa-Jenkinsa są stosowane znacznie rzadziej ze względu na gruntowną wiedzę statystyczną, której wymaga ich opracowywanie. Najmniej wymagające pod tym względem wydaje się modelowanie w przestrzeni stanów z wykorzystaniem filtru Kalmana.

Przedstawiono rozwiązanie komplementarne w stosunku do metod statystycznych. Są to neuronowe modele prognostyczne. Decyzja o wyborze właśnie tych modeli była uwarunkowana ich uniwersalnością i prostotą. Neuronowy model prognostyczny nadaje się do rozwiązywania bardzo wielu problemów z zakresu prognostyki. W tej pracy zostanie on zastosowany do krótkoterminowych prognoz stężeń zanieczyszczeń.

Z formalnego punktu widzenia, neuronowy model prognostyczny jest siecią neuronową. Wybór do tego celu konkretnego typu sieci jest uwarunkowany jej własnościami jak np. odwzorowywanie zależności liniowych lub nieliniowych, zdolności klasyfikacyjne, reprezentacja związków o lokalnym lub ogólnym charakterze. Neuronowe modele prognostyczne przejmują cechy sieci neuronowych. Najistotniejsze z nich to:

- niezwykła prostota w porównaniu z modelami statystycznymi, zwłaszcza serii czasowych,

- eliminacja problemu formalnej poprawności modelu prognostycznego,
- uniwersalność, gdyż sieci tego samego typu, różniące się tylko liczbą neuronów mogą realizować bardzo różne rozwiązania prognostyczne,
- zdolność do realizowania odwzorowań nieliniowych o nieznanej postaci, co prowadzi do poprawy rezultatów w stosunku do przybliżeń liniowych,
- umożliwienie oceny znaczenia nieliniowości w modelowanych zależnościach, wykorzystywanych do prognozowania.

Stanowiąc rozwiązanie różne od tradycyjnych, prognostyczny model neuronowy wpisuje się jednocześnie w ogólnie przyjęte koncepcje modelowania prognostycznego. Wspólna dla wszystkich sieci neuronowych struktura umożliwia budowę zarówno deterministycznych, niedeterministycznych jak i mieszanych modeli prognostycznych. Jedna metodologia neuronowa służy realizacji zupełnie różnych rozwiązań prognostycznych. Jest to jej niewątpliwa zaleta.

Jak zaznaczono w części wstępnej pracy, literatura przedmiotu nie opisuje przypadków analogicznych do rozważanego w pracy. Wyjątek stanowią badania Boznar (1993, 1995, 1996) Maklar (1994, 1996) reprezentujące właśnie podejście mieszane i w odniesieniu wyłącznie do stężeń dwutlenku siarki. W związku z tym trudno o materiał porównawczy dla przedstawionego. Z tego powodu, interpretację i ocenę wyników neuronowych modeli prognostycznych poprowadzono w kierunku wniosków, które wypływają z analizy wyników modelowania, raczej niż w kierunku porównań z innymi modelami neuronowymi czy statystycznymi.

Problemem prognostycznym, rozważanym przez autora jest prognoza czasowa 30-minutowych stężeń zanieczyszczeń jak SO_2 , NO_2 i pył zawieszony w powietrzu atmosferycznym. Modele prognostyczne opracowano dla danych pomiarowych pochodzących ze stacji monitoringu powietrza atmosferycznego w Bogatyni za lata 1997/98. Wybór tego punktu pomiarowego uzasadnia jego położenie względem źródeł emisji. Stacja znajduje się na wzniesieniu i jest eksponowana w kierunku napływu zanieczyszczeń emitowanych przez Elektrownię Turów. Jest położona centralnie, między wyrobiskiem Kopalni a zwałowiskiem popiołu, czyli dwoma powierzchniowymi źródłami pyłu (rozdz.3.1.1, rys.3.2). Już na etapie wstępnego opisu stacji ujawnia się duże prawdopodobieństwo rejestracji przez nią wpływu czynników zaburzających oddziaływanie Elektrowni. Usytuowanie w granicach miasta wymaga uwzględnienia w interpretacji wyników modelowania wpływu lokalnej emisji na obraz stężeń rejestrowany w Worku Żytawskim.

W związku z możliwością wykorzystania w prognostyce podejścia deterministycznego interesujące jest stwierdzenie statystycznie istotnej zależności zmiennych opisywanych od zmiennych opisujących. W naszym przypadku dotyczy to stężeń zanieczyszczeń i parametrów

meteorologicznych. Wyniki testu jednorodności wariancji Levene'a nie wykazały niezależności 30-minutowych stężeń zanieczyszczeń i parametrów meteorologicznych rejestrowanych równolegle na stacji pomiarowej.

Wobec powyższego faktu powstaje pytanie, jak wyraźna jest ta zależność i czy możliwa jest identyfikacja jej postaci. Statystyka proponuje miarę oceny współzależności zmiennych, w postaci współczynnika korelacji, jedynie w odniesieniu do związków liniowych. Wartości współczynników korelacji między stężeniami SO_2 (tab.3.1), NO_2 (tab.3.2) i pyłu (tab.3.3) a parametrami meteorologicznymi, rejestrowanymi równolegle, pozwalają postawić tezę o łączącej je bardzo słabej współzależności liniowej. Najwyższe wartości przyjmują współczynniki korelacji stężeń zanieczyszczeń gazowych z temperaturą. Stężenia pyłu wykazują ogólnie najslabszy związek z warunkami meteorologicznymi. Jednak w ich przypadku, znaczenie prędkości wiatru jest większe niż jej rola w kształtowaniu stężeń SO_2 czy NO_2 . Z analizy tabel 3.1, 3.2 i 3.3 wynika ponadto, że mimo ogólnie małych wartości współczynników korelacji, stężenia chwilowe dowolnego zanieczyszczenia są najsilniej związane z warunkami meteorologicznymi panującymi w punkcie rejestracji stężeń.

Bogatynia, którą wybrano jako obiekt modelowania, nie zajmuje pozycji skrajnej pod względem wartości współczynników korelacji wyznaczonych dla wszystkich stacji pomiarowych. W sensie stopnia uwarunkowania zmienności stężeń przez parametry meteorologiczne jest więc reprezentatywna dla całej sieci pomiarowej.

Zważywszy dynamiczny charakter ośrodka, jakim jest powietrze atmosferyczne, należy zastanowić się, czy określona wartość stężenia zanieczyszczenia w wybranym punkcie nie jest w większym stopniu rezultatem warunków meteorologicznych występujących wcześniej, niż równocześnie z wartością stężenia. Miarą liniowej zależności stężeń zanieczyszczeń od wartości parametrów przesuniętych w czasie jest współczynnik korelacji wzajemnej.

Tabela 3.1. Współczynniki korelacji pomiędzy stężeniem SO₂ a parametrami meteorologicznymi rejestrowanymi na poszczególnych stacjach monitoringowych

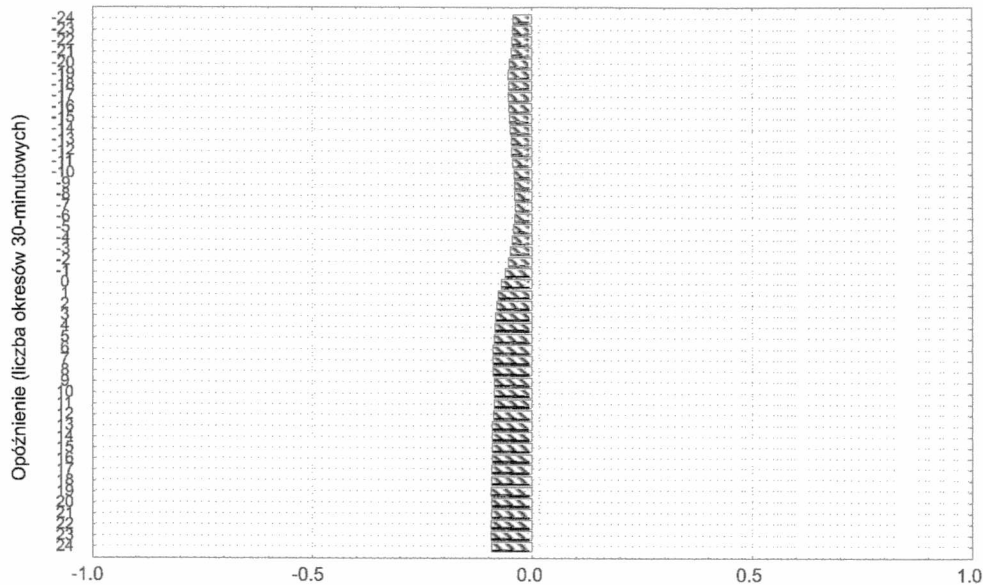
		Stężenie SO ₂							
		Bogatynia	Działoszyn	Jasna Góra	Radomierzyce	Sieniawka	Witka	Wyszków	Zgorzelec
Bogatynia	kierunek wiatru	0.02	0.00	-0.06	0.03	-0.09	0.09	-0.01	-0.06
	prędkość wiatru	-0.07	-0.03	-0.06	0.17	-0.05	0.11	-0.02	0.11
	wilgotność	0.06	0.04	0.10	0.00	0.10	-0.01	0.03	0.02
	temperatura	-0.39	-0.26	-0.37	-0.19	-0.39	-0.18	-0.36	-0.31
	ciśnienie	0.25	0.17	0.25	0.12	0.24	0.14	0.24	0.10
Działoszyn	kierunek wiatru	-0.08	-0.11	-0.13	-0.04	-0.19	-0.02	-0.11	-0.16
	prędkość wiatru	0.07	0.12	0.11	0.36	0.12	0.29	0.14	0.21
	wilgotność	0.05	0.03	0.09	0.04	0.09	0.03	0.03	0.05
	temperatura	-0.39	-0.25	-0.36	-0.20	-0.39	-0.19	-0.36	-0.31
	ciśnienie	0.23	0.15	0.22	0.10	0.21	0.12	0.22	0.08
Jasna Góra	kierunek wiatru	-0.11	-0.15	-0.21	-0.10	-0.27	-0.09	-0.15	-0.20
	prędkość wiatru	0.05	0.09	0.10	0.39	0.09	0.28	0.10	0.28
	wilgotność	0.05	0.02	0.08	0.03	0.08	-0.01	0.01	0.05
	temperatura	-0.38	-0.24	-0.35	-0.20	-0.38	-0.18	-0.35	-0.31
	ciśnienie	0.17	0.08	0.15	0.09	0.16	0.12	0.15	0.01
Radomierzyce	kierunek wiatru	0.02	0.07	0.04	0.11	0.09	0.12	0.04	0.09
	prędkość wiatru	0.05	0.06	0.07	0.28	0.08	0.21	0.09	0.24
	wilgotność	0.02	0.00	0.06	-0.04	0.06	-0.04	-0.02	-0.02
	temperatura	-0.38	-0.24	-0.35	-0.17	-0.38	-0.17	-0.34	-0.29
	ciśnienie	0.24	0.15	0.24	0.10	0.23	0.11	0.23	0.09
Sieniawka	kierunek wiatru	-0.10	-0.09	-0.10	-0.07	-0.15	-0.05	-0.10	-0.14
	prędkość wiatru	-0.12	-0.07	-0.10	-0.01	-0.09	-0.04	-0.07	-0.04
	wilgotność	0.07	0.05	0.12	0.04	0.11	0.04	0.05	0.05
	temperatura	-0.39	-0.26	-0.37	-0.20	-0.39	-0.20	-0.36	-0.31
	ciśnienie	0.15	0.12	0.17	0.08	0.17	0.09	0.18	0.07
Witka	kierunek wiatru	0.12	0.14	0.20	0.13	0.20	0.10	0.12	0.18
	prędkość wiatru	0.03	0.02	0.01	0.30	0.04	0.22	0.03	0.20
	wilgotność	0.03	0.01	0.07	-0.04	0.06	-0.02	-0.01	-0.02
	temperatura	-0.37	-0.24	-0.35	-0.18	-0.38	-0.17	-0.34	-0.29
	ciśnienie	0.25	0.16	0.24	0.11	0.23	0.13	0.23	0.09
Wyszków	kierunek wiatru	-0.03	-0.08	-0.08	-0.05	-0.17	-0.03	-0.06	-0.14
	prędkość wiatru	-0.03	0.02	-0.01	0.28	0.00	0.21	0.04	0.17
	wilgotność	0.07	0.04	0.10	0.02	0.09	0.01	0.04	0.03
	temperatura	-0.38	-0.24	-0.35	-0.20	-0.38	-0.18	-0.35	-0.30
	ciśnienie	0.22	0.14	0.22	0.11	0.22	0.13	0.21	0.08
Zgorzelec	kierunek wiatru	-0.05	-0.06	-0.10	0.01	-0.15	0.03	-0.06	-0.13
	prędkość wiatru	-0.02	-0.01	-0.04	0.16	0.00	0.10	0.05	0.10
	wilgotność	0.04	0.02	0.09	-0.02	0.08	-0.01	0.01	-0.01
	temperatura	-0.38	-0.24	-0.36	-0.18	-0.38	-0.17	-0.34	-0.29
	ciśnienie	0.25	0.16	0.25	0.11	0.24	0.12	0.24	0.10

Tabela 3.2. Współczynniki korelacji pomiędzy stężeniem NO₂ a parametrami meteorologicznymi rejestrowanymi na poszczególnych stacjach monitoringowych

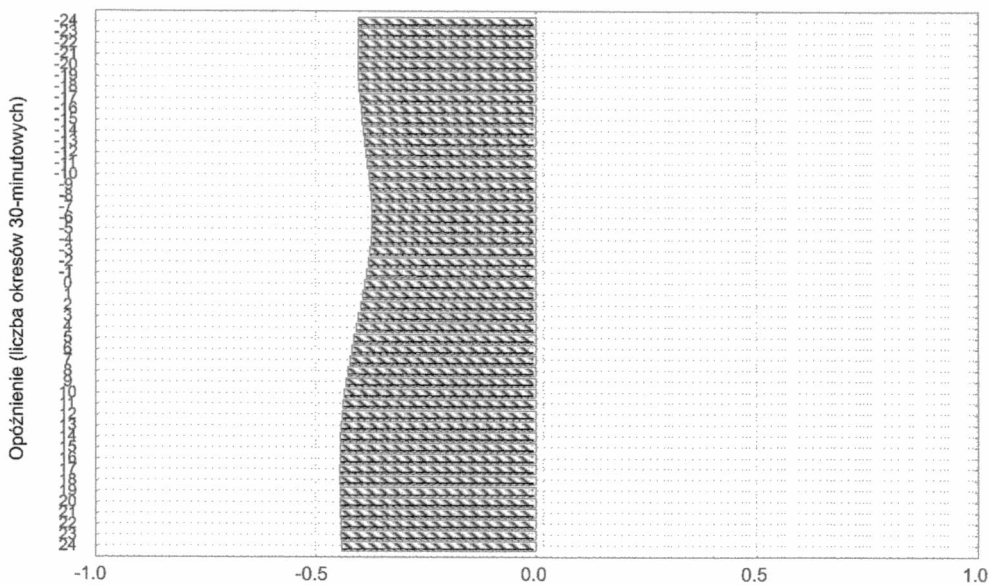
		Stężenie NO ₂							
		Bogatynia	Działoszyn	Jasna Góra	Radomierzyce	Sieniawka	Witka	Wyszków	Zgorzelec
Bogatynia	kierunek wiatru	0.09	0.04	-0.03	-0.09	-0.04	0.00	0.04	-0.07
	prędkość wiatru	-0.22	-0.12	-0.11	-0.12	-0.17	-0.05	-0.08	-0.10
	wilgotność	0.20	0.10	0.22	0.15	0.22	0.16	0.11	0.21
	temperatura	-0.31	-0.19	-0.39	-0.31	-0.42	-0.32	-0.24	-0.42
	ciśnienie	0.24	0.20	0.17	0.19	0.30	0.16	0.17	0.19
Działoszyn	kierunek wiatru	-0.03	0.07	-0.09	-0.12	-0.08	-0.09	0.02	-0.12
	prędkość wiatru	-0.02	-0.05	0.02	0.01	-0.03	0.10	0.00	0.01
	wilgotność	0.14	0.11	0.25	0.14	0.17	0.21	0.14	0.20
	temperatura	-0.30	-0.19	-0.40	-0.30	-0.40	-0.33	-0.24	-0.41
	ciśnienie	0.24	0.19	0.16	0.18	0.28	0.15	0.17	0.17
Jasna Góra	kierunek wiatru	-0.06	0.01	-0.14	-0.18	-0.13	-0.15	0.00	-0.17
	prędkość wiatru	-0.04	-0.02	0.04	0.10	-0.05	0.14	0.01	0.06
	wilgotność	0.13	0.06	0.25	0.12	0.16	0.16	0.12	0.18
	temperatura	-0.28	-0.18	-0.39	-0.28	-0.39	-0.31	-0.23	-0.40
	ciśnienie	0.07	0.00	-0.11	-0.03	0.11	-0.01	-0.09	0.03
Radomierzyce	kierunek wiatru	0.06	0.08	0.06	0.09	0.08	0.11	0.10	0.04
	prędkość wiatru	-0.06	-0.10	-0.01	-0.05	-0.05	0.04	-0.02	-0.01
	wilgotność	0.15	0.13	0.20	0.25	0.20	0.18	0.11	0.23
	temperatura	-0.29	-0.20	-0.39	-0.34	-0.42	-0.33	-0.23	-0.42
	ciśnienie	0.24	0.19	0.17	0.18	0.30	0.15	0.17	0.19
Sieniawka	kierunek wiatru	-0.07	0.02	-0.09	-0.10	-0.05	-0.10	-0.02	-0.10
	prędkość wiatru	-0.21	-0.13	-0.06	-0.13	-0.15	-0.13	-0.07	-0.14
	wilgotność	0.19	0.14	0.27	0.21	0.25	0.23	0.16	0.25
	temperatura	-0.31	-0.20	-0.40	-0.32	-0.42	-0.34	-0.24	-0.42
	ciśnienie	0.20	0.20	0.23	0.20	0.29	0.17	0.24	0.16
Witka	kierunek wiatru	0.08	0.04	0.13	0.16	0.16	0.17	0.04	0.18
	prędkość wiatru	-0.07	-0.04	-0.03	-0.01	-0.05	0.05	0.02	0.01
	wilgotność	0.16	0.12	0.20	0.22	0.20	0.18	0.10	0.23
	temperatura	-0.30	-0.19	-0.39	-0.33	-0.42	-0.33	-0.23	-0.42
	ciśnienie	0.24	0.20	0.17	0.18	0.30	0.16	0.17	0.19
Wyszków	kierunek wiatru	0.00	0.06	-0.06	-0.08	-0.04	-0.06	0.03	-0.07
	prędkość wiatru	-0.08	-0.04	-0.02	-0.02	-0.12	0.05	0.00	-0.03
	wilgotność	0.16	0.10	0.25	0.13	0.20	0.19	0.15	0.20
	temperatura	-0.28	-0.18	-0.39	-0.28	-0.40	-0.31	-0.24	-0.40
	ciśnienie	0.22	0.18	0.14	0.16	0.27	0.14	0.15	0.16
Zgorzelec	kierunek wiatru	0.01	0.08	-0.07	-0.08	-0.06	-0.04	0.05	-0.13
	prędkość wiatru	-0.14	-0.11	-0.08	-0.17	-0.11	-0.08	-0.04	-0.12
	wilgotność	0.17	0.14	0.22	0.22	0.22	0.20	0.13	0.25
	temperatura	-0.30	-0.20	-0.39	-0.33	-0.43	-0.33	-0.24	-0.43
	ciśnienie	0.24	0.19	0.18	0.19	0.31	0.16	0.17	0.20

Tabela 3.3. Współczynniki korelacji pomiędzy stężeniem pyłu a parametrami meteorologicznymi na poszczególnych stacjach monitoringowych

		Stężenie pyłu							
		Bogatynia	Działoszyn	Jasna Góra	Radomierzyce	Sieniawka	Witka	Wyszków	Zgorzelec
Bogatynia	kierunek wiatru	0.00	-0.12	-0.16	-0.17	-0.17	-0.16	-0.09	-0.20
	prędkość wiatru	-0.29	-0.25	-0.24	-0.22	-0.23	-0.19	-0.20	-0.19
	wilgotność	0.08	0.03	0.05	0.02	0.02	-0.02	0.06	0.01
	temperatura	-0.11	-0.05	-0.12	-0.16	-0.10	-0.16	-0.20	-0.16
	ciśnienie	0.26	0.21	0.19	0.22	0.22	0.23	0.20	0.19
Działoszyn	kierunek wiatru	-0.09	-0.20	-0.22	-0.21	-0.24	-0.26	-0.18	-0.28
	prędkość wiatru	-0.22	-0.15	-0.20	-0.18	-0.21	-0.12	-0.10	-0.16
	wilgotność	-0.06	-0.04	-0.01	-0.04	-0.07	-0.06	0.00	-0.05
	temperatura	-0.06	-0.02	-0.11	-0.16	-0.07	-0.15	-0.19	-0.15
	ciśnienie	0.25	0.20	0.19	0.22	0.21	0.23	0.19	0.19
Jasna Góra	kierunek wiatru	-0.10	-0.18	-0.20	-0.20	-0.21	-0.25	-0.16	-0.24
	prędkość wiatru	-0.21	-0.11	-0.21	-0.09	-0.17	-0.04	-0.08	-0.06
	wilgotność	-0.04	-0.05	0.03	-0.04	-0.04	-0.06	0.00	-0.04
	temperatura	-0.05	-0.01	-0.11	-0.14	-0.07	-0.13	-0.18	-0.13
	ciśnienie	0.13	0.08	0.03	0.08	0.06	0.10	0.07	0.06
Radomierzyce	kierunek wiatru	-0.01	0.05	0.00	0.01	0.00	0.06	0.03	0.02
	prędkość wiatru	-0.21	-0.17	-0.17	-0.17	-0.16	-0.09	-0.12	-0.12
	wilgotność	0.04	0.05	0.07	0.08	0.00	0.02	0.08	0.07
	temperatura	-0.10	-0.06	-0.14	-0.19	-0.09	-0.18	-0.21	-0.19
	ciśnienie	0.26	0.20	0.19	0.22	0.22	0.23	0.20	0.19
Sieniawka	kierunek wiatru	-0.10	-0.15	-0.14	-0.14	-0.23	-0.19	-0.13	-0.19
	prędkość wiatru	-0.20	-0.21	-0.18	-0.19	-0.18	-0.20	-0.20	-0.18
	wilgotność	0.03	0.03	0.06	0.03	0.02	0.00	0.07	0.02
	temperatura	-0.08	-0.04	-0.11	-0.16	-0.09	-0.16	-0.20	-0.16
	ciśnienie	0.19	0.14	0.14	0.16	0.17	0.16	0.14	0.12
Witka	kierunek wiatru	0.13	0.21	0.20	0.21	0.21	0.23	0.15	0.23
	prędkość wiatru	-0.21	-0.16	-0.19	-0.15	-0.18	-0.10	-0.14	-0.11
	wilgotność	0.04	0.04	0.06	0.06	0.00	0.00	0.07	0.05
	temperatura	-0.09	-0.05	-0.13	-0.18	-0.09	-0.17	-0.21	-0.17
	ciśnienie	0.26	0.21	0.20	0.23	0.22	0.24	0.20	0.20
Wyszków	kierunek wiatru	-0.05	-0.17	-0.14	-0.14	-0.18	-0.20	-0.13	-0.22
	prędkość wiatru	-0.25	-0.21	-0.27	-0.20	-0.26	-0.16	-0.17	-0.18
	wilgotność	-0.01	-0.02	0.03	-0.02	-0.01	-0.04	0.03	-0.03
	temperatura	-0.06	-0.02	-0.10	-0.14	-0.08	-0.14	-0.18	-0.14
	ciśnienie	0.23	0.17	0.16	0.18	0.18	0.19	0.17	0.15
Zgorzelec	kierunek wiatru	-0.13	-0.19	-0.25	-0.24	-0.26	-0.25	-0.17	-0.30
	prędkość wiatru	-0.24	-0.24	-0.21	-0.22	-0.20	-0.20	-0.15	-0.21
	wilgotność	0.05	0.06	0.05	0.05	0.00	0.01	0.08	0.06
	temperatura	-0.10	-0.06	-0.13	-0.18	-0.10	-0.17	-0.21	-0.18
	ciśnienie	0.27	0.21	0.20	0.24	0.23	0.24	0.21	0.20



Rysunek 3.4. Współczynniki korelacji wzajemnej szeregów czasowych 30-minutowych stężeń SO_2 i prędkości wiatru w funkcji opóźnienia prędkości wiatru, dla pomiarów wykonanych w Bogatyni



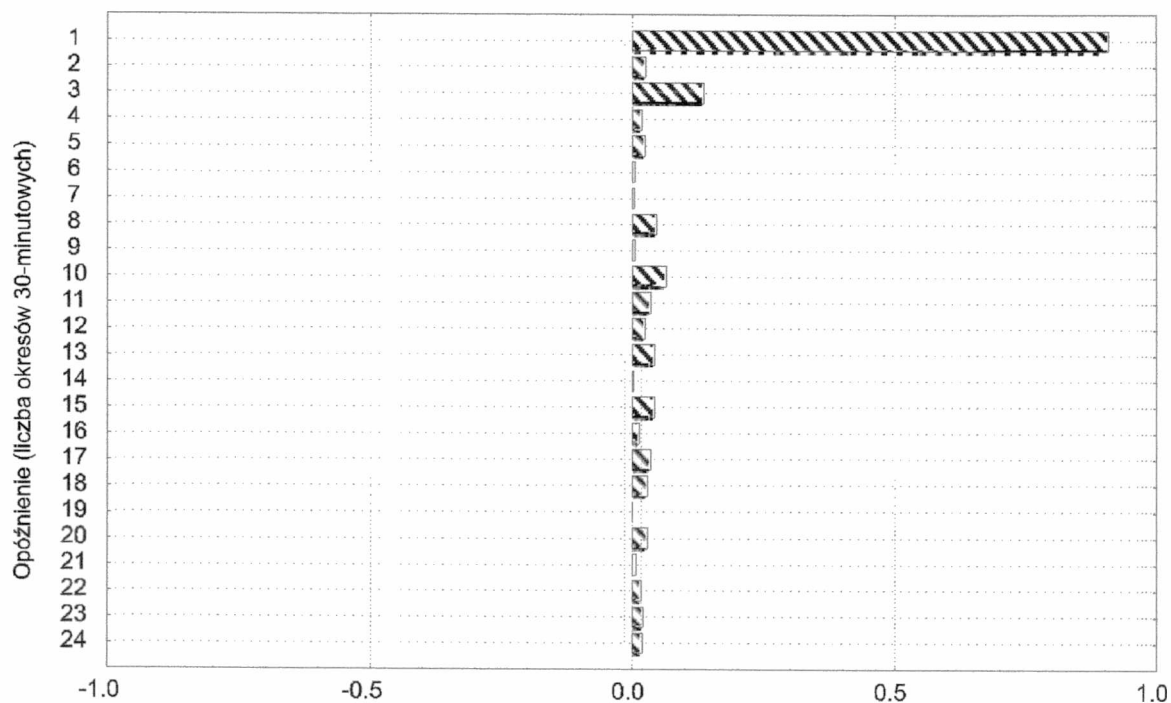
Rysunek 3.5. Współczynniki korelacji wzajemnej szeregów czasowych 30-minutowych stężeń SO_2 i temperatury w funkcji opóźnienia temperatury, dla pomiarów wykonanych w Bogatyni

Z wartości współczynników korelacji wzajemnej (rys.3.5) wynika np., że stężenie 30-minutowe SO_2 w Bogatyni silniej zależy od temperatury powietrza przed 12 godzinami, niż temperatury bieżącej. Ponadto bardziej wpływa na nie prędkość wiatru, która wystąpiła 2 godziny wcześniej (rys.3.4). Jest to przykład sytuacji, gdzie szeregi czasowe przesunięte względem siebie są silniej skorelowane, niż szeregi nie przesunięte. Niestety spostrzeżeń poczynionych dla SO_2

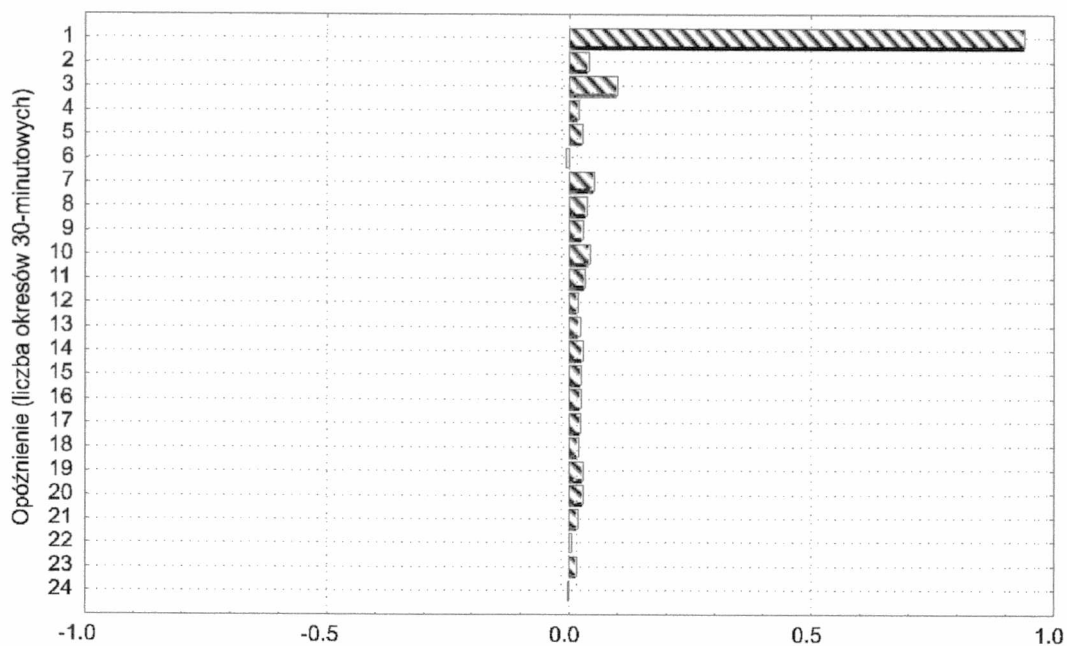
w Bogatyni nie można uogólnić na zestawy serii czasowych stężeń zanieczyszczeń i parametrów meteorologicznych pochodzące z pozostałych stacji monitoringowych. W każdym przypadku istnieją opóźnienia serii czasowych parametrów meteorologicznych względem serii czasowych stężeń, dla których współczynniki korelacji są wyraźnie wyższe od wartości podanych w tabelach (tab.3.1, tab.3.2, tab.3.3). Wprowadzenie do zbioru zmiennych objaśniających parametrów meteorologicznych przesuniętych w czasie w stosunku do zmiennej objaśnianej stwarza możliwość poprawy wyników modelowania deterministycznego. Wiąże się z tym również inna zaleta. Podczas wykonywania prognozy stężeń wykorzystywane są faktyczne wyniki pomiarów parametrów meteorologicznych nie zaś ich prognozy, co niewątpliwie wpływa na zmniejszenie błędu prognozy.

Jak wynika z analizy statystycznej, nie należy się spodziewać bardzo dobrych wyników modelowania stężeń zanieczyszczeń w układzie deterministycznym za pomocą modelu liniowego. Wartości maksymalne współczynników korelacji wzajemnej stężeń zanieczyszczeń i parametrów meteorologicznych nie przekraczają 0.5. Potencjał wyjaśniający zmiennych jak: prędkość i kierunek wiatru, wilgotność, temperatura oraz ciśnienie w odniesieniu do 30-minutowych stężeń zanieczyszczeń rejestrowanych wokół punkowego źródła zanieczyszczeń jest mały.

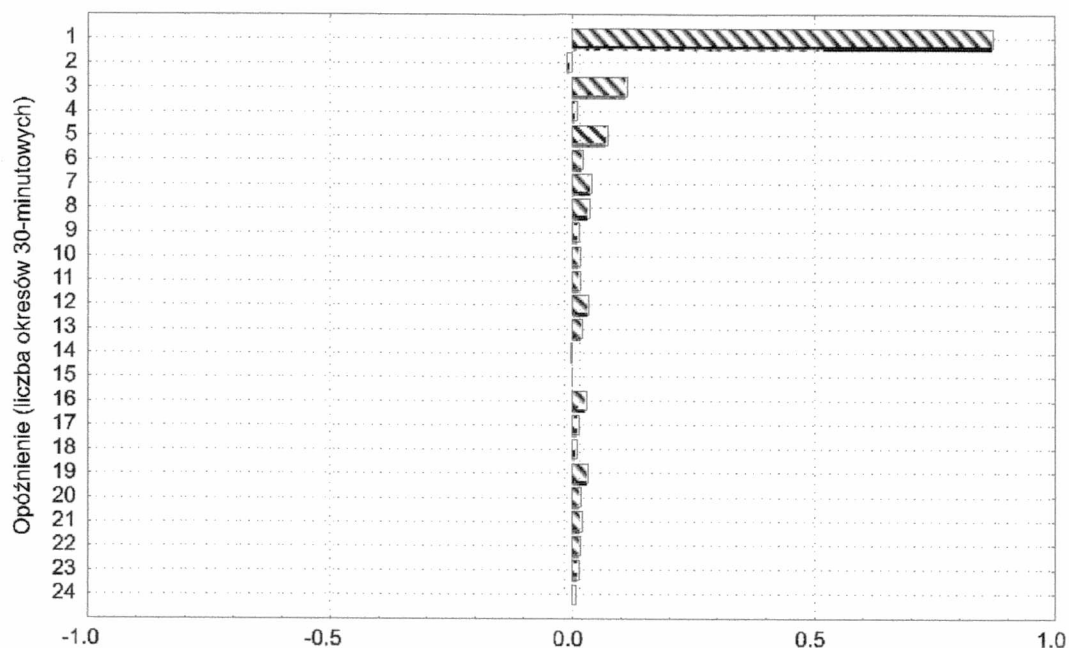
Choć bardzo naturalnym wydaje się wykorzystanie w prognostyce stężeń zanieczyszczeń ich zależności od zmiennych zewnętrznych w postaci np. parametrów meteorologicznych możliwe jest również inne rozwiązanie. Stanowi je model prognostyczny oparty wyłącznie na informacji zawartej w serii czasowej stężeń i nie korzystający z żadnych innych zmiennych. Model taki stosuje się z konieczności wówczas, gdy zmienne opisujące nie są znane np. nie podlegają pomiarom. W warunkach dostępności informacji o zmiennych zewnętrznych uzasadnieniem rezygnacji z wykorzystania tych zmiennych może być ich mały wpływ na zmienną objaśnianą lub to, że wartości zmiennych zewnętrznych nie odbiegają wyraźnie od pewnego zasadniczego kursu. Ich oddziaływanie na zmienną objaśnianą ma więc charakter ustalony.



Rysunek 3.6. Współczynniki autokorelacji cząstkowej dla serii czasowej stężeń SO_2 zarejestrowanych w Bogatyni w funkcji opóźnienia



Rysunek 3.7. Współczynniki autokorelacji cząstkowej dla serii czasowej stężeń NO_2 zarejestrowanych w Bogatyni w funkcji opóźnienia



Rysunek 3.8. Współczynniki autokorelacji cząstkowej dla serii czasowej stężeń pyłu zarejestrowanych w Bogatyni w funkcji opóźnienia

Powszechnie wiadomo, że serie czasowe stężeń zanieczyszczeń są w wysokim stopniu skorelowane wewnętrznie. Faktyczna zależność łączy jednak wyłącznie stężenia rejestrowane w kolejnych okresach pógodzinnych, a jest zaniedbywana dla większych odległości czasowych. Informują o tym wysokie wartości współczynnika autokorelacji cząstkowej, które wynoszą 0.95 dla serii czasowej stężeń NO_2 , 0.9 dla SO_2 i 0.87 dla pyłu, występujące przy opóźnieniu równym 30 minut. Dla pozostałych opóźnień występują małe wartości współczynnika autokorelacji cząstkowej równe, co najwyżej, 0.1 (rys. 3.6, rys.3.7, rys.3.8). Związek sąsiednich stężeń 30-minutowych jest zbliżony w przypadku serii czasowych wszystkich zanieczyszczeń, lecz najsilniej wypada dla NO_2 , a najslabiej dla pyłu.

3.3.2. Prognoza z wykorzystaniem neuronowego modelu deterministycznego

Modelowanie deterministyczne polega na analizowaniu zachowania zmiennej opisywanej w odpowiedzi na zmiany zmiennych opisujących. Warunkiem modelowania deterministycznego jest występowanie co najmniej współzależności, jeżeli nie zależności o charakterze przyczynowo skutkowym między zmiennymi opisującymi a zmienną opisywaną.

Przedmiotem rozważań w tym rozdziale są prognozy stężeń dwutlenku siarki, dwutlenku azotu i pyłu zawieszonego wykonywane za pomocą modelu deterministycznego w oparciu o wyniki pomiarów prędkości i kierunku wiatru, wilgotności, temperatury i ciśnienia atmosferycznego. Jako metodę rozwiązania problemu zaproponowano sieć neuronową z radialną funkcją transferu. Uzasadnieniem wyboru sieci radialnej jest proponowana interpretacja neuronów sieci jako reprezentantów zbliżonych zestawów parametrów meteorologicznych. Neuron radialny odpowiada pozytywnie na wektory wejściowe zbliżone do wektora wag neuronu, a jego odpowiedź słabnie w miarę wzrostu różnicy między wektorem wejściowym i wektorem wag neuronu. Każdemu wektorowi wejściowemu parametrów meteorologicznych towarzyszy wartość stężenia zanieczyszczenia podawana na wyjście sieci. Poszczególne neurony sieci uczą się więc zależności między warunkami meteorologicznymi a stężeniami zanieczyszczeń w obrębie zakresów warunków meteorologicznych, a cała sieć powinna opanować pełne spektrum zmienności warunków meteorologicznych i pełen zakres zmienności stężeń. Rozsądnym wydaje się założenie, że przy stabilnym polu emisji skutkiem zbliżonych warunków meteorologicznych są zbliżone wartości stężeń zanieczyszczeń. Wobec tego, na etapie wykonywania prognozy, jeżeli wektor wejściowy opisuje sytuację meteorologiczną bliską sytuacji reprezentowanej przez dowolny neuron sieci, odpowiedź w postaci wartości stężenia również powinna być bliska rzeczywistej.

3.3.2.1. Propozycja rozwiązania problemu

W literaturze najczęściej spotykane jest modelowanie stężeń zanieczyszczeń w oparciu o bieżące warunki meteorologiczne [np. Gardner, 1999; Shi, 1997; Comrie 1997]. Warto uświadomić sobie, że w warunkach rzeczywistych takie podejście wymaga dodatkowo wykonania prognozy warunków meteorologicznych, co obarcza prognozę stężeń dodatkowym błędem. Z tego powodu korzystniejsze jest posługiwanie się informacją o wcześniejszych warunkach meteorologicznych, bo ta informacja może pochodzić z pomiarów.

Analiza współczynników korelacji wzajemnej stężeń zanieczyszczeń i parametrów meteorologicznych wskazuje na istnienie statystycznie istotnej współzależności między nimi.

Jest to związek bardzo słaby, a jego natężenie zmienia się w funkcji przesunięcia czasowego między seriami czasowymi zmiennych (rozdz. 3.3.1). Maksimum zależności rzadko występuje przy jednym opóźnieniu, wspólnym dla wszystkich parametrów meteorologicznych, czyli dla sytuacji meteorologicznej. Wynikałoby stąd, że na płaszczyźnie analizy statystycznej zależność stężenia zanieczyszczenia od czegoś, co nazywamy sytuacją meteorologiczną jest słabsza niż od zestawu wartości parametrów meteorologicznych rozsuniętych w czasie. Opóźnienie musi być takie, żeby wartości współczynników korelacji wzajemnej między szeregami czasowymi poszczególnych parametrów meteorologicznych a szeregiem stężeń zanieczyszczeń były maksymalne (tab.3.4). Wykorzystanie zestawu wartości parametrów meteorologicznych, które wykazują maksymalny z możliwych związek ze stężeniem zanieczyszczenia stwarza szansę poprawy jakości prognozy w stosunku do rozwiązania z wykorzystaniem sytuacji meteorologicznej. Niestety, położenie optimum korelacji wzajemnej może się zmieniać wraz ze zmianą długości serii czasowych, czyli nie jest stałe w czasie dla danej pary stężenie zanieczyszczenia-parametr meteorologiczny. Wynika stąd, że położenie maksimum współczynników korelacji wzajemnej dla zbioru uczącego i dla zbioru testowego mogą się różnić. Wówczas model, który wydaje się w sposób optymalny wykorzystywać prawidłowości ujawniające się w danych uczących, w warunkach prognozy nie będzie się zachowywał optymalnie.

Tabela 3.4. Opóźnienie [godzina] serii czasowych warunków meteorologicznych w stosunku do stężeń zanieczyszczeń maksymalizujące współczynniki korelacji wzajemnej, dla danych ze stacji monitoringu powietrza atmosferycznego w Bogatyni za lata 1997 i 98

zanieczyszczenie	parametr meteorologiczny				
	kierunek wiatru	prędkość wiatru	wilgotność	temperatura	ciśnienie
SO ₂	2	12	12	12	0
NO ₂	0	0.5	0	3	0
pył	12	0.5	1	12	0

Model zależności stężeń zanieczyszczeń od parametrów meteorologicznych jest przygotowywany w celach prognostycznych, a rzeczywistą sytuację prognostyczną symuluje testowanie modelu na danych spoza zbioru uczącego. Jednakże możliwości prognostyczne modelu są uwarunkowane tym, jak dobra reprezentacja wektorów testowych znajduje się w zbiorze uczącym. Innymi słowy, zależy od występowania w zbiorze uczącym konfiguracji bliskich testowym co przygotowuje się do generalizacji i udzielania dobrych odpowiedzi na wektory testowe.

Jedną z propozycji konstrukcji zbiorów uczącego i testowego polega na arbitralnym podziale dostępnego zestawu danych na dwie części. Według jednych rozwiązań okres czasu, któremu odpowiada zbiór testujący następuje po okresie czasu, któremu odpowiada zbiór uczący np. [Boznar,93]. Według innych sieć nauczona na danych pochodzących np. z jednego roku jest testowana na danych pochodzących z innego roku np. [Gardner, Dorling,2000]. Może to również dotyczyć sezonów np. zima-zima, czy lato-lato. Obie strategie mają zapewnić obecność w zbiorze

uczącym wektorów zbliżonych do testowych. Faktyczna realizacja tego zadania w pierwszym przypadku wymaga w zasadzie ciągłego podprowadzania sieci neuronowej pod przypadek testowy poprzez systematyczne wprowadzanie do zbioru uczącego ostatnio zarejestrowanych zestawów danych: parametry meteorologiczne-stężenie zanieczyszczenia. Taki zabieg miałby zagwarantować, że ze względu na bliskość czasową zbiorów testowego i uczącego warunki zbliżone do testowych są uwzględnione w strukturze wag sieci. Uzasadnienie drugiego rozwiązania polega na założeniu podobieństwa warunków meteorologicznych występujących na analizowanym obszarze w kolejnych latach lub sezonach. Jak wiadomo to założenie nie jest mocne, gdyż kolejne lata lub sezony mogą się istotnie różnić pod względem ich charakterystyki meteorologicznej.

W związku z tym zaproponowano inne rozwiązanie. Polega ono na wydzieleniu z całego zbioru wektorów wejściowych wzorców, czyli grup wektorów podobnych do siebie. Zaletą tego podejścia jest możliwość budowy kilku sieci neuronowych, po jednej osobno dla każdego wzorca warunków meteorologicznych. Ze względu na profilowanie danych uczących wstępna klasyfikacja zbioru uczącego powinna ułatwić i przyspieszyć naukę sieci oraz poprawić jakość modeli. Ponadto, na etapie wykonywania prognozy, sprawdzenie do którego wzorca należy wektor testowy warunków meteorologicznych pozwala z dużym prawdopodobieństwem przyjąć, że wybrany model prognostyczny jest wyuczony zależności łączącej ten lub zbliżony wzorec ze stężeniem zanieczyszczenia. Proponowane rozwiązanie wydaje się mieć bardziej ogólny charakter, niż arbitralny podział zbioru danych na zbiór wektorów uczących i zbiór wektorów testowych.

Klasyfikację zbioru warunków meteorologicznych można przeprowadzić np. z wykorzystaniem samoorganizującej sieci neuronowej. Po przeprowadzeniu klasyfikacji dla każdego wyróżnionego wzorca przygotowywana jest odrębna sieć neuronowa, która oddaje charakter zależności między zestawami parametrów meteorologicznych należących do wzorca a odpowiadającymi im wartościami stężeń zanieczyszczeń. Procedura korzystania z modelu wygląda tak, że każdy testowy zestaw parametrów meteorologicznych jest najpierw sprawdzany ze względu na przynależność do wzorców. W rezultacie wybierana jest ta sieć, która najlepiej odzwierciedla przełożenie danego wektora testowego na stężenie zanieczyszczeń i prognoza stężeń jest wykonywana z wykorzystaniem tej sieci.

Kryterium różnicującym grupy warunków meteorologicznych jest kierunek wiatru. Ten parametr meteorologiczny charakteryzuje się największą zmiennością, czyli najsilniej wpływa na dynamikę stężeń zanieczyszczeń. Ponadto kierunek wiatru decyduje, skąd napływają zanieczyszczenia do punktu pomiarowego. Wyróżniono 14 grup warunków meteorologicznych.

Zważywszy standardowe 8 lub 12 sektorowe róże wiatrów, 14 grup umożliwia wyłonienie układów pośrednich, jeśli takie istnieją. Dla określenia w jaki sposób wydzielone grupy korespondują z tradycyjnie pojmowaną różą wiatrów na rys.3.9 pokazano rozmieszczenie 14 grup warunków meteorologicznych w sektorach ośmiosektorowej róży wiatrów. Więcej szczegółów zawiera tab.3.5. Sektorom 5, 6 i 7 reprezentowanym najczęściej w zbiorze danych pomiarowych przydzielono więcej niż jedną grupę (rys.3.9). Poszczególne grupy liczą od 1000 do 4000 przypadków.

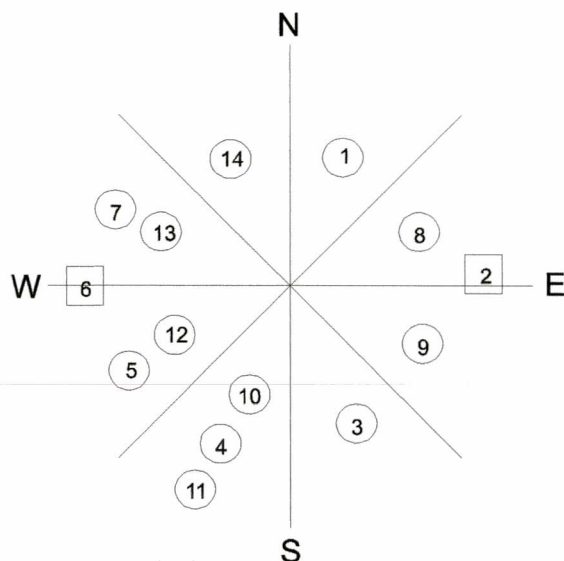
Jest interesujące, na ile wydzielenie grup warunków meteorologicznych według kierunku wiatru wpływa na jednolitość ich profilu pod względem pozostałych parametrów meteorologicznych, czyli prędkości wiatru, temperatury, wilgotności, ciśnienia atmosferycznego, lecz przede wszystkim pod względem stężeń zanieczyszczeń.

Tabela 3.5. Udział [%] kierunków wiatru z ośmiosektorowej róży wiatru w grupach warunków meteorologicznych

grupa	sektor kierunku wiatru							
	0-45°	45-90°	90-135°	135-180°	180-225°	225-270°	270-315°	315-360°
1	91	9						
2		61	39					
3				100				
4					100			
5					21	79		
6						51	49	
7							98	2
8	1	99						
9			80	20				
10			22	78				
11					84	16		
12						98	2	
13							100	
14							4	96

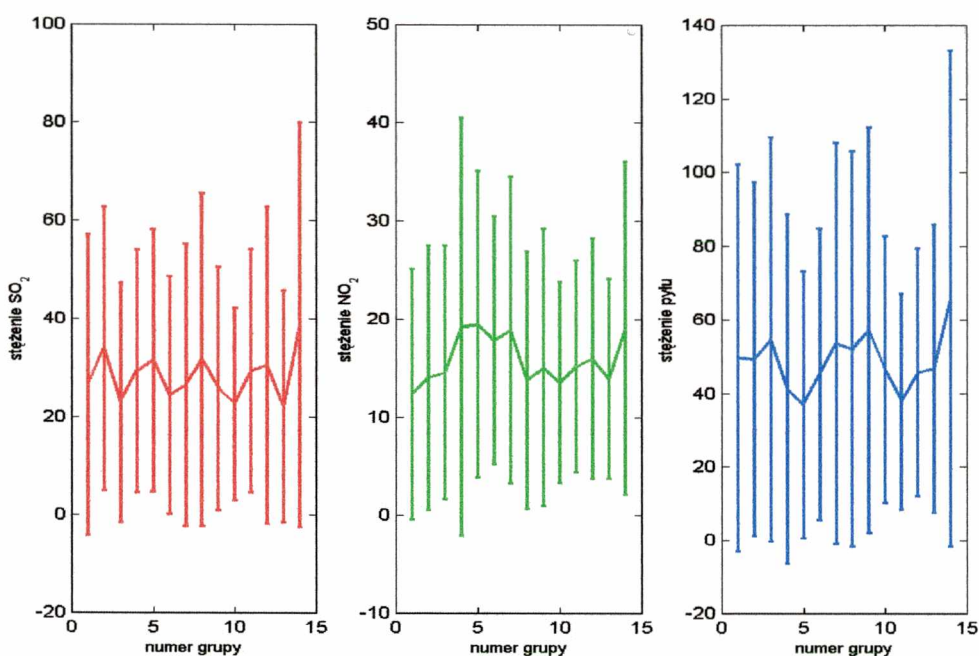
W obrębie wyróżnionych grup mniejsze prędkości wiatru (0-3m/s) dominowały w grupach (2, 3, 8, 9) odpowiadających wschodnim i południowo wschodnim sektorom wiatru i w grupach (5, 6, 7) odpowiadających sektorom zachodnim. W stosunku do nich grupy (4, 10, 11, 12, 13) pokrywające się z sektorami południowo-zachodnimi i zachodnimi cechuje wzrost udziału prędkości wiatru z zakresu 3-6m/s. Grupy 1 i 14 odpowiadające kierunkom północnym charakteryzują się równomiernym rozkładem prędkości wiatru między zakresami. Pod względem temperatury, w grupach cieplejszych (3, 10, 11, 12, 13) nieznacznie przeważają temperatury powyżej 10°C. W pozostałych grupach dominują temperatury poniżej 10°C. Udział temperatur poniżej (-10°C) szczególnie wzrasta w grupach (2, 8 i 9). Przeciętne wilgotności powietrza w zakresie 55-75% dominują w grupach (10, 11, 12 i 13), a w pozostałych wilgotność jest najczęściej większa od 75%. Najbardziej suche powietrze, o wilgotności <55% ma największy udział w grupach (1, 2 i 3), a w pozostałych grupach prawie nie występuje. Rozkład wartości

ciśnienia w grupach był dość równomierny. Ok. 70% przypadków to ciśnienia z przedziału 970-990hPa, a w pozostałych przypadkach ciśnienia są mniejsze od 970hPa, lub częściej większe od 990hPa.



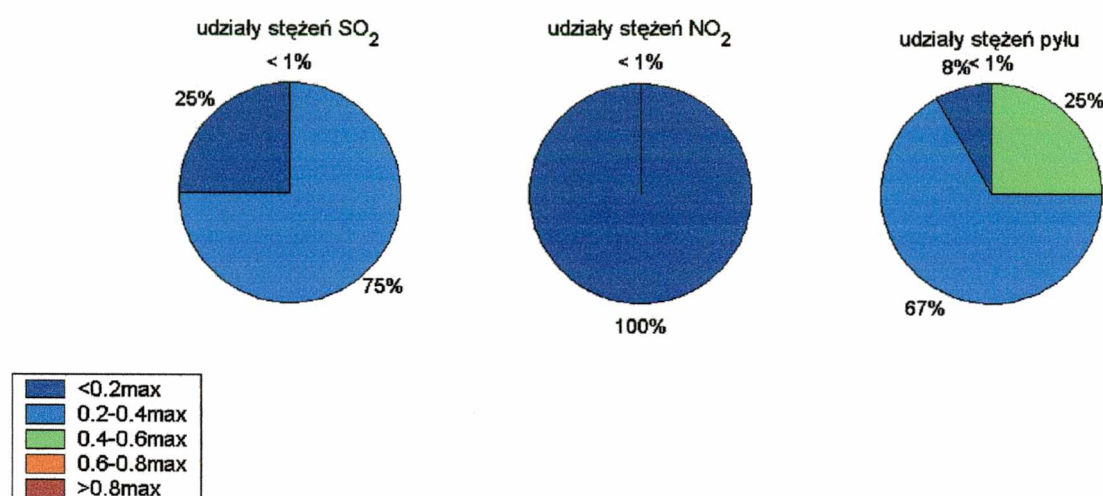
Rysunek 3.9. Rozmieszczenie grup warunków meteorologicznych w sektorach różnicy wiatrów

Wydzielone grupy nie mają więc wyraźnych profili meteorologicznych. Ciekawie przedstawia się natomiast ich profil pod względem zakresu zmienności stężeń jak też rozkładu ilościowego przypadków w różnych zakresach stężeń. Poszczególne grupy różnią się zakresem zmienności stężeń. Nie są to jednak różnice definitywne, gdyż większa część zakresu zmienności stężeń jest wspólna dla grup (rys.3.10).



Rysunek 3.10. Średnie stężenia zanieczyszczeń [$\mu\text{g}/\text{m}^3$] i ich odchylenia standardowe, w różnych grupach warunków meteorologicznych, na przykładzie grup dla stężeń bieżących

Obok ogólnego zakresu zmienności stężeń w grupach warunków meteorologicznych bardzo ważny jest rozkład ilościowy wartości stężeń w tych grupach. W 99% przypadków stężenia NO_2 są mniejsze od 20% wartości stężenia maksymalnego NO_2 . Wszystkie grupy warunków meteorologicznych mają prawie jednakową strukturę pod względem stężeń tego zanieczyszczenia. Dla SO_2 99% przypadków rozkłada się w grupach w różnym stopniu między stężenia w zakresie 0-20% wartości maksymalnej i 20-40% wartości maksymalnej stężenia SO_2 . Przeważają jednak grupy z dominującym pierwszym zakresem. Ilościowy rozkład stężeń pyłu w grupach charakteryzuje jeszcze większa złożoność. Występowaniu wymienionych już zakresów towarzyszy wyraźna reprezentacja przedziałów 40-60% i 60-80% wartości maksymalnej. Grupy warunków meteorologicznych różnią się proporcjami między ilością przypadków stężeń w poszczególnych zakresach. Przykład częstości występowania stężeń SO_2 , NO_2 i pyłu w obrębie grupy warunków meteorologicznych przedstawiono na rys.3.11.



Rysunek 3.11. Częstość występowania stężeń SO_2 , NO_2 i pyłu w zadanych przedziałach stężeń maksymalnych dla pierwszej grupy warunków meteorologicznych. Dotyczy stężeń bieżących.

3.3.2.2. Dyskusja wyników neuronowego prognozowania deterministycznego

Przeanalizowano wyniki prognoz 30-minutowych stężeń SO_2 , NO_2 i pyłu na podstawie zestawu 30-minutowych parametrów meteorologicznych. W skład zestawu wchodzi: prędkość i kierunek wiatru, wilgotność, temperatura i ciśnienie powietrza atmosferycznego. Rozważono przypadki, gdy podstawę prognozy stanowiły dane meteorologiczne rejestrowane równolegle z prognozowanym stężeniem, wyprzedzające prognozowane stężenie o 3, 6, 12, 24 i 48 godzin oraz przesunięte o okres czasu uwarunkowany maksymalnym współczynnikiem korelacji wzajemnej stężeń zanieczyszczeń i poszczególnych parametrów meteorologicznych.

Prognozy wykonywano w obrębie 14 grup parametrów meteorologicznych wyróżnionych w wyniku klasyfikacji pełnego zbioru parametrów meteorologicznych z zastosowaniem mapy Kohonena. Powstało w ten sposób 14 grup modeli. Ponieważ punktem odniesienia analiz są warunki meteorologiczne, modele stężeń różnych zanieczyszczeń należące do tej samej grupy są zbudowane w oparciu o ten sam zestaw warunków meteorologicznych. Modele tego samego zanieczyszczenia dla różnych opóźnień operują na różnych zestawach stężeń. Modele zbudowane dla zestawu warunków meteorologicznych, który maksymalizuje współczynniki ich korelacji wzajemnej ze stężeniami zanieczyszczeń różnią się od pozostałych zarówno pod względem składu zbioru warunków meteorologicznych jak i zbioru stężeń.

Wyniki modelowania analizowano w funkcji horyzontu prognozy i w funkcji grup warunków meteorologicznych. Poszukiwano odpowiedzi na pytanie o optymalną odległość czasową między momentem wystąpienia stężenia prognozowanego a zestawem warunków meteorologicznych stanowiących podstawę prognozy. Zadaniem analizy było również stwierdzenie, czy jakość prognozy i modeli prognostycznych zależy od wzorca warunków meteorologicznych, dla którego zbudowano model. Mogło to doprowadzić do wytypowania grup warunków meteorologicznych umożliwiających wykonywanie najlepszych prognoz stężeń zanieczyszczeń oraz wskazanie grup najgorszych z prognostycznego punktu widzenia.

Tabela 3.6. Zakres zmienności parametrów charakteryzujących jakość deterministycznych modeli prognostycznych oraz jakość prognoz stężeń SO_2 , NO_2 i pyłu

parametr modelu/prognozy	zanieczyszczenie		
	SO_2	NO_2	pył
współczynnik determinacji	0.15÷0.63	0.21÷0.57	0.04÷0.42
RMAE	0.46÷0.92	0.56÷1.06	0.63÷1.48
RMSE	14÷33	5÷21	26÷58
błąd modelu	0.29÷0.79	0.34÷0.68	0.34÷0.83
różnica średnich stężeń rzeczywistych i z modelu	-1.1÷3.5	-0.9÷0.8	-3.7÷5

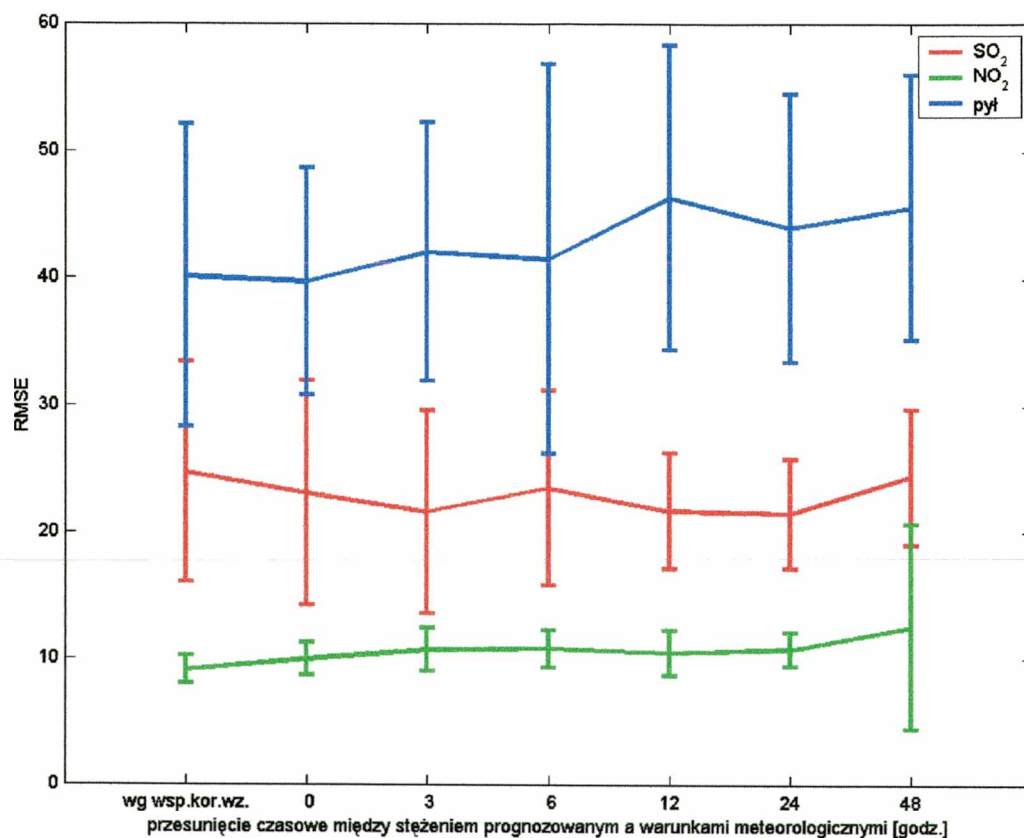
Tabela 3.7. Wartości dolnego kwantyla współczynnika determinacji i górnych kwantyli pozostałych parametrów dla deterministycznych modeli prognostycznych stężeń SO_2 , NO_2 i pyłu

parametr modelu/prognozy	zanieczyszczenie		
	SO_2	NO_2	pył
współczynnik determinacji	0.2÷0.32	0.27÷0.42	0.06÷0.26
RMAE	0.68÷0.78	0.72÷0.92	0.99÷1.18
RMSE	23÷31	10÷12	43÷53
błąd modelu	0.6÷0.71	0.5÷0.6	0.67÷0.79
różnica średnich stężeń rzeczywistych i z modelu	0.73÷2.4	-0.14÷0.6	1.2÷4.4

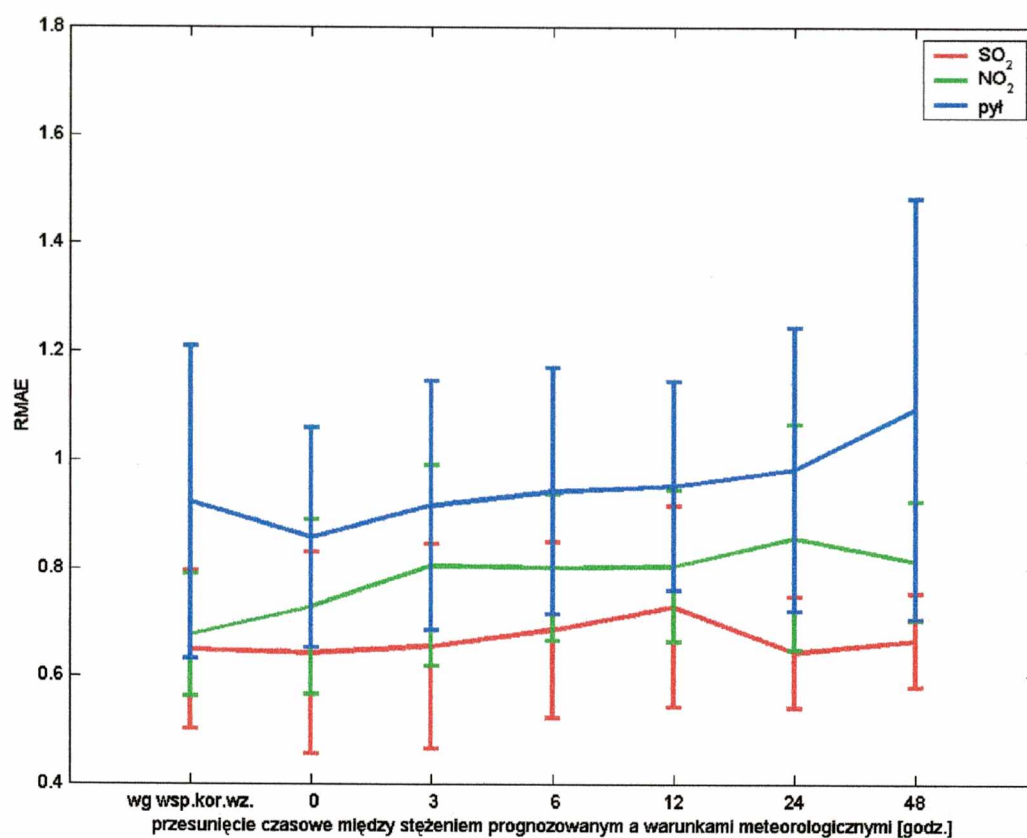
Na rysunkach (rys.3.12 do rys.3.17) przedstawiono parametry charakteryzujące jakość prognoz oraz jakość modeli prognostycznych stężeń SO_2 , NO_2 i pyłu w funkcji horyzontu prognozy. Są to RMAE, RMSE, współczynnik determinacji, błąd modelu oraz różnice średnich stężeń rzeczywistych i wyliczonych z modelu, uśrednione dla wszystkich 14 grup warunków meteorologicznych. W tab.3.6 podano pełne zakresy zmienności parametrów zgodnie z rysunkami (rys.3.12 do rys.3.17). Zarówno z rysunków jak i tabel wynika, że parametry zmieniają się w różnych zakresach dla poszczególnych zanieczyszczeń. W celu ułatwienia porównania ze sobą modeli zbudowanych dla różnych zanieczyszczeń wyliczono wartości dolnego kwantyla współczynników determinacji oraz górnego kwantyla pozostałych parametrów modelu i zamieszczono je w tab.3.7. Kwantyle dostarczają informacji o tym, od jakiej wartości są większe współczynniki determinacji 75% wszystkich wykonywanych prognoz oraz od jakiej wartości są mniejsze ich błędy. Zakres zmienności kwantyli parametrów jest mniejszy, niż pełen zakres zmienności parametrów i pokazuje, które wartości parametrów przeważają.

Spośród trzech zanieczyszczeń, które przeanalizowano najgorszymi parametrami charakteryzują się modele prognostyczne i prognozy stężeń pyłu. Uśrednione po grupach, w ramach kolejnych opóźnień, błędy prognozy zarówno RMSE, RMAE jak i część błędu, za którą odpowiada model są większe dla prognoz stężeń pyłu niż dla stężeń SO_2 i NO_2 (niebieska krzywa na rys.3.12, rys.3.13 i rys.3.16 leży najwyżej). Współczynniki determinacji dla modelu prognostycznego stężeń pyłu są mniejsze niż dla stężeń SO_2 i NO_2 (niebieska krzywa na rys.3.15 leży najniżej), natomiast różnice między średnimi rzeczywistych stężeń zanieczyszczeń a średnimi stężeń pochodzących z modelu na rys.3.17 przyjmują wartości porównywalne z różnicami dla NO_2 czy SO_2 .

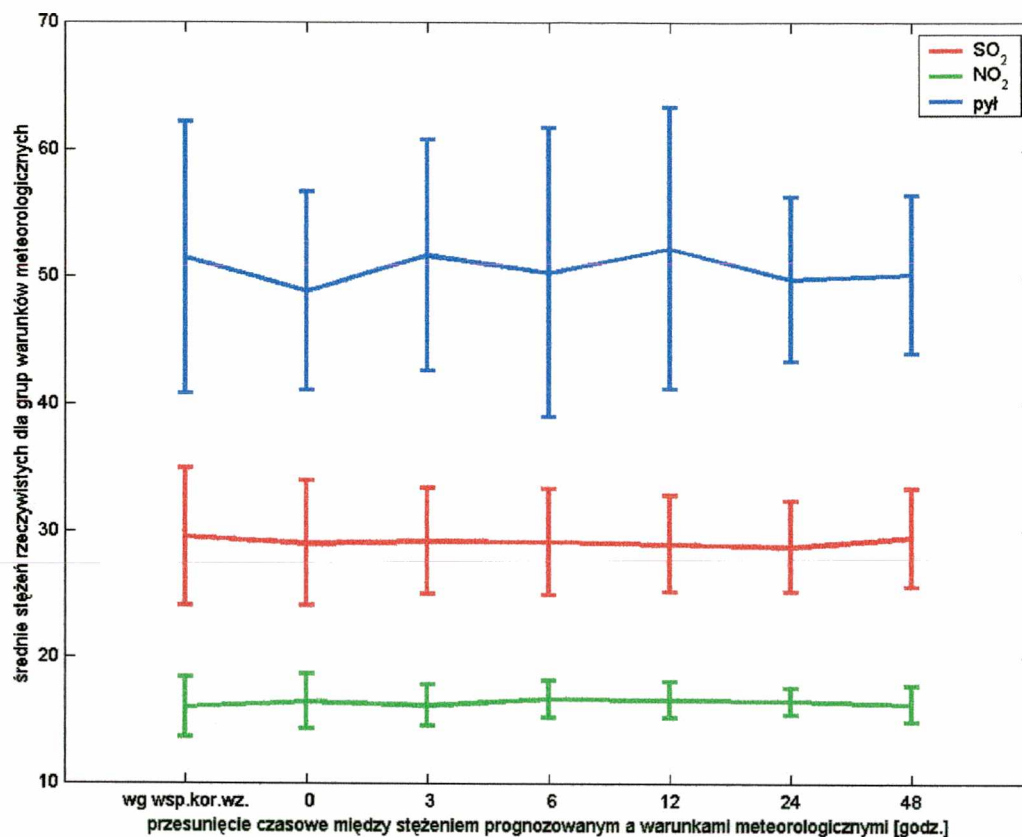
Jak wynika z rezultatów przedstawionych w tabelach (tab.3.6 i tab.3.7) oraz na rysunkach (rys.3.12 do rys.3.17), utrudniona jest jednoznaczna ocena przewagi modeli prognostycznych i prognoz stężeń NO_2 nad SO_2 lub odwrotnie. W całym zakresie opóźnień współczynniki determinacji dla modeli NO_2 są wyraźnie większe, niż dla pyłu i nieznacznie większe, niż dla modeli prognostycznych stężeń SO_2 (rys.3.15). Średni błąd kwadratowy prognozy jest najniższy dla prognozy stężeń NO_2 (rys.3.12). Część tego błędu, za którą odpowiada model jest zawsze mniejsza niż dla pyłu, a w większości przypadków mniejsza od błędu modelu stężeń SO_2 (rys.3.16). Z zestawienia błędów względnych RMAE (rys.3.13) dla prognoz stężeń NO_2 i SO_2 wynika jednak, że w kategoriach błędu względnego wyniki prognoz stężeń SO_2 są lepsze niż NO_2 . Te rezultaty są zgodne z wynikami analizy przygotowawczej do modelowania deterministycznego, gdzie stwierdzono, że najsłabszy związek łączy parametry meteorologiczne i stężenia pyłu. Związek stężeń NO_2 i SO_2 z tymi parametrami charakteryzuje podobne natężenie choć jest on wyraźnie silniejszy niż dla pyłu.



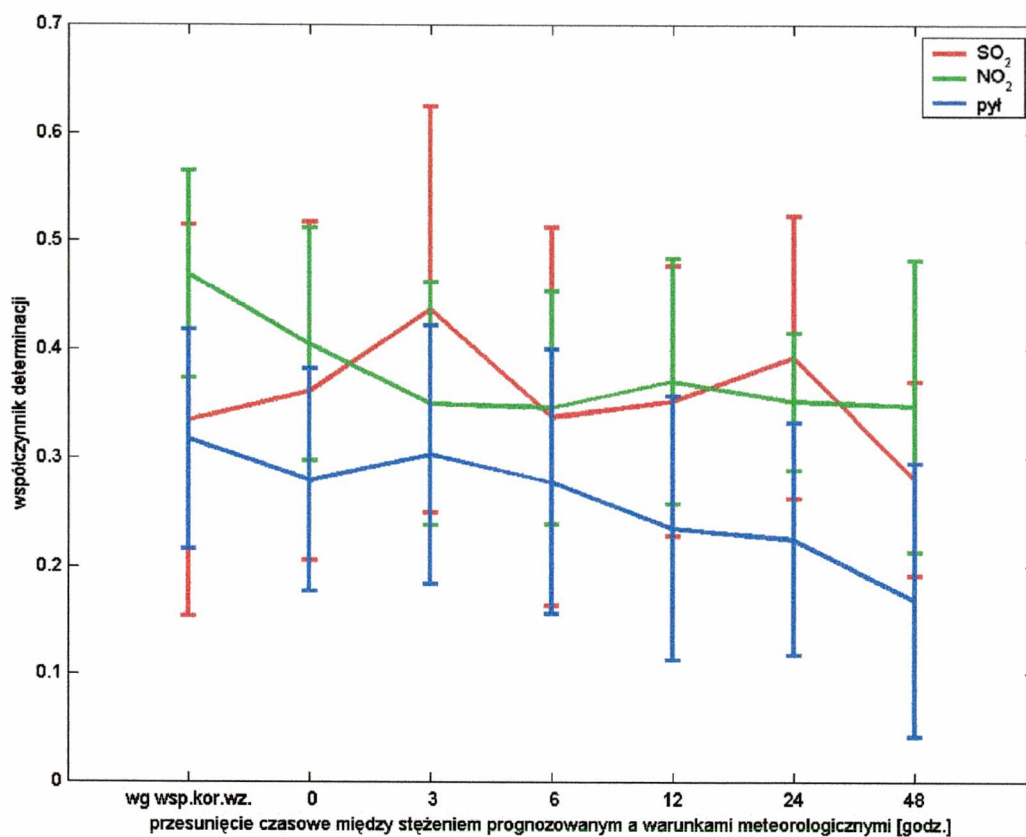
Rysunek 3.12. Zakres zmienności średniego błędu kwadratowego prognozy 30-minutowych stężeń SO₂, NO₂ i pyłu [μg/m³] w funkcji opóźnienia zestawu warunków meteorologicznych



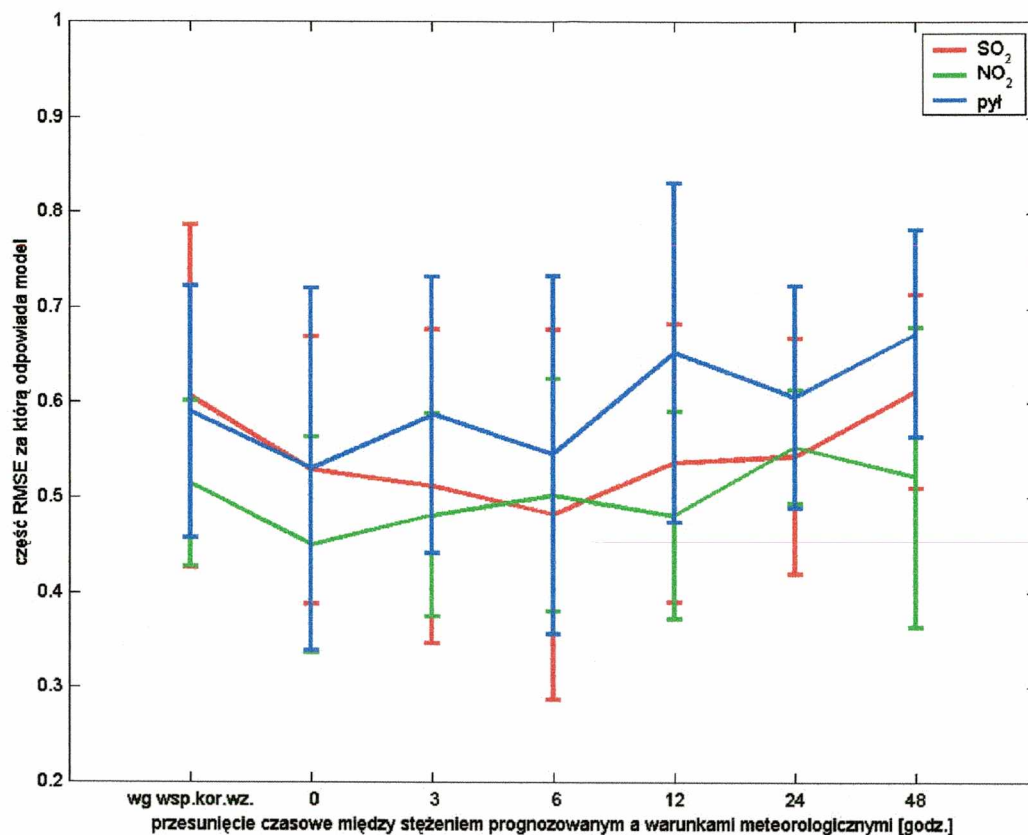
Rysunek 3.13. Zakres zmienności względnego średniego błędu absolutnego prognozy 30-minutowych stężeń SO₂, NO₂ i pyłu w funkcji opóźnienia zestawu warunków meteorologicznych



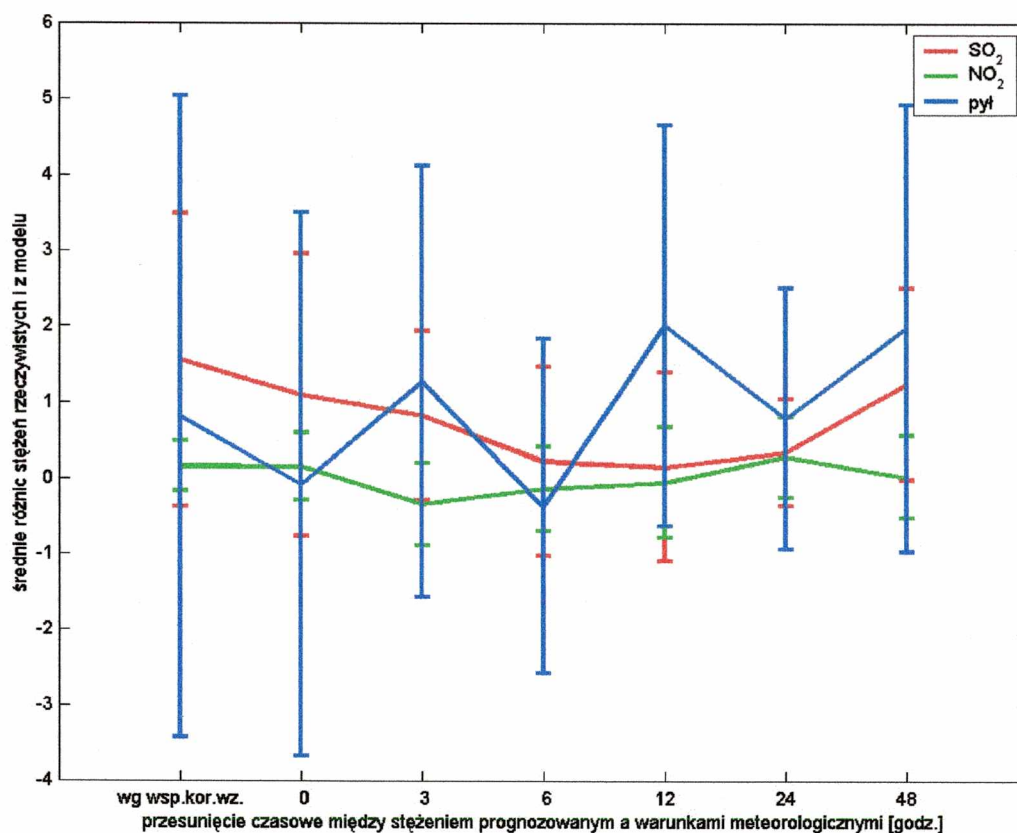
Rysunek 3.14. Zakres zmienności średnich rzeczywistych stężeń SO_2 , NO_2 i pyłu [$\mu\text{g}/\text{m}^3$] w zbiorach testowych dla grup warunków meteorologicznych, w funkcji opóźnienia zestawu warunków meteorologicznych



Rysunek 3.15. Zakres zmienności współczynnika determinacji modeli prognostycznych 30-minutowych stężeń SO_2 , NO_2 i pyłu w funkcji opóźnienia zestawu warunków meteorologicznych



Rysunek 3.16. Zakres zmienności części średniego błędu kwadratowego, za którą odpowiadają modele prognostyczne 30-minutowych stężeń SO_2 , NO_2 i pyłu w funkcji opóźnienia zestawu warunków meteorologicznych



Rysunek 3.17. Zakres zmienności średnich różnic między stężeniami rzeczywistymi a pochodzącymi z modeli prognostycznych stężeń SO_2 , NO_2 i pyłu [$\mu\text{g}/\text{m}^3$] w funkcji opóźnienia zestawu warunków meteorologicznych

Jedno z pytań, na które poszukiwano odpowiedzi analizując jakość modeli prognostycznych i prognoz, dotyczy optymalnego horyzontu prognozy. W przypadku SO_2 optymalnymi są model i prognoza stężeń na podstawie warunków meteorologicznych sprzed doby. Co do wartości współczynnika determinacji (0.39 ± 0.13), RMAE (0.64 ± 0.1), RMSE (21 ± 4), jak też części błędu prognozy za którą odpowiada model (0.54 ± 0.12), jest ona prawie identyczna z prognozą o horyzoncie 3 godzin. Jednakże oparcie prognozy stężeń SO_2 na warunkach meteorologicznych sprzed 24 godzin umożliwia większe zbliżenie średniej stężeń rzeczywistych do średniej stężeń uzyskiwanych z modelu (0.34 ± 0.7 / 1.2 ± 1.9) oraz wiąże się z uzyskaniem mniejszej wariancji błędu względnego i kwadratowego, niż w przypadku pozostałych opóźnień.

Na tle prognoz wykonywanych w oparciu o sytuację meteorologiczną, niekorzystnie wypada prognoza bazująca na zestawie warunków meteorologicznych maksymalizującym współczynniki korelacji wzajemnej stężeń SO_2 i parametrów meteorologicznych. Przyczyną tego jest prawdopodobnie fakt, że maksima współczynników korelacji wzajemnej dla pełnego zbioru warunków meteorologicznych są inne, niż w jego podzbiorach. W efekcie rezultaty prognozy są nawet gorsze, niż w przypadku prognozy opartej na warunkach meteorologicznych pochodzących sprzed dwóch dni.

Rezultaty prognoz stężeń NO_2 są najlepsze, jeżeli podstawę prognozy stanowi zestaw warunków meteorologicznych maksymalizujący współczynniki korelacji wzajemnej stężeń NO_2 i poszczególnych parametrów meteorologicznych. Parametry modelu i prognoz o najwyższej jakości są następujące: współczynnik determinacji (0.47 ± 0.1), RMAE (0.68 ± 0.11), RMSE (9 ± 1), części błędu prognozy, za którą odpowiada model (0.51 ± 0.09), odchylenie średniej stężeń rzeczywistych od średniej stężeń pochodzących z modelu (0.16 ± 0.32). Wydaje się, że przesunięcia czasowe, którym odpowiadają maksima współczynników korelacji wzajemnej mają to samo położenie w podokresach pełnego zbioru warunków meteorologicznych jak w całym zbiorze.

Parametry zbliżone do optymalnych, aczkolwiek gorsze, charakteryzują model prognostyczny stężeń NO_2 bazujący na warunkach meteorologicznych rejestrowanych równoległe z prognozowanym stężeniem. Wraz ze zwiększaniem odległości czasowej między prognozą a występowaniem bazowych warunków meteorologicznych jakość modelu prognostycznego i prognozy ulega pogorszeniu. Tę tendencję można odczytać z rys.3.12, rys.3.13, rys.3.15 i rys.3.16. Wyraża ją spadek wartości współczynników determinacji, a wzrost wartości błędów RMAE, RMSE oraz błędu, za który odpowiada model. Zmiany parametrów modeli w funkcji opóźnienia są małe mają łagodny przebieg.

Najlepsze prognozy stężeń pyłu są oparte na bieżącym zestawie parametrów meteorologicznych lub na zestawie minimalizującym współczynniki korelacji wzajemnej. Pierwsze rozwiązanie charakteryzują nieznacznie mniejsze błędy wraz z ich odchyleniami RMAE (0.9 ± 0.2 / 0.92 ± 0.29), RMSE (40 ± 9 / 40 ± 11), błąd modelu (0.5 ± 0.19 / 0.59 ± 0.13) i mniejsza przeciętna różnica między stężeniami rzeczywistymi i uzyskanymi z modelu (-0.09 ± 3.9 / 0.8 ± 4.2). Jednakże współczynnik determinacji tego modelu jest nieznacznie mniejszy (0.28 ± 0.1 / 0.32 ± 0.1). Z podanych informacji wynika, że za najlepsze można uznać model prognostyczny i prognozę stężeń pyłu w oparciu o bieżące warunki meteorologiczne.

Parametry optymalnych modeli prognostycznych i prognoz stężeń SO_2 , NO_2 i pyłu zebrano i przedstawiono w tab.3.8. Wydaje się, że optymalny horyzont prognozy powinien podlegać identyfikacji osobno w odniesieniu do każdego zanieczyszczenia i lokalizacji punktu pomiarowego.

Tabela 3.8. Wartości parametrów charakteryzujących jakość optymalnych deterministycznych modeli prognostycznych stężeń SO_2 , NO_2 i pyłu oraz jakość prognoz wykonywanych za pomocą tych modeli

parametr modelu/prognozy	zanieczyszczenie		
	SO_2 (24h)	NO_2 (współ. korel.)	pył (bież)
współczynnik determinacji	0.39 ± 0.13	0.47 ± 0.1	0.28 ± 0.1
RMAE	0.64 ± 0.1	0.68 ± 0.11	0.86 ± 0.2
RMSE	21 ± 4	9 ± 1	40 ± 9
błąd modelu	0.54 ± 0.12	0.51 ± 0.09	0.53 ± 0.19
różnica średnich stężeń rzeczywistych i z modelu	0.34 ± 0.7	0.16 ± 0.32	-0.09 ± 3.59

Warto zestawić parametry charakteryzujące optymalne neuronowe modele prognostyczne ze współczynnikami korelacji wzajemnej stężeń zanieczyszczeń i parametrów meteorologicznych stanowiących wstępny wskaźnik prawdopodobnych rezultatów prognozowania, przy wykorzystaniu modelu liniowego. Zgodnie z tab.3.9 ich wartości maksymalne są co do wartości bezwzględnej bardzo małe, generalnie poniżej 0.1, za wyjątkiem temperatury, w którym to przypadku mogą sięgać 0.4. Kwadrat współczynnika korelacji między zmiennymi jest równy współczynnikowi determinacji jednej zmiennej przez drugą, czyli części wariancji jednej zmiennej, którą można wytłumaczyć za pomocą innej zmiennej. Gdyby parametry meteorologiczne były od siebie niezależne, o czym wiemy, że jest nieprawdą, maksymalne oczekiwane wartości współczynników determinacji modelu liniowego, stanowiłyby sumy kwadratów wartości zamieszczonych w tabeli, wyliczone dla jej wierszy. Wynosiłyby odpowiednio 0.14 dla NO_2 , 0.19 dla SO_2 i 0.03 dla pyłu. Rzeczywiste wartości są mniejsze w związku z wzajemną współzależnością

parametrów meteorologicznych. Jak wynika z tab.3.8, model neuronowy pozwala uzyskać wyraźnie wyższe wartości współczynników determinacji, czyli lepiej wykorzystać informację meteorologiczną dla określenia zmienności stężeń zanieczyszczeń. Wiąże się to prawdopodobnie z identyfikacją i wykorzystaniem przez ten model nieliniowej części zależności między parametrami meteorologicznymi a stężeniami zanieczyszczeń.

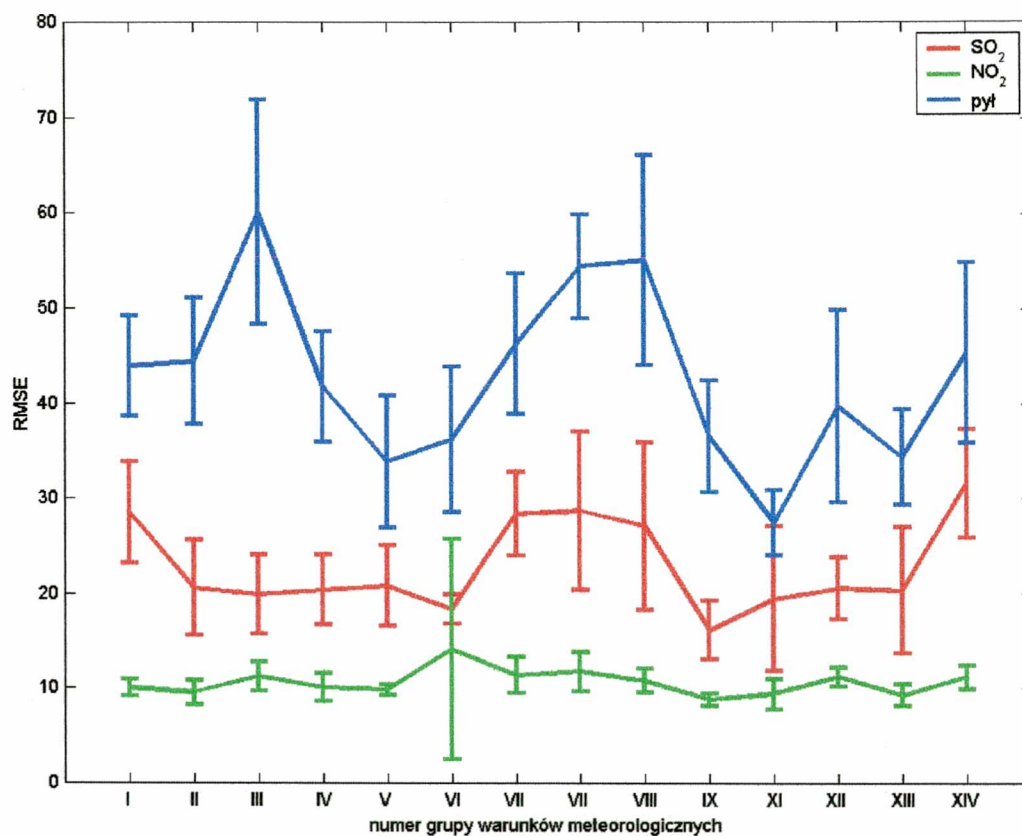
Tabela 3.9. Maksymalne wartości współczynników korelacji wzajemnej między stężeniami zanieczyszczeń a parametrami meteorologicznymi dla zbioru danych poddanego modelowaniu

zanieczyszczenie	parametr meteorologiczny				
	kierunek wiatru	prędkość wiatru	wilgotność	temperatura	ciśnienie
NO ₂	0.01	0.15	0.15	0.3	0.01
SO ₂	0.07	0.1	0.1	0.4	0.01
pył	0.01	0.11	-0.1	0.05	0.1

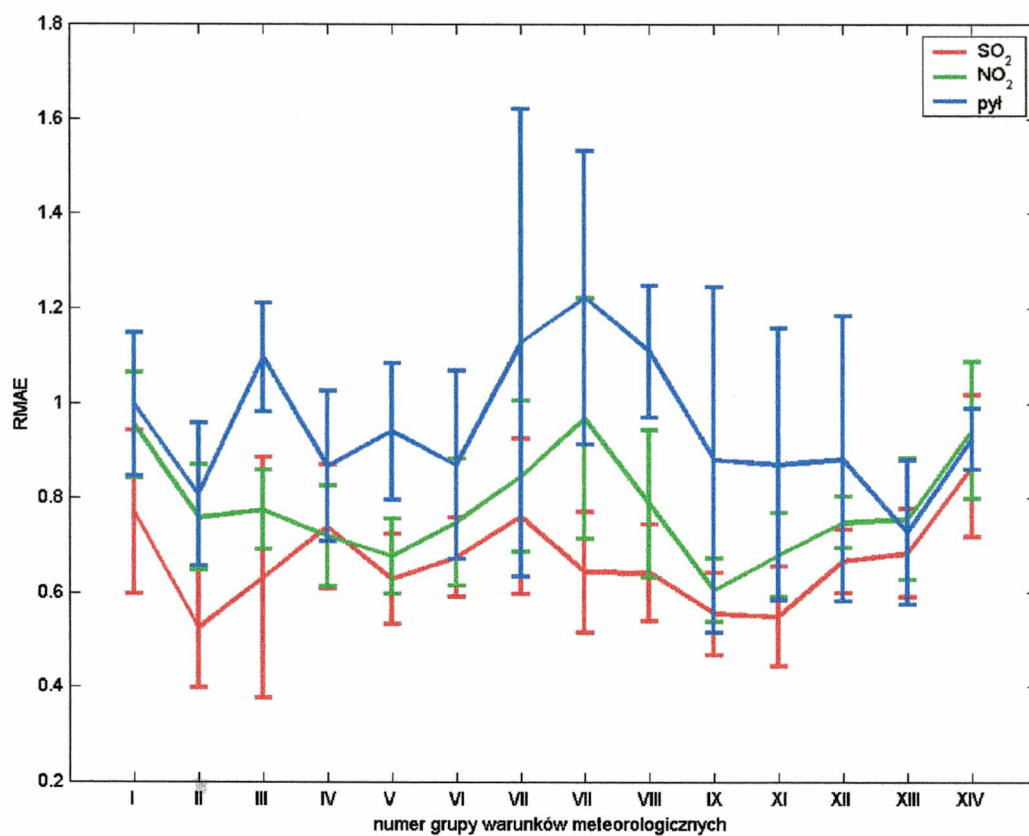
Kolejne wnioski wynikają z analizy wyników modelowania stężeń zanieczyszczeń na podstawie warunków meteorologicznych w funkcji grup warunków meteorologicznych. Pełny zakres zmienności parametrów charakteryzujących jakość modeli oraz jakość prognoz przedstawiono w tab.3.10. Korelujące z tabelą wykresy zamieszczono na rysunkach od rys.3.18 do rys.3.23. Tym razem wartości parametrów są uśredniane dla opóźnień w ramach każdej grupy, a wykresy przedstawiają wyliczoną średnią parametru i zakres jej zmienności w obrębie kolejnych grup warunków meteorologicznych.

Tabela 3.10. Zakres zmienności parametrów charakteryzujących jakość deterministycznych modeli prognostycznych stężeń SO₂, NO₂ i pyłu oraz jakość prognoz wykonywanych za pomocą tych modeli. Parametry dla opóźnień uśredniono w ramach grup

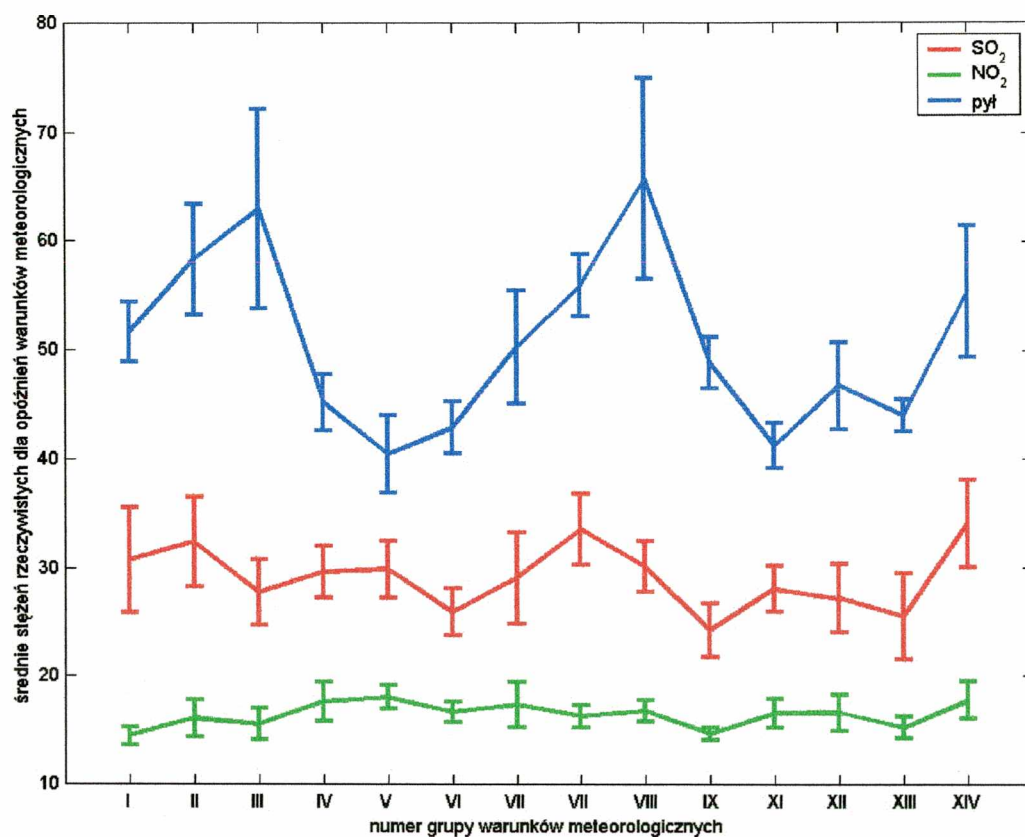
parametr modelu/prognozy	zanieczyszczenie		
	SO ₂	NO ₂	pył
współczynnik determinacji	0.07÷0.70	0.16÷0.57	0.07÷0.53
RMAE	0.38÷1.02	0.54÷1.22	0.52÷1.62
RMSE	12÷37	3÷26	24÷72
błąd modelu	0.24÷0.77	0.23÷0.73	0.25÷0.84
różnica średnich stężeń rzeczywistych i z modelu	-1.4÷4.1	-1÷1	-7.5÷6.4



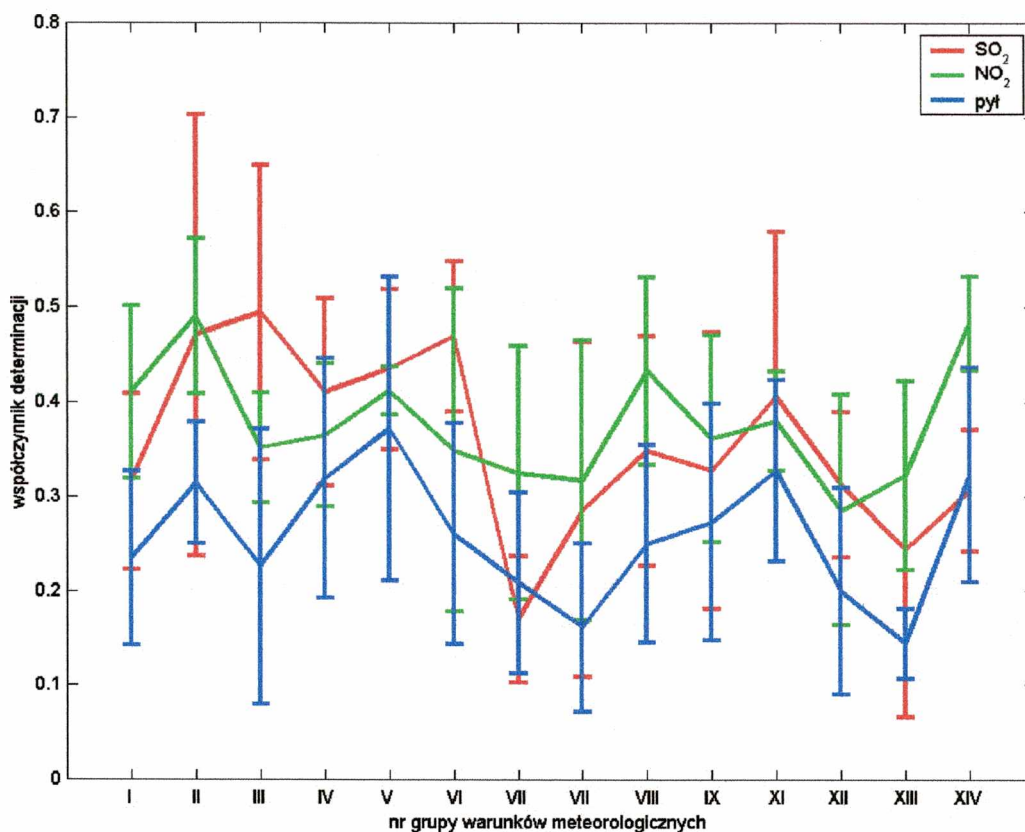
Rysunek 3.18. Zakres zmienności średniego błędu kwadratowego prognozy 30-minutowych stężeń SO_2 , NO_2 i pyłu [$\mu\text{g}/\text{m}^3$] dla 14 grup warunków meteorologicznych



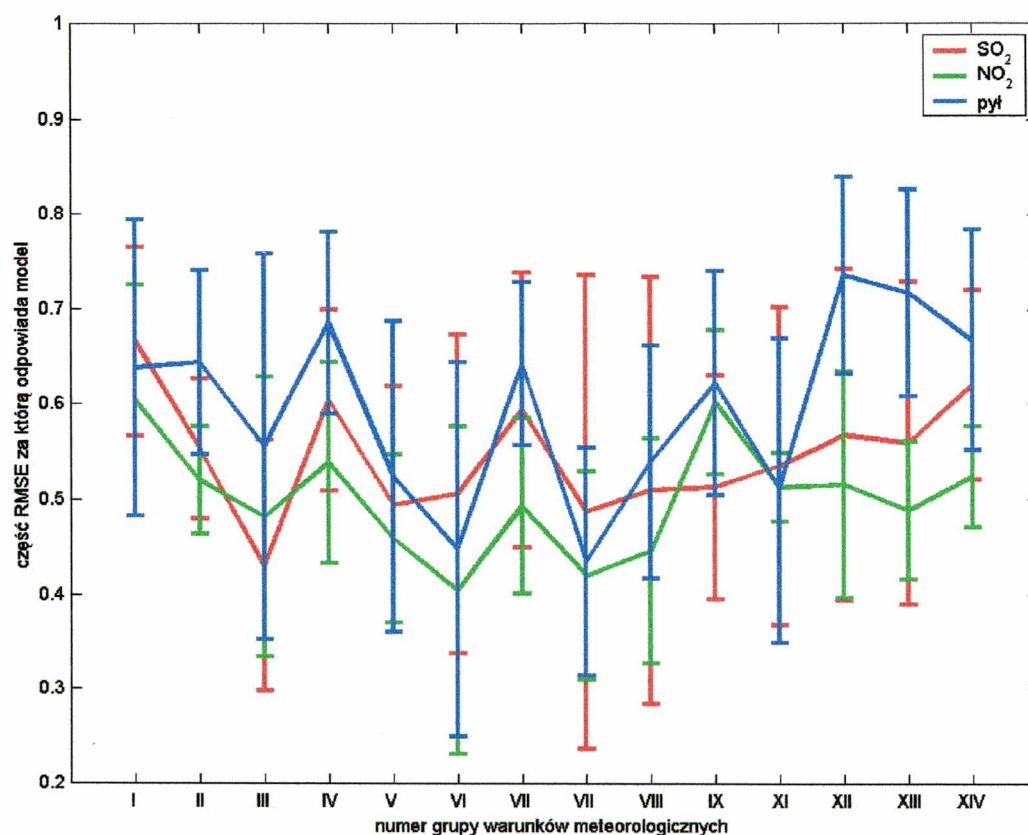
Rysunek 3.19. Zakres zmienności względnego średniego błędu absolutnego prognozy 30-minutowych stężeń SO_2 , NO_2 i pyłu dla 14 grup warunków meteorologicznych



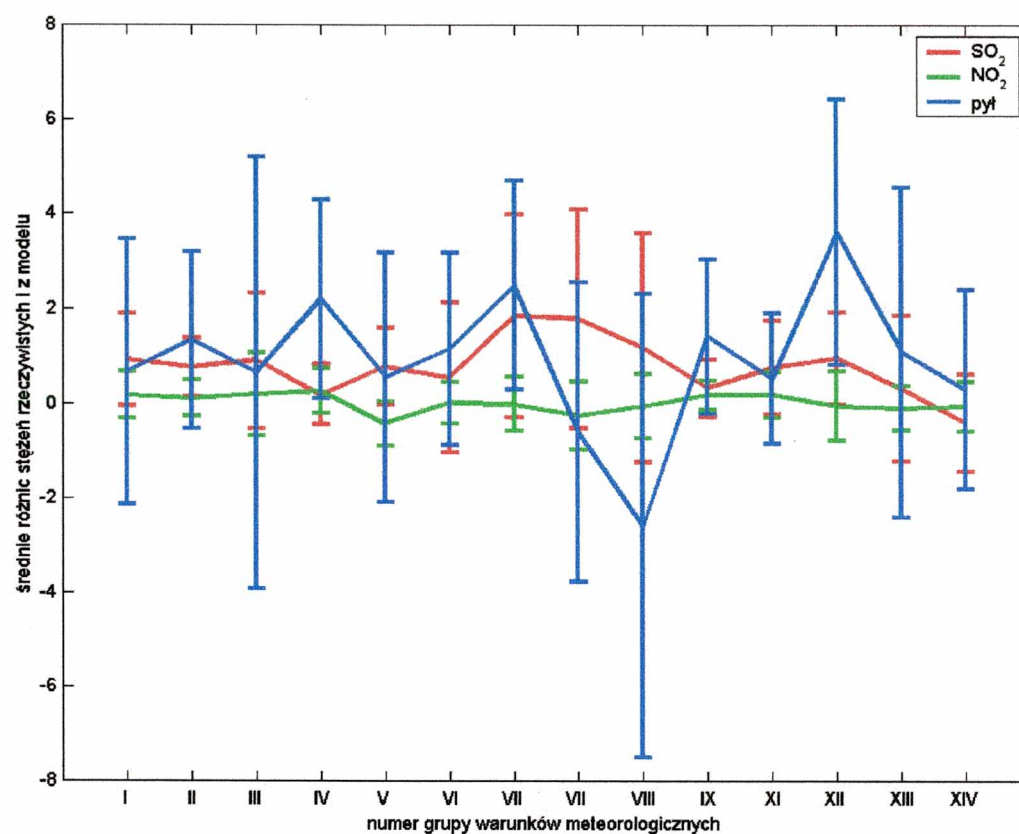
Rysunek 3.20. Zakres zmienności średnich rzeczywistych stężeń SO_2 , NO_2 i pyłu w zbiorach testowych [$\mu\text{g}/\text{m}^3$] dla różnych opóźnień warunków meteorologicznych dla 14 grup warunków meteorologicznych



Rysunek 3.21. Zakres zmienności współczynnika determinacji modeli prognostycznych 30-minutowych stężeń SO_2 , NO_2 i pyłu dla 14 grup warunków meteorologicznych



Rysunek 3.22. Zakres zmienności części błędu, za którą odpowiadają modele prognostyczne 30-minutowych stężeń SO₂, NO₂ i pyłu dla 14 grup warunków meteorologicznych



Rysunek 3.23. Zakres zmienności średnich różnic między stężeniami rzeczywistymi a pochodzącymi z modeli prognostycznych 30-minutowych stężeń SO₂, NO₂ i pyłu [μg/m³] dla 14 grup warunków meteorologicznych

Z wykresów zamieszczonych na rysunkach (rys.3.18 do rys.3.23) wynika, że wartości RMSE, RMAE, współczynnika determinacji, błędu modelu i odległości stężeń rzeczywistych od wyliczonych z modelu nie są jednakowe dla wszystkich grup warunków meteorologicznych. Wartości tych współczynników zwykle rosną lub maleją wspólnie dla wszystkich zanieczyszczeń. Grupy (1, 7, 8, 9, 11 i 14) charakteryzują się np. dużymi wartościami błędu RMSE i RMAE dla wszystkich zanieczyszczeń. Jakość prognoz wykonywanych dla parametrów meteorologicznych należących do tych grup jest najgorsza. Natomiast do najlepszych należą prognozy wykonywane za pomocą modeli zbudowanych dla grup (2, 4, 5, 6, 10, 11, 12 i 13), o czym świadczą odpowiednio małe wartości błędów względnego i średniego kwadratowego. Wszystkie grupy, dla których wystąpiły najlepsze prognozy, za wyjątkiem drugiej odpowiadają południowo-zachodnim i zachodnim kierunkom wiatru. Należą tu grupy charakteryzujące się mniejszymi (0-3m/s) prędkościami wiatru jak (2, 5 i 6) oraz większymi (3-6m/s) jak grupy (4, 10, 11, 12 i 13). Grupy (10, 11, 12 i 13) cechuje przewaga temperatur powyżej 10°C i wilgotności w zakresie 55-75%. Modele odpowiadają za najmniejszą część średniego błędu kwadratowego prognozy w przypadku grup (3, 6, 8, 11). Z tymi grupami związane są najniższe stężenia zanieczyszczeń. Współczynniki determinacji osiągają największe wartości w grupach (2, 5, 9, 11 i 14) co oznacza, że stężenia wyliczone z modeli opartych na warunkach meteorologicznych należących do tych grup są najsilniej skorelowane z rzeczywistymi stężeniami zanieczyszczeń. Są to warunki związane z małymi prędkościami wiatru na kierunkach wschodnim do południowo-wschodniego oraz południowo-zachodnim i nieco większymi prędkościami wiatru z kierunku północnego. Cechują się przewagą niskich temperatur i małą wilgotnością (<55%), jak w przypadku grup (1 i 2) lub przewagą wilgotności wysokiej (>75%). Należą tu grupy stężeń wysokich (2, 9 i 14) oraz niskich (5 i 11). Jak widać trudno podać przejrzyste kryteria meteorologiczne przewagi jednych prognoz lub modeli prognostycznych nad innymi.

Ogólne podobieństwo zmian jakości prognoz i modeli stężeń wszystkich zanieczyszczeń przy zmianie grupy warunków meteorologicznych sugeruje, że warunki te są nośnikiem informacji wspólnej dla zmienności stężeń wszystkich zanieczyszczeń. Wynika stąd, że układy warunków meteorologicznych wpływają na stężenia różnych zanieczyszczeń w podobny sposób. Jeśli następuje poprawa jakości prognoz, to dotyczy ona z reguły wszystkich zanieczyszczeń. Podobnie jest z jakością modelu. Największe podobieństwo zmian jakości modelu w funkcji grup warunków meteorologicznych występuje między pyłem i NO₂ (rys.3.21 i rys.3.22). Odstępstwo SO₂ prawdopodobnie nie wynika stąd, że warunki meteorologiczne inaczej wpływają na zmienność stężeń SO₂ niż NO₂ i pyłu. Jest to raczej skutek wystąpienia lokalnej emisji SO₂, która

nakłada się na tło adwekcyjne i wpływ wspólnego źródła wszystkich zanieczyszczeń - Elektrowni Turów, powodując, że stężenia tego zanieczyszczenia zachowują się w tych samych warunkach meteorologicznych inaczej niż NO_2 i pyłu.

Z drugiej strony dany układ parametrów meteorologicznych w różnym stopniu determinuje zmienność stężeń poszczególnych zanieczyszczeń. Informują o tym różnice między wartościami parametrów modeli dla poszczególnych zanieczyszczeń w obrębie grup. Modele prognostyczne stężeń pyłu cechują najgorsze parametry. Porównywalną jakością charakteryzują się modele stężeń dwutlenku azotu i dwutlenku siarki.

Zestawienie zmienności RMSE i RMAE z zakresem zmienności międzygrupowej stężeń zanieczyszczeń nasuwa wniosek, że jakość prognozy jest uwarunkowana poziomem stężeń modelowanych oraz wariancją tych stężeń. W tym miejscu istotną jest uwaga, że charakter zmienności stężeń w zbiorach testowych przedstawiony na rys.3.19 jest reprezentatywny dla zakresu zmienności stężeń w zbiorach uczących. Grupy testowe o niskich średnich stężeniach pochodzą z grup uczących o niskich średnich stężeniach i grupy testowe o małej wariancji stężeń pochodzą z grup testowych o małej wariancji stężeń. Im niższe stężenia i im mniejsza ich wariancja (rys.3.20) tym lepsza jakość prognozy identyfikowalna błędami RMSE i RMAE (rys.3.18 i rys.3.19). Jeżeli tak, to można postawić tezę, że jakości prognoz stężeń pyłu są najgorsze z tego powodu, że stężenia tego zanieczyszczenia odznaczają się najwyższymi wartościami i największą zmiennością, a nie wynikają stąd, że warunki meteorologiczne najsłabiej determinują zmienność stężeń pyłu.

Nie jest jednak prawdziwe twierdzenie, że parametry meteorologiczne w tym samym stopniu wpływają na stężenia wszystkich zanieczyszczeń, a to, że prognozy w kategoriach RMSE i RMAE dla jednych wypadają lepiej, niż dla drugih jest tylko skutkiem charakteru zmienności stężeń danego zanieczyszczenia. Wynika to z niemożności utrzymania tezy o zależności współczynnika determinacji modelu od poziomu i wariancji stężeń zanieczyszczeń. Ten współczynnik jest zaś miarą identyfikowanego przez model wpływu parametrów meteorologicznych na kształtowanie się stężeń zanieczyszczeń. Jego wartości są wyraźnie najniższe dla pyłu co sugeruje, że z warunków meteorologicznych można pozyskać najmniej informacji dotyczących tego zanieczyszczenia (tab.3.6, tab.3.7, tab.3.8 i tab.3.10, rys.3.15 i rys.3.21). Wyraźnie wyższe są współczynniki determinacji modeli opracowanych dla stężeń NO_2 . To zanieczyszczenie silniej podlega wpływowi warunków meteorologicznych. Podlega mu jednak w sposób identyczny jak pył. Krzywe niebieska (pył) i zielona (NO_2) na rys.3.21 są prawie dokładnie swoimi równoległymi przesunięciami. Inaczej układa się krzywa reprezentująca zmienność współczynnika determinacji stężeń SO_2 .

3.3.3. Prognoza z wykorzystaniem modelu niedeterministycznego

Modelowanie niedeterministyczne opiera się na wykorzystaniu informacji zawartej w szeregu czasowym zmiennej modelowanej bez posługiwania się żadnymi innymi zmiennymi objaśniającymi. Stężenie zanieczyszczenia jest prognozowane w oparciu o wartości, które przyjmowało w przeszłości. Ze względu na to, że w modelu niedeterministycznym nie uwzględnia się żadnych zmiennych zewnętrznych wpływających na wartość zmiennej opisywanej, konieczne jest przyjęcie założenia o ich niezmienności. W przeciwnym przypadku bazowanie wyłącznie na prawidłowości ujawniającej się w szeregu czasowym zmiennej jest nieuprawnione. Stąd przydatność modelowania niedeterministycznego wyłącznie w prognozowaniu krótkoterminowym.

Modelowanie niedeterministyczne serii czasowych stężeń zanieczyszczeń można zrealizować za pomocą sieci neuronowej. Wymaga to nauczania sieci prognozowania przyszłych wartości stężeń na podstawie ich wartości przeszłych. Zbiór uczący należy więc zbudować z fragmentów szeregu czasowego stężeń, podając równocześnie na wyjście sieci, jako wartości docelowe, stężenia wyprzedzające wektor wejściowy o żądany horyzont prognozy.

3.3.3.1. Propozycja rozwiązania problemu

Pierwsze zagadnienie, związane z zastosowaniem podejścia niedeterministycznego, w modelowaniu dotyczy postaci zależności między przesuniętymi względem siebie wartościami stężeń zanieczyszczeń w ich szeregu czasowym. W ogólnym przypadku sieć neuronowa nie dostarcza odpowiedzi. Pozwala jednak ocenić charakter zależności tzn. stwierdzić jej liniowość bądź nieliniowość. Wniosek jest istotny, gdyż może stanowić argument za stosowaniem lub odrzuceniem modeli liniowych w modelowaniu serii czasowych stężeń zanieczyszczeń. Dla zbadania problemu zaproponowano trzy metody neuronowe modelowania serii czasowych stężeń. Od najprostszej, w formie neuronowego filtra liniowego poprzez sieć neuronową z neuronem o liniowej funkcji przejścia, a skończywszy na sieci zbudowanej z neuronów o nieliniowych funkcjach transferu. Jeżeli nieliniowość decyduje o kształcie zależności między stężeniami następującymi po sobie w szeregu czasowym, to sieć zbudowana z neuronów o nieliniowych funkcjach przejścia poprawi wynik uzyskany za pomocą sieci z liniową funkcją transferu.

Drugie zagadnienie dotyczy długości fragmentu serii czasowej, która wystarcza do opisu prognozowanego stężenia następującego po tym fragmencie. Analiza współczynników autokorelacji i autokorelacji cząstkowej wykazała brak okresowości w serii czasowej stężeń

30-minutowych i istotną zależność stężeń od wartości bezpośrednio je poprzedzających. Opierając się na tych spostrzeżeniach dobrano długości wektorów uczących. Neuronowy filtr liniowy uczono na podstawie wektorów wybieranych z szeregu czasowego stężeń 30-minutowych przez okno o długości od 1 do 24 elementów. Długość wektorów uczących prezentowanych sieci z liniową oraz nieliniową funkcją transferu zmieniano w zakresie od 1 do 96 elementów. W ten sposób przeanalizowano możliwość wpływu na stężenie prognozowane jego historii do dwóch dób wstecz. Neuronowy filtr liniowy budowano w dwóch wersjach: dla danych z dwóch dni oraz tygodnia. Pozostałe sieci uczono na miesiącu danych.

Trzecie zagadnienie stanowi horyzont prognozy. Jak już powiedziano, metody niedeterministyczne nadają się do prognozowania krótkoterminowego. Optymalny horyzont prognozy jest związany z okresem czasu, po którym w serii czasowej stężeń zauważalny jest wpływ zmiany czynników zewnętrznych. Jakość prognozy wykonywanej na podstawie modelu niedeterministycznego dla dłuższych horyzontów czasowych jest bardzo słaba. Przeanalizowano szereg horyzontów prognozy 30-minutowych stężeń SO_2 , NO_2 i pyłu w poszukiwaniu horyzontów optymalnych. W przypadku neuronowego filtra liniowego rozważano horyzonty 0.5, 1, 3, 6 i 9 godzin, a w związku z sieciami o liniowych i nieliniowych funkcjach transferu jeszcze dodatkowo horyzont 12 godzin.

Dodatkowe pytanie dotyczyło skuteczności prognozowania w warunkach łagodnego przebiegu stężeń i przebiegu obejmującego epizody. Poszukując odpowiedzi przeprowadzono test filtra neuronowego na dwóch różnych zbiorach. Jeden charakteryzował się łagodną zmiennością stężeń, a w drugim wystąpił ich gwałtowny skok. Pozostałe sieci testowano na okresie 1 dnia wspólnego dla wszystkich zanieczyszczeń. Jednak prognozowane przebiegi różniły się wyraźnie swoim charakterem.

Zaproponowano też dwa różne tryby uczenia sieci. Tryb adaptacyjny w przypadku filtra neuronowego i tryb batch dla sieci o liniowych i nieliniowych funkcjach transferu. Rozwiązanie adaptacyjne wdrożono w takiej formie, że filtr testowano na jednym przypadku testowym odległym o horyzont prognozy od ostatniego przypadku włączonego do zbioru uczącego. W trybie nauki sieć podprowadzano na odległość horyzontu prognozy do przypadku testowego korzystając z okna o długości wektora uczącego przesuwającego się wzdłuż szeregu czasowego stężeń. Każde przesunięcie okna wiązało się z uaktualnieniem wag sieci, co jest cechą nauki sieci w trybie adaptacyjnym. Wobec tego utrwalony w strukturze sieci obraz zależności między jej wejściami i wyjściami odzwierciedlał silniej zależność występującą w końcowej części zbioru uczącego, niż jego fragmencie początkowym. Natomiast uczenie w trybie batch zastosowane dla

sieci z liniowym oraz nieliniowym neuronem wiązało się z korektą wag sieci po prezentacji całego zbioru uczącego. Wszystkie przypadki prezentowane sieci były więc równoprawne. Wobec tego sieć utrzymywała generalną prawidłowość łączącą wejścia i wyjścia sieci ujawniającą się w całym zbiorze uczącym. Ostatni przypadek występujący w zbiorze uczącym nie miał dominującego wpływu na profil wag sieci.

3.3.3.2. Dyskusja wyników neuronowego modelowania niedeterministycznego

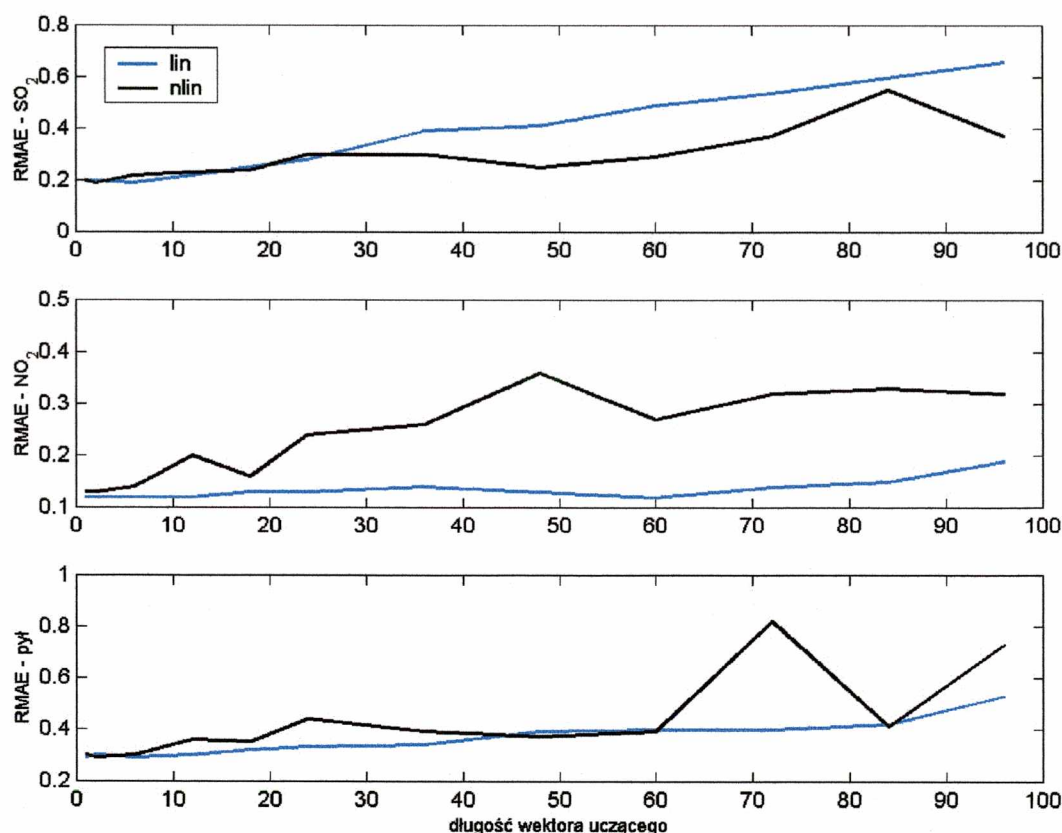
Wyniki modelowania wykazały, że walory prognostyczne wszystkich trzech typów sieci neuronowych są zbliżone dla krótkich wektorów uczących i krótkich horyzontów prognozy (tab.3.11). Dotyczy to zwłaszcza modelu nieliniowego i liniowego uczonych w trybie batch. Jakość wszystkich modeli spada w miarę wydłużania wektora uczącego. Pogorszenie jakości jest jednak wyraźnie najszybsze dla neuronowego filtra liniowego. W tabeli (tab.3.11) pokazano, że gdy zmiana długości wektora uczącego z 1 na 2, a następnie 6 elementów wywołuje bezkierunkowe wahania w parametrach modeli uczonych w trybie batch, to wskaźniki efektywności filtra liniowego ulegają pogorszeniu. Zaobserwowano, że w przypadku serii czasowej 30-minutowych stężeń SO₂ okno o długości 36 przypadków uniemożliwia zbudowanie modelu zależności za pomocą filtra neuronowego.

Tabela 3.11. Parametry niedeterministycznych modeli prognostycznych dla 30-minutowych stężeń SO₂ dla horyzontu prognozy 0.5 godziny przy długości okien uczących 1, 2 i 6 elementów szeregu czasowego stężeń

parametr	liniowy filtr adaptacyjny			sieć liniowa			sieć nieliniowa		
	1	2	6	1	2	6	1	2	6
RMSE	14.74	16.72	17.09	14.18	14.38	14.21	14.58	13.05	14.27
RMAE	0.24	0.24	0.24	0.20	0.20	0.19	0.20	0.19	0.22
współczynnik determinacji	0.74	0.72	0.70	0.74	0.73	0.75	0.74	0.78	0.75
średnia rzeczywista	42.59	42.59	42.59	42.59	42.59	42.59	42.59	42.59	42.59
średnia z modelu	46.02	46.29	45.93	43.64	43.61	44.27	43.31	42.50	43.52
wariancja rzeczywista	744	744	744	744	744	744	744	744	744
wariancja z modelu	761	834	949	717	726	744	777	760	797
błąd modelu	0.11	0.15	0.19	0.09	0.09	0.08	0.05	0.05	0.04
błąd danych	0.89	0.85	0.81	0.91	0.91	0.92	0.95	0.95	0.96

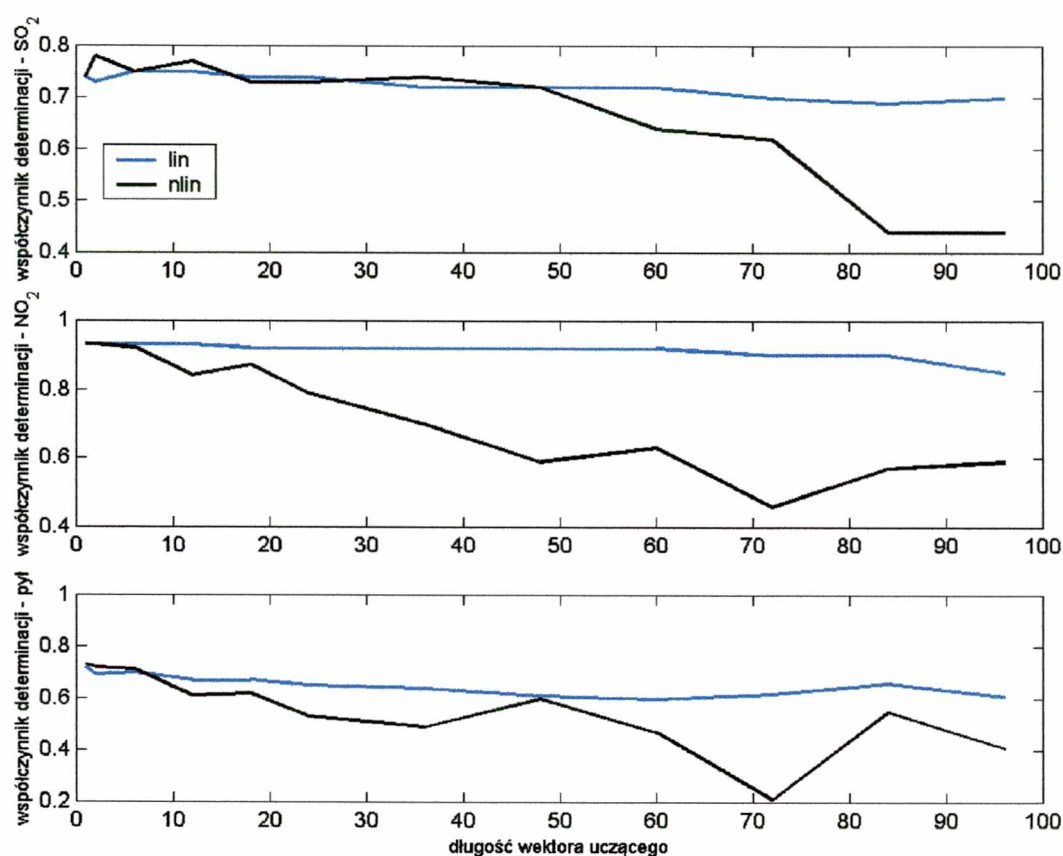
Modele liniowy i nieliniowy uczone w trybie batch dostarczają prognoz o zbliżonej jakości dla wektorów uczących o długości 1 do 6 elementów (tab.3.11). Wraz z dalszym wydłużaniem wektora uczącego jakość modelu i wyniki prognozowania systematycznie się pogarszają. Jak wynika z rysunków (rys.3.24 i rys.3.25) względny błąd prognozy wzrasta, a współczynnik determinacji modelu maleje w funkcji długości wektora uczącego w odniesieniu do wszystkich zanieczyszczeń.

Skoro dla zadanego horyzontu prognozy najlepsze rezultaty można uzyskać korzystając z jednoelementowego wektora uczącego, składającego się ze stężenia odległego o horyzont prognozy od stężenia prognozowanego oznacza to, że zawiera ono najistotniejszą informację dla wykonania prognozy. Wydłużanie wektora uczącego, poprzez rozbudowywanie go o informację dotyczącą przeszłości, nie poprawia rezultatów modelowania w stosunku do wektora jednoelementowego, a powyżej pewnej wartości ok. 6 elementów wpływa na ich pogorszenie (rys.3.24 i rys.3.25). Ta prawidłowość jest przewidywalna na podstawie wzorców autokorelacji ujawniających się w szeregach czasowych 30-minutowych stężeń zanieczyszczeń, które przedstawiono na rys.3.26. Przebieg wartości współczynnika autokorelacji w zależności od opóźnienia ma charakter funkcji malejącej. Odstępstwo pojawiające się w przypadku pyłu dotyczy dużego opóźnienia. Tak długi horyzont prognozy stężeń 30-minutowych jest nieosiągalny w przypadku modelu niedeterministycznego ze względu na dokładność prognozy. Malejący charakter zależności współczynnika od opóźnienia powoduje, że bez względu na horyzont prognozy, zawsze element wektora uczącego najbliższy w czasie wartości prognozowanej niesie o niej maksimum informacji w całym wektorze uczącym. Gdyby omawiana funkcja miała charakter okresowy, można się spodziewać istotnego wkładu w wartość prognozowaną opóźnień innych, niż jeden krok. [Perez,2000]

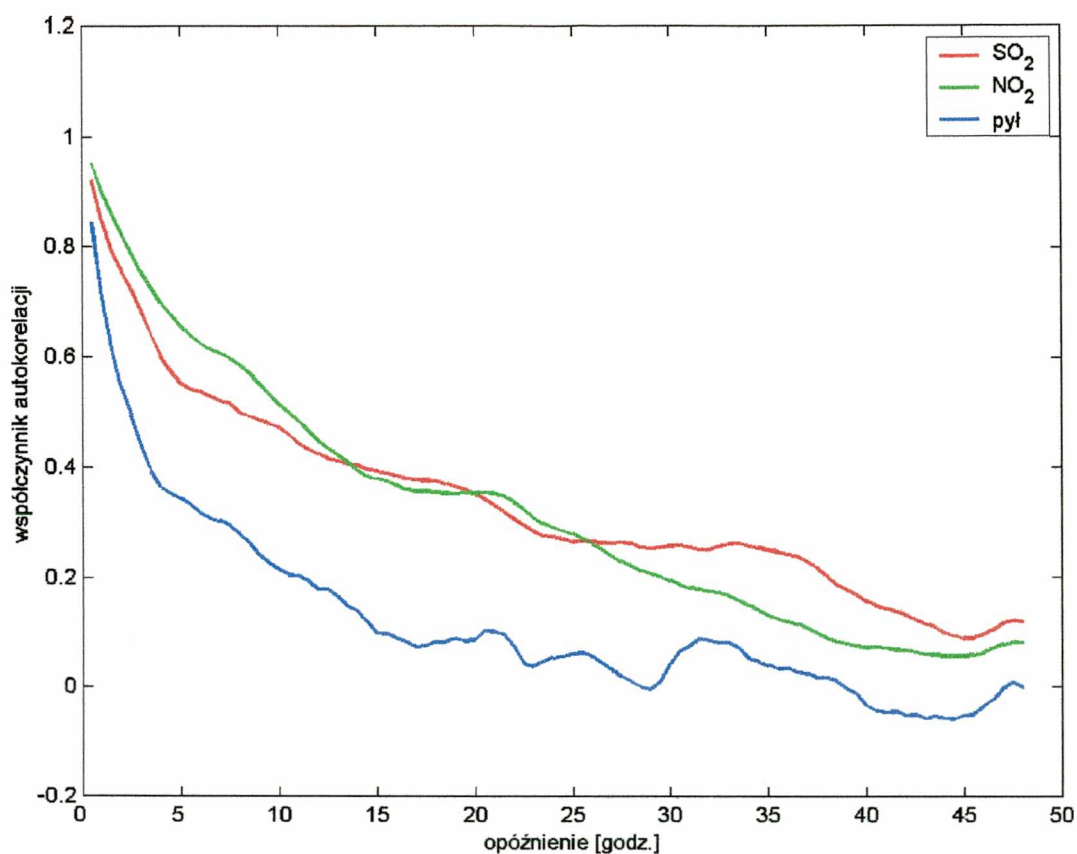


Rysunek 3.24. Zależność względnego błędu absolutnego prognozy od długości wektora uczącego dla horyzontu prognozy 0.5 godziny, dla zbioru testowego

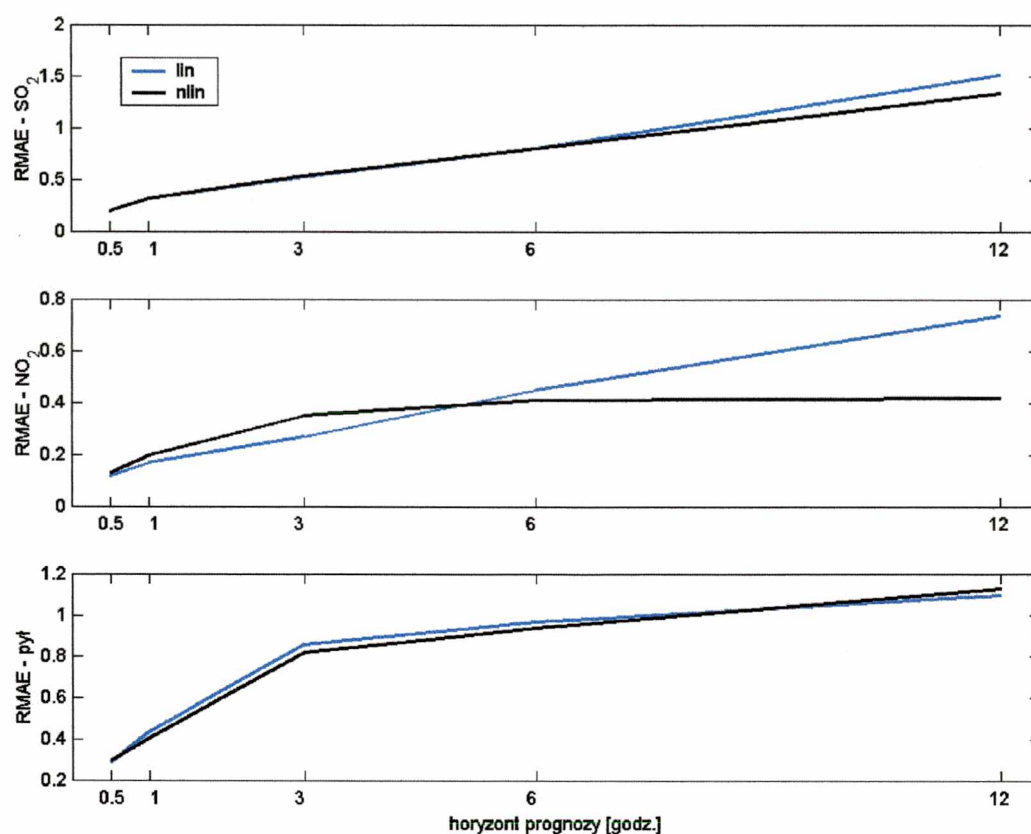
Współczynnik autokorelacji dostarcza informacji o zależności natury liniowej. Jak wynika z porównania parametrów modeli nieliniowych i liniowych, to właśnie liniowa forma zależności między stężeniami następującymi po sobie dominuje w szeregu czasowym 30-minutowych stężeń zanieczyszczeń. Model nieliniowy nie dostarcza wyników lepszych, niż model liniowy. Dla jednoelementowego wektora uczącego i różnych horyzontów prognozy parametry obu modeli są bardzo zbliżone, co wynika z rysunków (rys.3.24 i rys.3.25). W miarę wydłużania wektora podawanego na wejście sieci ujawnia się stopniowo przewaga modelu liniowego. Można na tej podstawie wnioskować, że ewentualne formy nieliniowości wykrywane przez sieć z neuronami nieliniowymi stanowią raczej zaburzenie, niż prawidłowość wnoszącą wkład w poprawę prognozy.



Rysunek 3.25. Zależność współczynnika determinacji od długości wektora uczącego dla horyzontu prognozy 0.5 godziny, dla zbioru testowego



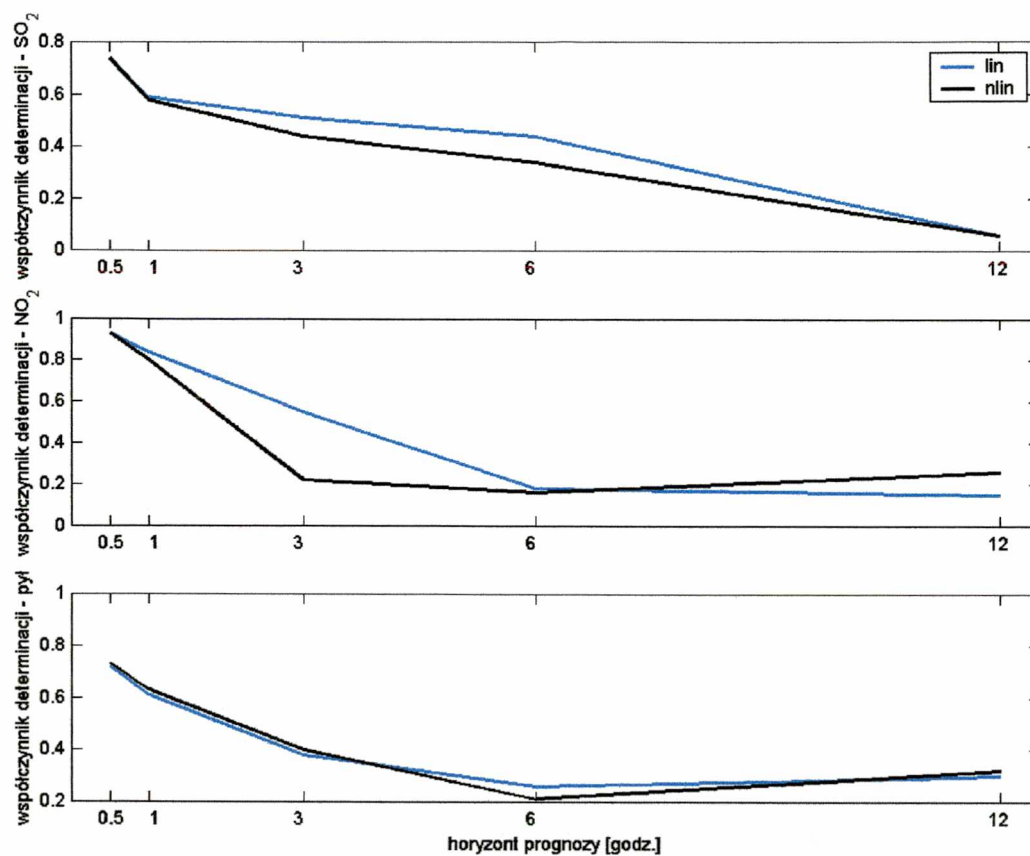
Rysunek 3.26. Wartość współczynnika autokorelacji szeregu czasowego stężeń SO_2 , NO_2 i pyłu w funkcji opóźnienia



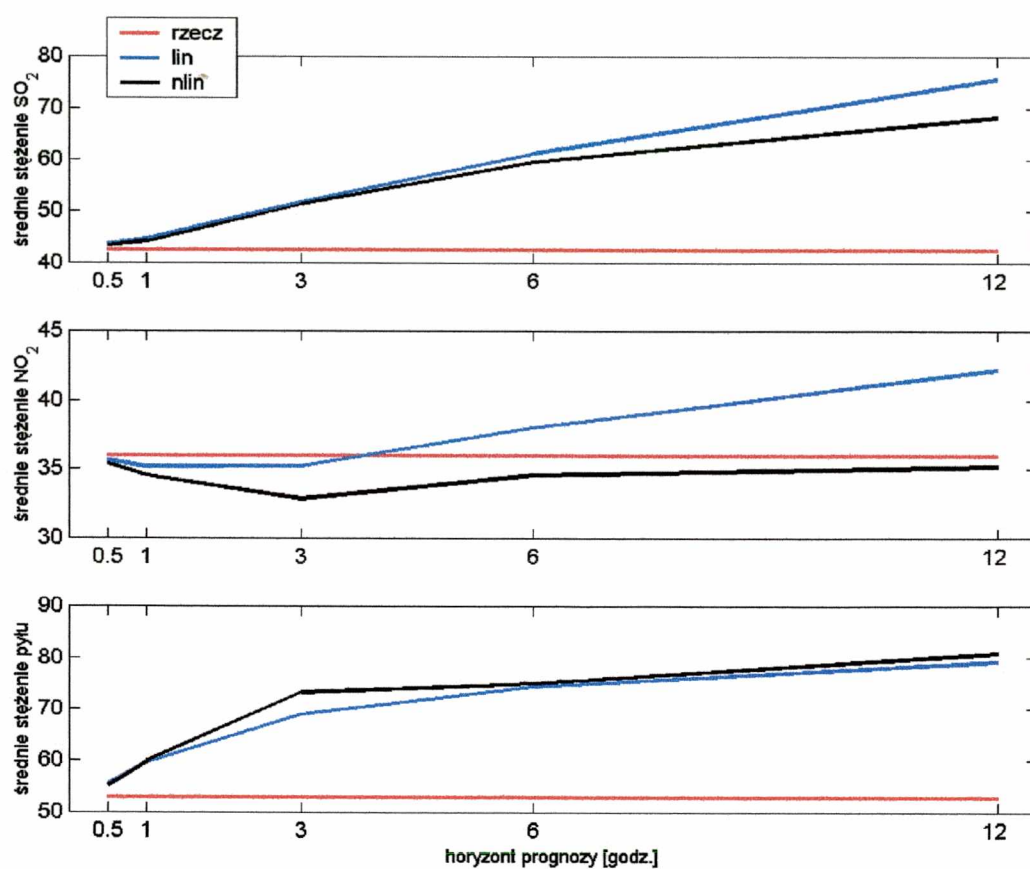
Rysunek 3.27. Zależność względnego błędu absolutnego prognozy od długości horyzontu prognozy, dla zbioru testowego

Jakość prognozy bardzo silnie zależy od jej horyzontu czasowego. Z rys.3.27 wynika, że błąd prognozy wzrasta z wydłużaniem horyzontu czasowego. W przypadku prognoz wykonywanych dla dwutlenku siarki oraz pyłu już horyzont prognozy na poziomie 3 godzin wiąże się z błędem względnym przekraczającym 50%, co dyskwalifikuje prognozę jako wartościową wypowiedź o przyszłości. Jeśli chodzi o prognozę stężeń dwutlenku azotu, akceptowalną jest prognoza z wyprzedzeniem 6 godzin. Zależność współczynników determinacji od horyzontu prognozy przedstawiona na rys.3.28 sugeruje jednak, że należałoby ograniczyć horyzont prognozy do jednej godziny w odniesieniu do wszystkich zanieczyszczeń. W tym horyzoncie zmienność stężeń wszystkich zanieczyszczeń jest wyjaśniana najpełniej, bo w 84% dla NO_2 , w 63% model dla pyłu i 58% dla SO_2 . Modelowanie serii czasowej 30-minutowych stężeń dwutlenku azotu w układzie niedeterministycznym dla krótkich horyzontów prognozy i krótkich wektorów uczących prowadzi do lepszych rezultatów w porównaniu z dwutlenkiem siarki i pyłem. Można się było tego spodziewać na podstawie analizy współczynników autokorelacji serii czasowych stężeń rozważanych zanieczyszczeń zamieszczonych na rys.3.26. O pogarszaniu się jakości modelu w miarę wydłużania horyzontu prognozy świadczą również inne wskaźniki jakości modelu. Jak pokazano na rys.3.29 w miarę wydłużania horyzontu prognozy średnia z odpowiedzi modelu na zbiór testowy oddala się od średniej z rzeczywistych wartości należących do tego zbioru. Jednocześnie następuje spadek wariancji stężeń prognozowanych, jak wynika z rys.3.30. Obie te prawidłowości oznaczają, że wraz z wydłużaniem horyzontu prognozy odpowiedzi modelu są coraz wyraźniej zawyżane w stosunku do rzeczywistych wartości stężeń prognozowanych. Zawyżanie jest coraz bardziej konsekwentne, bo towarzyszy mu spadek wariancji odpowiedzi modelu. Model oddala się od prawidłowości, którą powinien odzwierciedlać. Wykazuje to bardzo wyraźnie błąd modelu, który stopniowo przejmując coraz większą część całego błędu prognozy (rys.3.31).

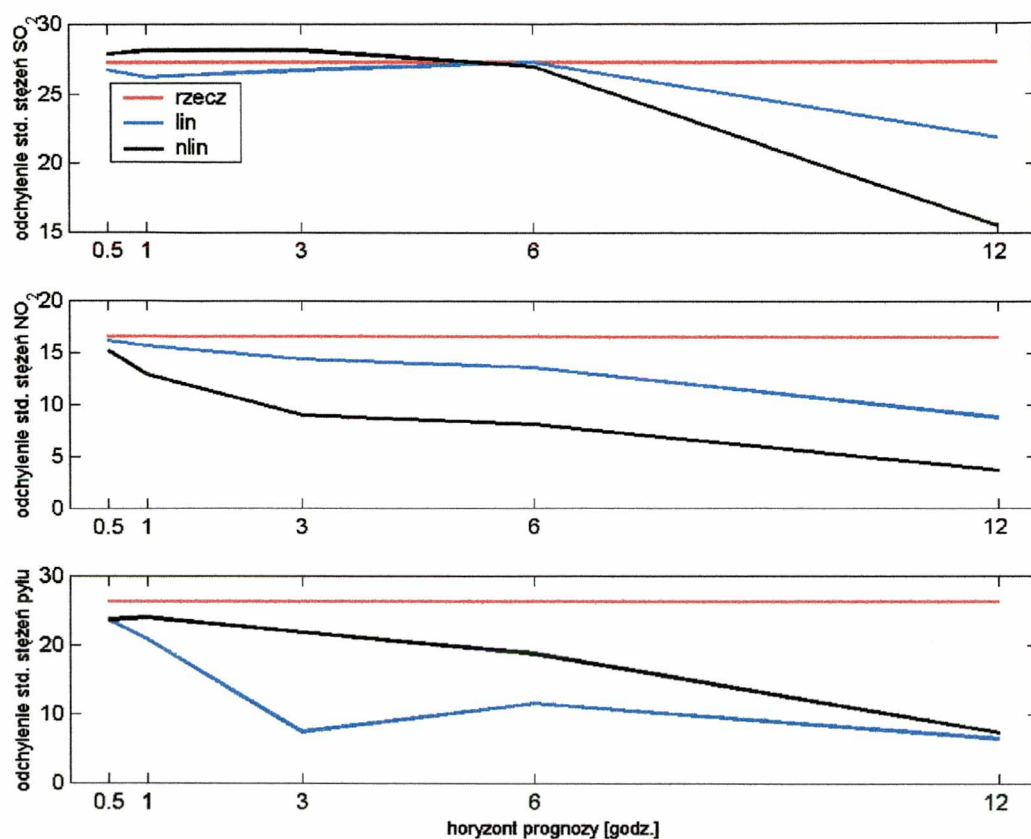
Stwierdzono zależność walorów prognostycznych modelu od charakteru zmienności stężeń występującej w zbiorze testującym. Ten sam model prognozuje szereg czasowy o łagodnej zmienności stężeń znacznie lepiej, niż szereg, do którego należy epizod stężeń. W związku z tym dla szeregów charakteryzujących się łagodną zmiennością stężeń możliwe są o wiele dłuższe horyzonty prognozy, niż w przypadku szeregów o dużej zmienności stężeń.



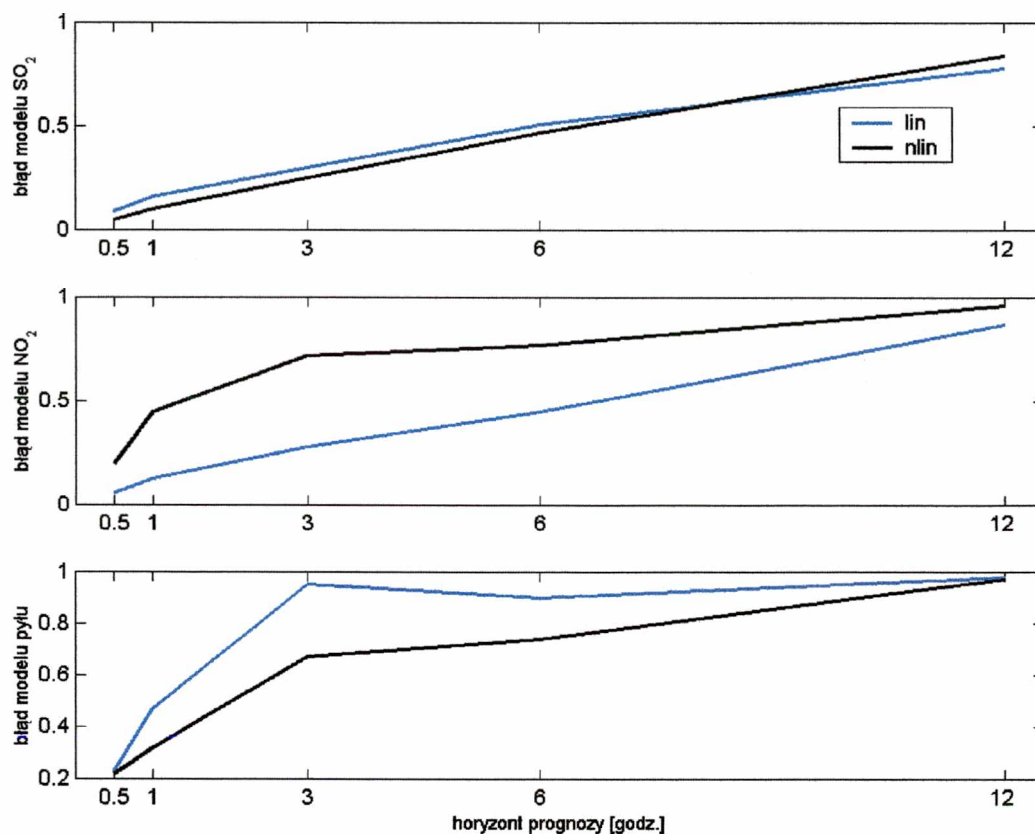
Rysunek 3.28. Zależność współczynnika determinacji modelu od długości horyzontu prognozy, dla zbioru testowego



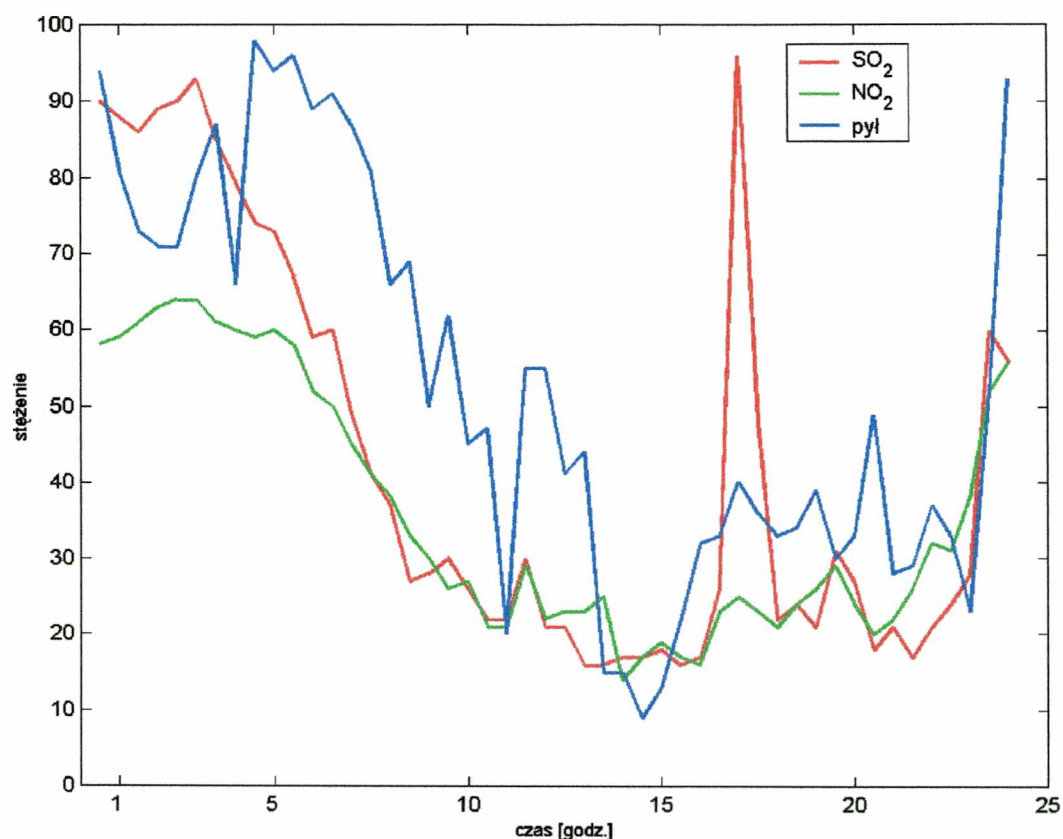
Rysunek 3.29. Średnia wartości stężeń rzeczywistych [$\mu\text{g}/\text{m}^3$], odpowiedzi modelu liniowego i nieliniowego w funkcji horyzontu prognozy dla zbioru testowego



Rysunek 3.30. Odchylenie standardowe stężeń rzeczywistych [$\mu\text{g}/\text{m}^3$], odpowiedzi modelu liniowego i nieliniowego w funkcji horyzontu prognozy, dla zbioru testowego



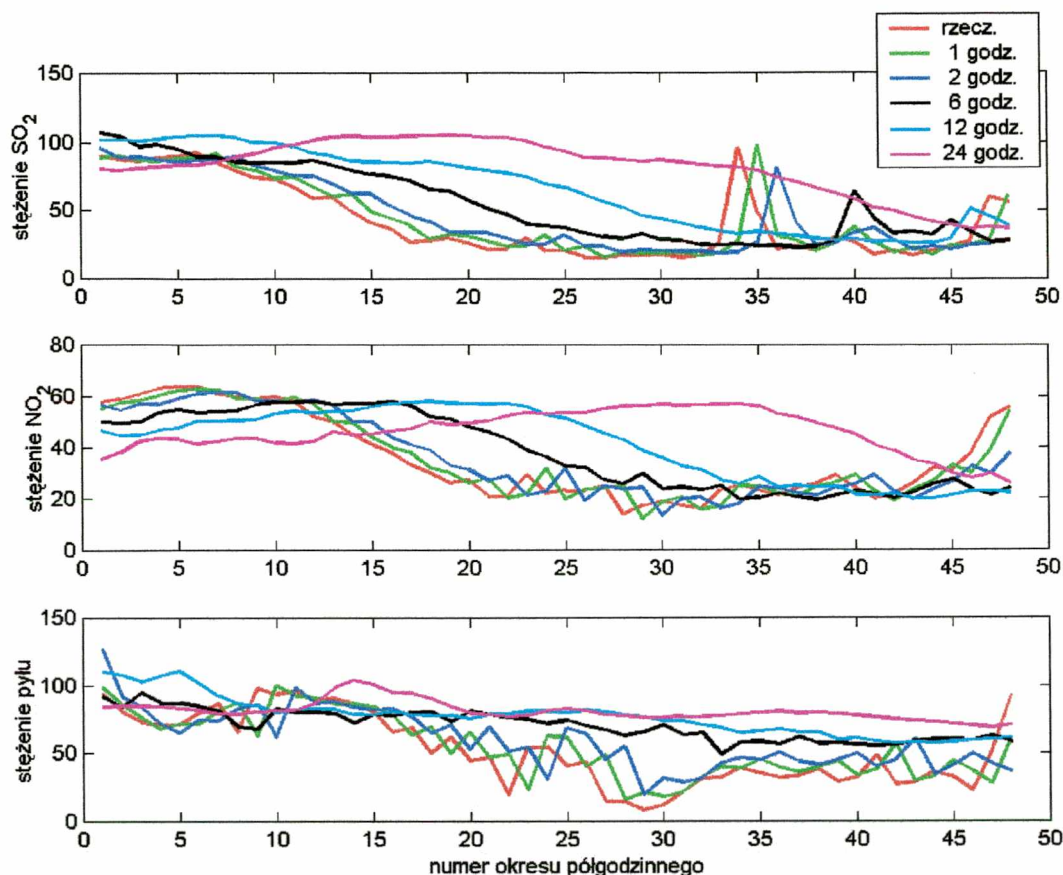
Rysunek 3.31. Część RMSE, za którą odpowiada model liniowy i nieliniowy, w funkcji horyzontu prognozy



Rysunek 3.32. Szeregi czasowe stężeń zanieczyszczeń stanowiące przedmiot testowania modeli prognostycznych [$\mu\text{g}/\text{m}^3$]

Prawidłowość tę zauważono w przypadku neuronowych modeli liniowych i nieliniowych. Na rys.3.32 pokazano dobowe przebiegi 30-minutowych stężeń SO_2 , NO_2 i pyłu, stanowiące przedmiot prognozy dla omawianych modeli. Szereg czasowy stężeń dwutlenku azotu odznacza się najmniejszą zmiennością, wyraźny epizod stężeń pojawia się w szeregu stężeń SO_2 , który poza tym jest porównywalny co do gładkości z przebiegiem NO_2 . Szereg stężeń pyłu cechują największe różnice pomiędzy sąsiednimi wartościami stężeń. Taka kolejność szeregów testowych pod względem charakteru zmienności stężeń odpowiada kolejności modeli prognostycznych pod względem jakości uzyskiwanych prognoz. Najmniejszy względny absolutny błąd prognozy stężeń NO_2 wynosi 12%, dla SO_2 przyjmuje wartość 20% a dla pyłu 29%. W zakresie współczynników determinacji, które wynoszą odpowiednio 93%, 74% i 72%, rozbieżność między modelem dla dwutlenku siarki i pyłu jest mniejsza podczas, gdy model zbudowany dla NO_2 jest wyraźnie najlepszy. Wszystkie wartości dotyczą prognozy wykonanej z horyzontem 0.5 godziny w oparciu o jednoelementowy zbiór uczący.

Jakość prognozy stężenia 30-minutowego nie zależy od długości zbioru uczącego w zakresie od 1 do 8 dób. Do wykonania dobrej prognozy z wykorzystaniem modelu niedeterministycznego wystarczy 1 doba danych za okres bezpośrednio poprzedzający prognozę.



Rysunek 3.33. Prognozowane przebiegi 30-minutowych stężeń zanieczyszczeń [$\mu\text{g}/\text{m}^3$]. Prognozy wykonano za pomocą nieliniowego filtra neuronowego dla różnych horyzontów czasowych. Wyniki uzyskano przy sześcioelementowym wektorze uczącym.

Wydaje się, że w krótkim horyzoncie czasowym działanie modelu prognostycznego polega na odtwarzaniu wartości stężenia z tej pozycji w wektorze wejściowym, która w toku procesu uczenia okazała się być najbliższa wartości podawanej na wyjście sieci. Połączenie neuronowe tego wejścia z wyjściem sieci charakteryzuje się największą wagą. W miarę wydłużania horyzontu prognozy podobieństwo wyjścia sieci do najbliższego jej w czasie elementu wektora wejściowego maleje. Sieć jest stopniowo pozbawiana wejścia podobnego do wyjścia, wobec tego musi się uczyć odwzorowywać wyjście na podstawie wejść coraz słabiej z nim skorelowanych. Pojawia się tendencja do uśredniania. Jak wynika z rys.3.33, najwyraźniej występuje ona w przypadku stężeń pyłu. Układ prognozowanych stężeń między 25 a 35 elementem zbioru testowego przy horyzontach czasowych 6, 12 i 24 ilustruje zanik odtwarzania stężeń z wektora wejściowego. Następuje spłaszczenie przebiegu. Dla opóźnień 6, 12 i 24 w przypadku stężeń pyłu wartości współczynnika autokorelacji są już na bardzo niskim poziomie ok. 0.25 (rys.3.26). Współczynnik autokorelacji dla serii czasowych stężeń SO₂ i NO₂ cechuje stały spadek jednak dla opóźnienia 24 jego wartość osiąga dopiero do 0.5. Dlatego prognoza 30-minutowych stężeń SO₂ i NO₂ w horyzoncie czasowym równym 12 godzin wiąże się nadal z odtwarzaniem stężenia ze zbioru uczącego najbliższego w sensie czasowym stężeniu prognozowanemu (rys.3.33).

3.4. Interpolacja przestrzenna stężeń imisyjnych

Najbardziej czytelnym sposobem obrazowania stanu zanieczyszczenia powietrza atmosferycznego nad wybranym terytorium jest przestrzenny rozkład stężeń zanieczyszczeń. Ma on wiele pozytywnych właściwości, których wspólnym mianownikiem są prostota i komunikatywność przekazu połączona z bardzo dużą pojemnością informacyjną. Dane stanowiące podstawę generowania takich rozkładów pochodzą z pomiarów prowadzonych równocześnie w co najmniej kilku punktach pomiarowych. Często ich źródłem są systemy monitoringu powietrza atmosferycznego.

Z badawczego punktu widzenia analiza rozkładów przestrzennych stężeń zanieczyszczeń na tle topografii terenu, warunków meteorologicznych i rozmieszczenia źródeł zanieczyszczeń otwiera pole dla analizy udziału tych elementów w kształtowaniu jakości powietrza atmosferycznego. Jednak warunkiem przeprowadzenia takich badań jest osiągnięcie poprawności generowanych rozkładów przestrzennych tzn. jak najlepsze przybliżenie sztucznym rozkładem rzeczywistego pola stężeń. Oznacza to konieczność wyboru optymalnej metody interpolacji.

W tym rozdziale zaproponowano metodę interpolacji z wykorzystaniem radialnej sieci neuronowej. Metodę zastosowano do interpolacji przestrzennej 30-minutowych stężeń zanieczyszczeń.

3.4.1. Analiza przygotowawcza

Jakość interpolacji jest uzależniona od postaci pola wielkości mierzonej, rozmieszczenia punktów pomiarowych i metody interpolacji. Jeśli wiedza o prawdopodobnej postaci pola wielkości mierzonej jest dostępna na etapie tworzenia systemu pomiarowego, pomaga ona właściwie rozmieścić punkty pomiarowe. Układ punktów pomiarowych zoptymalizowany ze względu na to kryterium dostarcza kompletnej informacji o przestrzennym rozkładzie stężeń. Jakość wykonywanej interpolacji praktycznie nie zależy wówczas od zastosowanej metody interpolacji. W zasadzie dowolna metoda umożliwia uzyskanie dobrych rezultatów.

Jeśli prawdopodobna postać pola wielkości mierzonej nie jest znana, teoria interpolacji proponuje rozkład punktów pomiarowych na regularnej siatce w celu zidentyfikowania tego pola. Gęstość siatki jest zwykle uwarunkowana względami ekonomicznymi. Niestety punkty pomiarowe większości systemów monitoringu powietrza atmosferycznego nie są rozmieszczane pod tym kątem, lecz najczęściej według tzw. kryterium reprezentatywności. Oznacza to, że każdy punkt ma reprezentować jakość powietrza w obszarze o określonych cechach np. zabudowie, zaludnieniu, natężeniu ruchu drogowego, szacie roślinnej. Służy to realizacji zadań sieci pomiarowych, którymi

są zazwyczaj kontrola dotrzymania norm jakości powietrza, ocena ekspozycji populacji ludzkich, ekosystemów i dóbr materialnych.

Jeżeli punkty pomiarowe nie są rozmieszczone optymalnie z punktu widzenia interpolacji przestrzennej uzyskiwanych wyników pomiaru, to główne zadanie interpolacji sprowadza się do roli elementu rekompensującego niedostatek danych. Dzieje się to kosztem dokładności odtworzenia brakującej informacji. W przypadkach skrajnych, dokładność interpolacji może w niewielkim stopniu zależeć od jakości zastosowanej metody interpolacji. Wówczas maleje pojemność informacyjna generowanych rozkładów i słabnie moc wniosków formułowanych na podstawie ich analizy.

Poszukiwanie metody interpolacji, która w pełni wykorzystuje informację zawartą w danych pomiarowych, można prowadzić zasadniczo na dwa sposoby. Pierwszy wychodzi od analizy jednorodności przestrzennej informacji zebranej z różnych punktów pomiarowych. Wynik takiej analizy kieruje w stronę metod czułych na lokalną zmienność stężeń lub też metod dobrze odzwierciedlających zmienność o charakterze regionalnym. Drugi sposób polega na przetestowaniu wyników zastosowania różnych metod interpolacji i wyborze metody dającej wyniki najlepsze. Podejście może się wydawać nieco instrumentalne, nie mniej jednak pozwala wybrać metodę interpolacji faktycznie najlepszą. W pracy skojarzono oba podejścia. Najpierw podjęto próbę oceny jednorodności przestrzennej pól stężeń zanieczyszczeń SO_2 , NO_2 i pyłu. Następnie przebadano metodę interpolacji wykorzystującą neuronowe sieci radialne w różnych układach parametrów, poszukując potwierdzenia lub zaprzeczenia wyników wstępnej analizy statystycznej.

Dla rozpoznania jednorodności przestrzennej stężeń średniorocznych lub sezonowych pomocne może się okazać kryterium reprezentatywności. Zgodnie z definicją: "Skala przestrzennej reprezentatywności jest to cecha stanowiska pomiarowego określająca wielkość obszaru, w którym stężenie zanieczyszczenia może być reprezentowane przez wartości pomierzone na stacji monitoringowej. Skala nawiązuje do średniej odległości (d_R) pomiędzy skrajnymi punktami obszaru, w którym stężenia zanieczyszczenia nie różnią się wyraźnie od wartości pomierzonych na stacji monitoringu. (Zakłada się, że różnice nie powinny przekraczać 20% stężenia na stanowisku pomiarowym w odniesieniu do wartości średniorocznych lub sezonowych). Reprezentatywność punktu pomiarowego uzależniona jest od gradientu średnich stężeń w jego otoczeniu." ... [Bibl. Monit. Środ.]

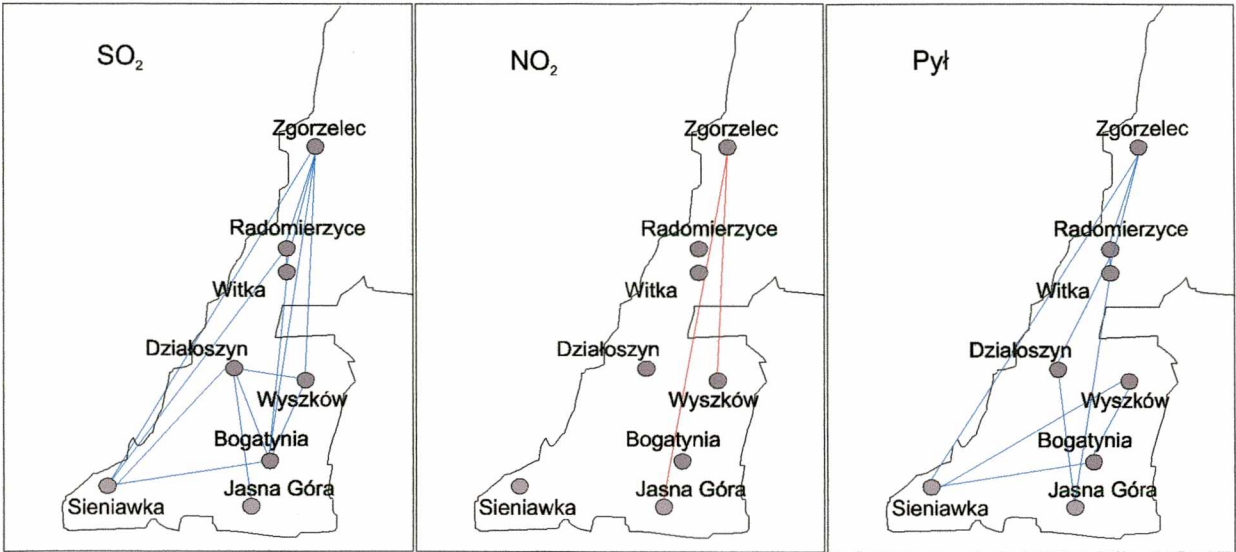
Jeżeli sąsiednie punkty pomiarowe znajdują się w granicach swoich obszarów reprezentatywności tzn. różnice między rejestrowanymi tam stężeniami są bardzo niewielkie, jako

metodę interpolacji stężeń średniorocznych można zaproponować metodę średnich ważonych z zastosowaniem jednakowych wag dla wszystkich punktów. Wartość wagi wynosi $1/N$ gdzie N jest liczbą punktów pomiarowych. W ten sposób w każdym punkcie o współrzędnych (x,y) , gdzie przebiega interpolacja, wpływ wszystkich punktów pomiarowych o współrzędnych (x_i,y_i) jest równouprawniony, a wartość stężenia w tym punkcie opisuje równanie:

$$s(x,y) = \sum_{i=1}^N \left(\frac{1}{N} \cdot s(x_i,y_i) \right) \quad (3.8)$$

Jak łatwo wywnioskować z równania (3.8) proponowana metoda interpolacji prowadzi do uzyskania płaskiej powierzchni interpolującej, która każdemu punktowi o współrzędnych (x,y) przyporządkowuje tę samą wartość s , będącą uśrednieniem wartości dla wszystkich punktów pomiarowych. Należy jednak zauważyć, że taki układ stacji jest zaprzeczeniem optymalności sieci monitoringowej. Poszczególne stacje dublują informację pomiarową i nie wnoszą nic nowego do obrazu jakości powietrza nad zadany terenem.

Powyższe kryterium zastosowane dla stężeń średniorocznych SO_2 , NO_2 i pyłu w Worku Żytawskim wykazało, że istnieją pary, jak również większe grupy stacji, dla których stężenia zanieczyszczeń nie różnią się od siebie bardziej niż o 20%. Takie grupy stacji zilustrowano na rysunku (rys.3.34) łącząc przynależne do nich stacje niebieskimi odcinkami. W przypadku stężeń SO_2 i pyłu trudno jednak wyznaczyć obszary reprezentatywności większe niż obejmujący dwie stacje: Radomierzyce i Witka. Obszary reprezentatywności punktów o zbliżonych wartościach stężeń są rozdzielone obszarami reprezentatywności innych punktów pomiarowych. Inaczej przedstawia się sytuacja w przypadku NO_2 . Czerwone linie łączą stacje, dla których średnioroczne stężenia różnią się bardziej niż o 20%. Jak widać, całą pozostałą część Worka Turoszowskiego cechują bardzo podobne wartości stężeń i zlokalizowane tam stacje leżą wzajemnie w swoich obszarach reprezentatywności. W przypadku tego zanieczyszczenia można swobodnie zredukować liczbę punktów pomiarowych do Zgorzelca i dowolnej innej stacji położonej na terenie Worka Żytawskiego. Najodpowiedniejszą zaś metodą interpolacji średniorocznych stężeń NO_2 wydaje się metoda średnich ważonych z zastosowaniem jednakowych wag dla wszystkich punktów. Dane dotyczące reprezentatywności punktów pomiarowych stężeń SO_2 i pyłu nie są wystarczające do zaproponowania tej samej metody interpolacji w odniesieniu do stężeń wymienionych zanieczyszczeń.



Rysunek 3.34. Stacje, dla których różnice stężeń średniorocznych nie przekraczają 20% połączone niebieskimi liniami w przypadku stężeń SO₂ i pyłu. Stacje, dla których różnice stężeń średniorocznych przekraczają 20% połączone czerwonymi liniami w przypadku stężeń NO₂.

Obok kryterium reprezentatywności, które dotyczy wyłącznie stężeń długookresowych, zaproponowano dodatkowe miary podobieństwa informacji zawartej w danych pomiarowych, użyteczne również w odniesieniu do zadania interpolacji stężeń krótkookresowych.

Wskaźnik pierwszy jest medianą wszystkich ilorazów maksymalnej i minimalnej spośród wartości zmiennej zarejestrowanych równocześnie w sieci pomiarowej w ciągu 2 lat. Wskaźnik pokazuje ilukrotnie maksymalna wartość np. stężenia SO₂ w danym punkcie pomiarowym przewyższa przeciętnie najmniejszą wartość stężenia SO₂ zmierzoną równocześnie na innej stacji.

Wskaźnik drugi jest medianą wszystkich różnic maksymalnej i minimalnej wartości zmiennych rejestrowanych w Worku Żytawskim na przestrzeni lat 1997 i 98. Wskaźnik pokazuje, o ile jednostek maksymalna wartość np. stężenia SO₂ zanotowana w jakimś punkcie pomiarowym jest przeciętnie większa od najmniejszej wartości stężenia SO₂ zmierzonej w tym samym czasie, lecz na innej stacji.

Wartości obu wskaźników wyznaczono dla wszystkich zmiennych poddanych pomiarom dla okresów uśredniania 30 minut, doba i rok. Wyniki obliczeń zamieszczono w tabelach (tab.3.12 do tab.3.14).

Tabela 3.12. Zestawienie wskaźników podobieństwa informacji dla wartości 30-minutowych parametrów mierzonych przez ZSM-ET

Parametr mierzony	Wskaźnik 1	Wskaźnik 2
Ciśnienie atmosferyczne	1.0	27.2 [Pa]
Temperatura powietrza	1.7	1.8 [K]
Wilgotność powietrza	1.4	20.0 [%]
Prędkość wiatru	6.9	4.6 [m/s]
Kierunek wiatru	15.0	188.0 [deg]
Stężenie pyłu	7.7	61.8 [µg/m ³]
Stężenie SO ₂	10.8	67.0 [µg/m ³]
Stężenie NO ₂	6.1	20.2 [µg/m ³]

Tabela 3.13. Zestawienie wskaźników podobieństwa informacji dla wartości średniodobowych stężeń zanieczyszczeń mierzonych przez ZSM-ET

Parametr mierzony	Wskaźnik 1	Wskaźnik 2
Stężenie pyłu	3.1	36.9 [µg/m ³]
Stężenie SO ₂	4.7	33.7 [µg/m ³]
Stężenie NO ₂	3.4	12.4 [µg/m ³]

Tabela 3.14. Zestawienie wskaźników podobieństwa informacji dla wartości średniorocznych stężeń zanieczyszczeń i parametrów meteorologicznych mierzonych przez ZSM-ET

Parametr mierzony	Wskaźnik 1	Wskaźnik 2
Stężenie pyłu	1.8 (2)	23.6 (26.4min) [µg/m ³]
Stężenie SO ₂	1.8 (1.7)	18 (17.2min) [µg/m ³]
Stężenie NO ₂	1.6 (1.6)	5.8 (6.1min) [µg/m ³]

Na podstawie tab.3.12 można stwierdzić dużą jednorodność w obrębie chwilowych pól ciśnienia atmosferycznego, wilgotności i temperatury powietrza na terenie Worka Żytawskiego. Maksymalne różnice rejestrowanych tam 30-minutowych wartości ciśnienia kształtują się na poziomie 27Pa, różnice temperatury na poziomie 1.8°, a wilgotności w granicach 20%. Zmienność przestrzenna tych parametrów w obrębie Worka Żytawskiego jest więc niewielka. Wydłużenie okresu uśredniania powoduje w naturalny sposób jej dalsze zmniejszenie. Spośród parametrów meteorologicznych wziętych pod uwagę, jedynie prędkość i kierunek wiatru wykazują niejednorodność podobną jak stężenia zanieczyszczeń. Jedynie one, spośród wszystkich parametrów meteorologicznych, mogą być rozważane jako czynniki warunkujące zmienność przestrzenną chwilowych stężeń zanieczyszczeń.

Na poziomie stężeń chwilowych występują duże różnice między stacjami, gdzie rejestrowano wartości maksymalne a stacjami, gdzie pomiar wykazał minimalną wartość stężenia (tab.3.12). Dotyczy to wszystkich rozważanych zanieczyszczeń. Największe krotności stężeń minimalnych są związane z SO₂ (11) w stosunku do 8 dla pyłu i 6 dla NO₂. Według *wskaźnika1* przestrzenne zróżnicowanie stężeń NO₂ i pyłu są do siebie zbliżone bez względu na okres uśredniania, choć w miarę wydłużania tego okresu różnice między wszystkimi zanieczyszczeniami ulegają zatarciu. Mała wartość *wskaźnika2* dla NO₂ jest skutkiem ogólnie niższych wartości stężeń tego zanieczyszczenia w stosunku do SO₂ i pyłu. Z analizy tabel (tab.3.12 do tab.3.14) wynika, że wraz z wydłużaniem okresu uśredniania stężeń wzrasta jednorodność pól stężeń wyliczonych na podstawie pomiarów. Ze wzrostem okresu uśredniania pola stężeń nad obszarem Worka Turowskiego stają się coraz bardziej gładkie wskutek przesunięcia od reprezentacji wpływów lokalnych do odzwierciedlenia oddziaływań o większej skali.

Idąc tym tokiem myślenia warto zastanowić się, jaki udział w kształtowaniu stężeń na terenie Worka Turoszowskiego mają czynniki ściśle lokalne, ograniczające swój wpływ do jednej, może dwóch stacji, a jakiego rzędu jest wpływ czynników o większej skali oddziaływania widoczny w całym rozważanym terenie. Szukając odpowiedzi na tak postawione pytanie przeprowadzono analizę czynnikową kolejno na ciągach czasowych stężeń SO_2 , NO_2 i pyłu uwzględniając każdorazowo wszystkie stacje monitoringowe. Wyniki przedstawiono w tabelach (tab.3.15 do tab.3.17).

Tabela 3.15. Wyniki analizy metodą czynników głównych dla 30-minutowych stężeń SO_2 rejestrowanych w ośmiu punktach pomiarowych

Punkt pomiarowy	Zasób zmienności wspólnej		Ładunki czynnikowe	
	czynnik1	czynnik2	czynnik1	czynnik2
Bogatynia	0.57	0.01	0.76	0.10
Działoszyn	0.44	0.00	0.66	-0.05
Jasna Góra	0.59	0.00	0.77	0.01
Radomierzyce	0.56	0.01	0.75	-0.11
Sieniawka	0.65	0.00	0.81	0.01
Witka	0.01	0.96	-0.11	0.98
Wyszków	0.51	0.02	0.71	0.14
Zgorzelec	0.45	0.00	0.67	0.06
% wariancji całkowitej tłumaczony przez czynnik	0.47	0.13		

Tabela 3.16. Wyniki analizy metodą czynników głównych dla 30-minutowych stężeń NO_2 rejestrowanych w ośmiu punktach pomiarowych

Punkt pomiarowy	Zasób zmienności wspólnej		Ładunki czynnikowe	
	czynnik1	czynnik2	czynnik1	czynnik2
Bogatynia	0.35	0.05	0.59	-0.22
Działoszyn	0.51	0.00	0.72	0.05
Jasna Góra	0.00	0.75	0.04	-0.87
Radomierzyce	0.15	0.10	0.39	-0.32
Sieniawka	0.47	0.09	0.68	0.29
Witka	0.44	0.08	0.66	0.28
Wyszków	0.22	0.04	0.46	-0.21
Zgorzelec	0.35	0.01	0.59	-0.07
% wariancji całkowitej tłumaczony przez czynnik	0.31	0.14		

Tabela 3.17. Wyniki analizy metodą czynników głównych dla 30-minutowych stężeń pyłu rejestrowanych w ośmiu punktach pomiarowych

Punkt pomiarowy	Zasób zmienności wspólnej	Ładunki czynnikowe
	czynnik1	czynnik1
Bogatynia	0.18	0.42
Działoszyn	0.31	0.56
Jasna Góra	0.38	0.61
Radomierzyce	0.48	0.70
Sieniawka	0.42	0.65
Witka	0.51	0.72
Wyszków	0.37	0.61
Zgorzelec	0.44	0.66
% wariancji całkowitej tłumaczony przez czynnik	0.39	

Z rezultatów PCA przedstawionych w tabelach wynika, że wspólne czynniki wpływające na zmienność stężeń zanieczyszczeń w punktach pomiarowych ujawniają się zarówno dla SO_2 , NO_2 , jak i pyłu. Jeden wyróżniający się czynnik zaznaczający swój wpływ na wszystkich stacjach występuje w przypadku pyłu. Tłumaczy on średnio 39% wspólnej zmienności stężeń pyłu na rozważanym obszarze. W przypadku zanieczyszczeń gazowych ujawniają się po dwa czynniki. Pierwsze odpowiadają za wspólną zmienność stężeń w siedmiu punktach pomiarowych na poziomie 47% w odniesieniu do SO_2 i 31% w odniesieniu do NO_2 . Drugie czynniki główne odpowiadają za ok. 13% wspólnej zmienności stężeń SO_2 i 14% zmienności stężeń NO_2 sugerując występowanie osobnej głównej przyczyny zmienności stężeń SO_2 w Witce i NO_2 w Jasnej Górze. Wynika stąd, że czynniki odpowiedzialne za wspólną zmienność stężeń nie mają dominującego wpływu na kształtowanie przestrzennych rozkładów stężeń rozważanych zanieczyszczeń. To spostrzeżenie uzasadnia tezę o indywidualnych przyczynach zmienności stężeń zanieczyszczeń na poszczególnych stacjach pomiarowych.

Tabela 3.18. Współczynniki korelacji między szeregami czasowymi stężeń 30-minutowych zanieczyszczeń zarejestrowanych w różnych punktach pomiarowych

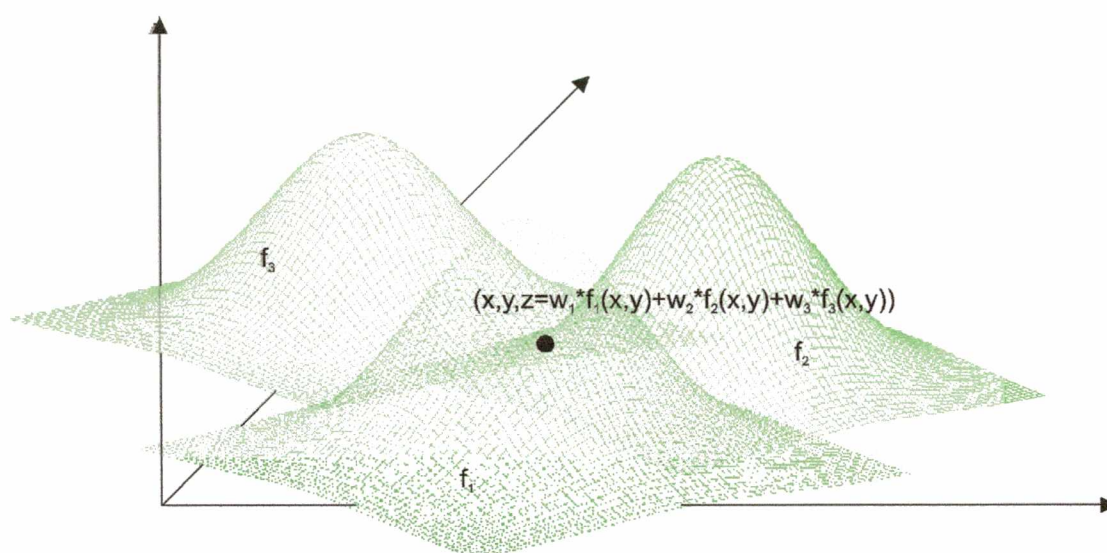
SO_2	Bogatynia	Działoszyn	Jasna Góra	Sieniawka	Radomierzycze	Witka	Wyszków	Zgorzelec
Bogatynia	1.00	0.62	0.75	0.37	0.66	0.39	0.69	0.50
Działoszyn	0.62	1.00	0.63	0.40	0.60	0.43	0.58	0.49
Jasna Góra	0.75	0.63	1.00	0.39	0.67	0.41	0.68	0.50
Radomierzycze	0.37	0.40	0.39	1.00	0.36	0.42	0.37	0.45
Sieniawka	0.66	0.60	0.67	0.36	1.00	0.38	0.66	0.50
Witka	0.39	0.43	0.41	0.42	0.38	1.00	0.37	0.33
Wyszków	0.69	0.58	0.68	0.37	0.66	0.37	1.00	0.47
Zgorzelec	0.50	0.49	0.50	0.45	0.50	0.33	0.47	1.00
NO_2								
Bogatynia	1.00	0.70	0.71	0.61	0.71	0.63	0.73	0.67
Działoszyn	0.70	1.00	0.63	0.58	0.66	0.65	0.66	0.61
Jasna Góra	0.71	0.63	1.00	0.62	0.58	0.63	0.71	0.64
Radomierzycze	0.61	0.58	0.62	1.00	0.56	0.65	0.57	0.68
Sieniawka	0.71	0.66	0.58	0.56	1.00	0.58	0.78	0.65
Witka	0.63	0.65	0.63	0.65	0.58	1.00	0.61	0.62
Wyszków	0.73	0.66	0.71	0.57	0.78	0.61	1.00	0.66
Zgorzelec	0.67	0.61	0.64	0.68	0.65	0.62	0.66	1.00
pył								
Bogatynia	1.00	0.51	0.57	0.50	0.58	0.55	0.48	0.47
Działoszyn	0.51	1.00	0.55	0.54	0.50	0.60	0.44	0.57
Jasna Góra	0.57	0.55	1.00	0.56	0.58	0.61	0.52	0.55
Radomierzycze	0.50	0.54	0.56	1.00	0.53	0.74	0.52	0.67
Sieniawka	0.58	0.50	0.58	0.53	1.00	0.58	0.57	0.53
Witka	0.55	0.60	0.61	0.74	0.58	1.00	0.57	0.73
Wyszków	0.48	0.44	0.52	0.52	0.57	0.57	1.00	0.52
Zgorzelec	0.47	0.57	0.55	0.67	0.53	0.73	0.52	1.00

Z punktu widzenia interpolacji ważna jest ocena stopnia współzależności między stężeniami rejestrowanymi w poszczególnych punktach pomiarowych. Współczynniki korelacji między stacjami wyliczone dla stężeń SO_2 , NO_2 i pyłu zamieszczono w tab.3.18. Wynika z nich, że najslabiej skorelowane przestrzennie są stężenia SO_2 , nieco lepiej stężenia pyłu, a zdecydowanie najbardziej stężenia NO_2 .

3.4.2. Propozycja rozwiązania

Opracowano metodę interpolacji przestrzennej z wykorzystaniem radialnych sieci neuronowych. Zasadniczą zaletą metody polega na jej elastyczności, gdyż zależnie od wartości parametrów umożliwia reprezentację zmienności przestrzennej o charakterze lokalnym, regionalnym lub pośrednim. Dzięki temu po odpowiednim skonfigurowaniu metoda nadaje się do rozwiązywania bardzo różnych problemów interpolacji.

Metoda interpolacji przestrzennej wykorzystująca radialną sieć neuronową jest rozwiązaniem nowym. W literaturze przedmiotu nie spotkano się z propozycją takiego zastosowania sieci neuronowej. Funkcjonowanie i możliwości metody przedstawiono na przykładzie problemu interpolacji przestrzennej 30-minutowych stężeń zanieczyszczeń na terenie objętym pomiarami sieci monitoringu powietrza atmosferycznego.



Rysunek 3.35. Idea metody interpolacji z wykorzystaniem radialnej sieci neuronowej

Na tle innych metod interpolacji radialną sieć neuronową można traktować jako hybrydę metod funkcyjnych i metod z grupy średniej ważonej. Wykorzystanie metody interpolacji bazującej na neuronowych sieciach radialnych wymaga przyjęcia założenia, że w otoczeniu punktów pomiarowych rozkład stężeń ma postać funkcji transferu przypisanych neuronom radialnym. W rozważanym przypadku jest to powierzchnia dzwonowa Gaussa rozpięta na dwóch wymiarach przestrzennych x i y . Wartość funkcji jest stężeniem zanieczyszczenia. Każdemu punktowi pomiarowemu odpowiada jeden neuron i zmienność stężeń wokół punktu opisuje funkcja transferu tego neuronu. Maksimum poszczególnych funkcji wypada w punktach pomiarowych.

Wartość funkcji transferu maleje wraz z oddalaniem się od punktu pomiarowego oddając spadek wpływu punktu na jego otoczenie wraz ze wzrostem odległości od punktu. Stężenia w przestrzeni między punktami pomiarowymi są wyliczane w wyniku składania wartości funkcji Gaussa przyporządkowanych punktom pomiarowym (rys.3.35). Dzieje się to proporcjonalnie do wag nadanych wyjściom z neuronów warstwy radialnej w neuronie liniowym, który pełni rolę sumatora i stanowi warstwę wyjściową sieci. Dzięki niemu przedział zmienności funkcji Gaussa $(0,1>$ może być przeskalowany na rzeczywisty zakres zmienności stężeń. Składanie powierzchni Gaussa jest odpowiednikiem sumowania otoczeń punktów pomiarowych w metodach z grupy średniej ważonej, lecz te wagi nie zmieniają się wraz z odległością między punktami.

Lokalny lub regionalny charakter uwarunkowań rozkładu stężeń reprezentuje rozpiętość powierzchni dzwonowej Gaussa. Krzywe o małej podstawie odzwierciedlają wpływ czynników o charakterze lokalnym. Im większy promień tym bardziej regionalny charakter czynnika odpowiedzialnego za rozkład stężeń zanieczyszczeń. Zadaniem badacza jest dobór powierzchni Gaussa o rozpiętości odpowiedniej dla postawionego problemu interpolacji. Za parametr charakteryzujący wielkość krzywej Gaussa przyjęto odległość między punktem pomiarowym a punktem, gdzie wykonywana jest interpolacja, dla której wartość funkcji Gaussa wynosi 0.5. Jest to tzw. spread. Im większy spread, tym bardziej rozległa przestrzennie powierzchnia dzwonowa. Stwierdzenie na podstawie wyników interpolacji, która wartość spread jest optymalna, umożliwia zweryfikowanie wstępnych przesłanek na temat charakteru przyczyn zmienności przestrzennej stężeń. Takich przesłanek dostarcza wstępna analiza statystyczna danych pomiarowych.

$$f = \exp\left(-\frac{((w-p) \cdot b)^2}{2}\right) \quad (3.9)$$

Gdzie:

w - wektor wag neuronu radialnego

p - wektor wejściowy

b - parametr skalujący

$$spread = \frac{0.8326}{b} \quad (3.10)$$

W celu znalezienia wartości spread optymalnej dla interpolacji przestrzennej stężeń zanieczyszczeń zaproponowano wykonanie interpolacji dla wybranych wartości spread. Ta wartość spread, która pozwala na uzyskanie najlepszych wyników interpolacji jest wartością optymalną. Wartości pochodziły z przedziału od minimalnej odległości między punktami pomiarowymi ($omin$) do odległości maksymalnej ($omax$). Przyjęte wartości spread podano w tab.3.19.

Tabela 3.19. Wartości spread, dla których przeanalizowano jakość interpolacji z wykorzystaniem sieci radialnych

Lp	Wartości spread
1	omin
2	$omin+0.15(omax-omin)$
3	$omin+0.25(omax-omin)$
4	$omin+0.50(omax-omin)$
5	$omin+0.75(omax-omin)$

Do oceny jakości metody interpolacji zastosowano procedurę cross-validation. Polega ona na estymacji wartości mierzonej w każdym spośród n punktów pomiarowych po kolei na podstawie pomiarów z pozostałych $n-1$ punktów. Dla potrzeb metody zakłada się chwilowo nieznaną wartość wielkości mierzonej w jednym z punktów i wylicza się ją wykorzystując dane pochodzące z pozostałych punktów pomiarowych. Procedura jest powtarzana dla każdego punktu pomiarowego. Wartości wyliczone są dalej porównywane z rzeczywistymi wartościami wielkości mierzonej.

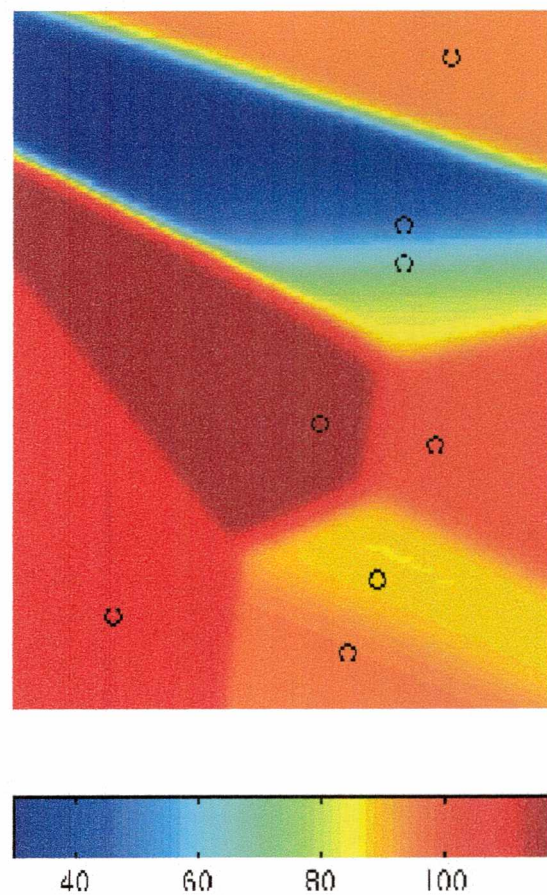
W związku z tym testowanie metody interpolacji przeprowadza się na zbiorach wyników pomiarów, które zawierają komplet pomiarów dla wszystkich punktów pomiarowych. Ta spośród metod interpolacji, dla której porównanie wyników uzyskanych z obliczeń i wartości rzeczywistych wypadło najlepiej jest uznawana za optymalną.

3.4.3. Dyskusja wyników neuronowej interpolacji stężeń

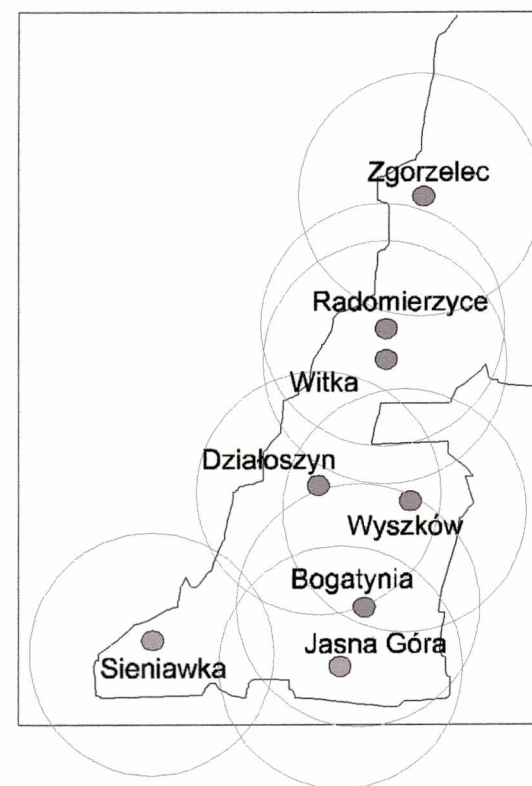
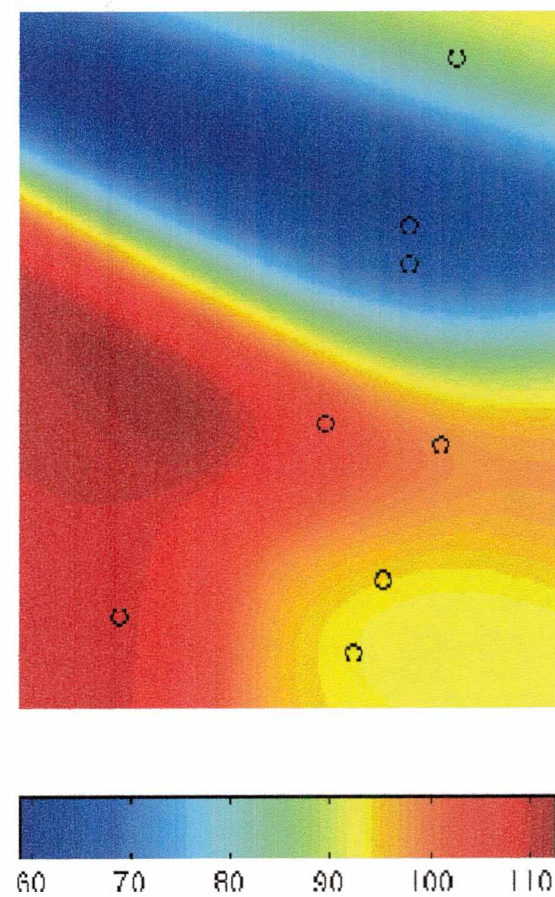
Wzrost spread, który oznacza powiększanie rozpiętości powierzchni Gaussa rozmieszczonych w punktach pomiarowych powoduje stopniowe spłaszczanie generowanego rozkładu przestrzennego stężeń. Rysunek 3.36 przedstawia pole stężeń charakteryzujące się dużym urozmaiceniem i dokładnym odtwarzaniem stężeń w punktach pomiarowych. Kolejne rysunki (rys.3.37 do rys.3.41) przedstawiają pola coraz bardziej jednorodne, zbliżające się do płaszczyzny reprezentującej poziom stężenia średniego dla obszaru jak na rys.3.41. Informuje o tym również zakres stężeń podany na skali kolorów pod każdym polem stężeń. Rozkłady przestrzenne stężeń skojarzono z rysunkami przedstawiającymi obszary wokół poszczególnych punktów pomiarowych o promieniach równych odpowiednim wartościom spread. Pokazano w ten sposób wpływ liczby punktów, w których otoczeniu znajduje się punkt, gdzie wykonywana jest interpolacja, na postać generowanego rozkładu przestrzennego stężeń.

Wyliczono miary jakości interpolacji 30-minutowych stężeń zanieczyszczeń z wykorzystaniem radialnych sieci neuronowych w postaci średniego błędu kwadratowego, średniego błędu względnego i współczynnika determinacji. Wskaźniki poprawności metody interpolacji zostały wyznaczone zgodnie z procedurą cross-validation dla wszystkich punktów pomiarowych przy kolejno zadawanych, według tab.3.19, wartościach spread. Analizy zostały przeprowadzone na pełnym zestawie danych z lat 1997 i 1998 jak również z podziałem na okres grzewczy i letni. Na rys.3.42 do rys.3.50 przedstawiono rezultaty dotyczące interpolacji stężeń SO_2 , NO_2 i pyłu rejestrowanych w okresie grzewczym. Ten okres czasu ujawnia specyficzną cechę przestrzennego rozkładu stężeń SO_2 . Odpowiednie wykresy dotyczące okresu letniego zamieszczono w załączniku.

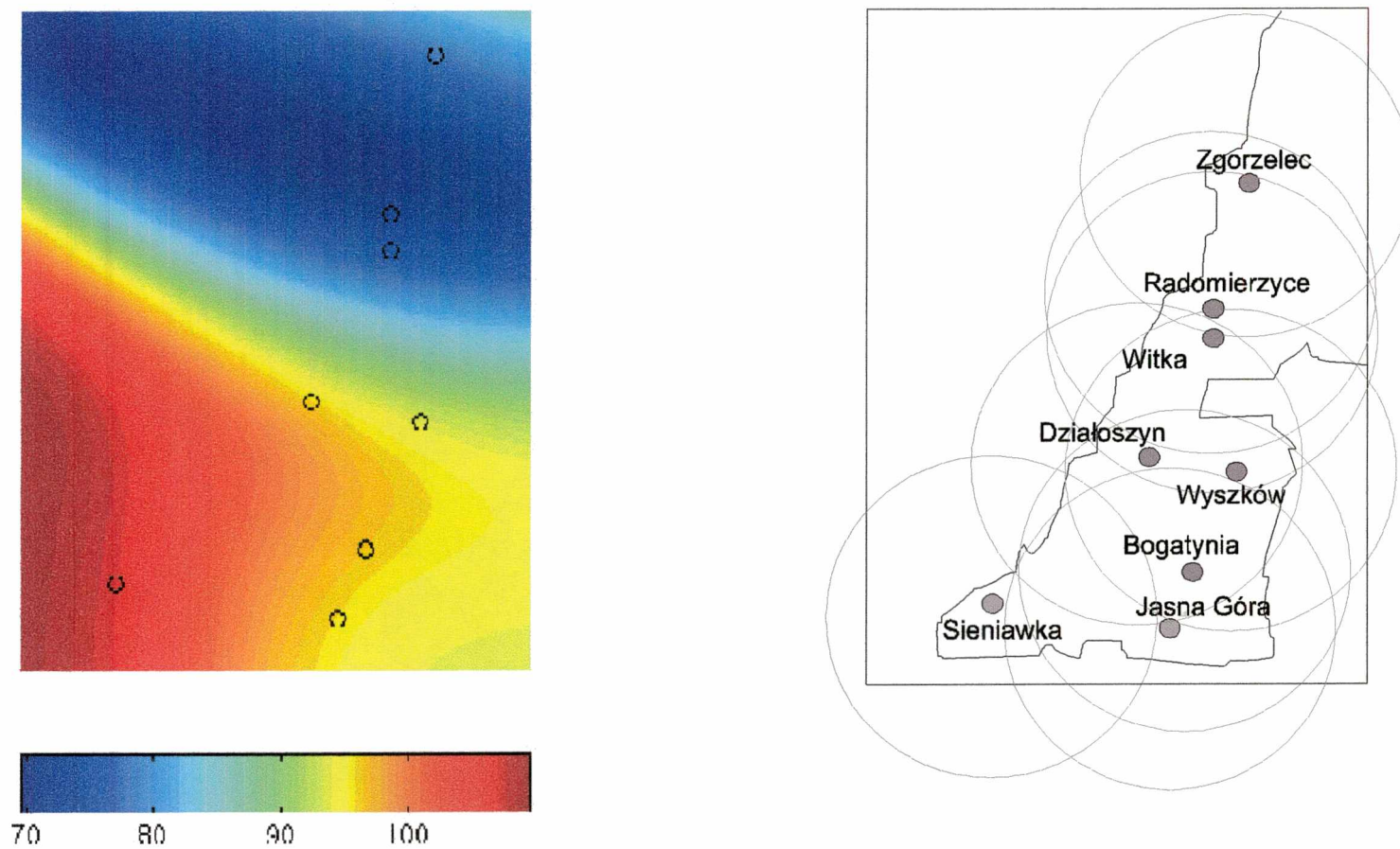
Należałoby się spodziewać, że zanieczyszczenia, których rozkłady przestrzenne są uwarunkowane szeregiem czynników lokalnych będą z większą dokładnością interpolowane przy małych wartościach spread. Natomiast pola zanieczyszczeń, w kształtowaniu których istotną rolę odgrywa wspólny czynnik o charakterze regionalnym, będą reprezentowane lepiej z zastosowaniem większych wartości spread.



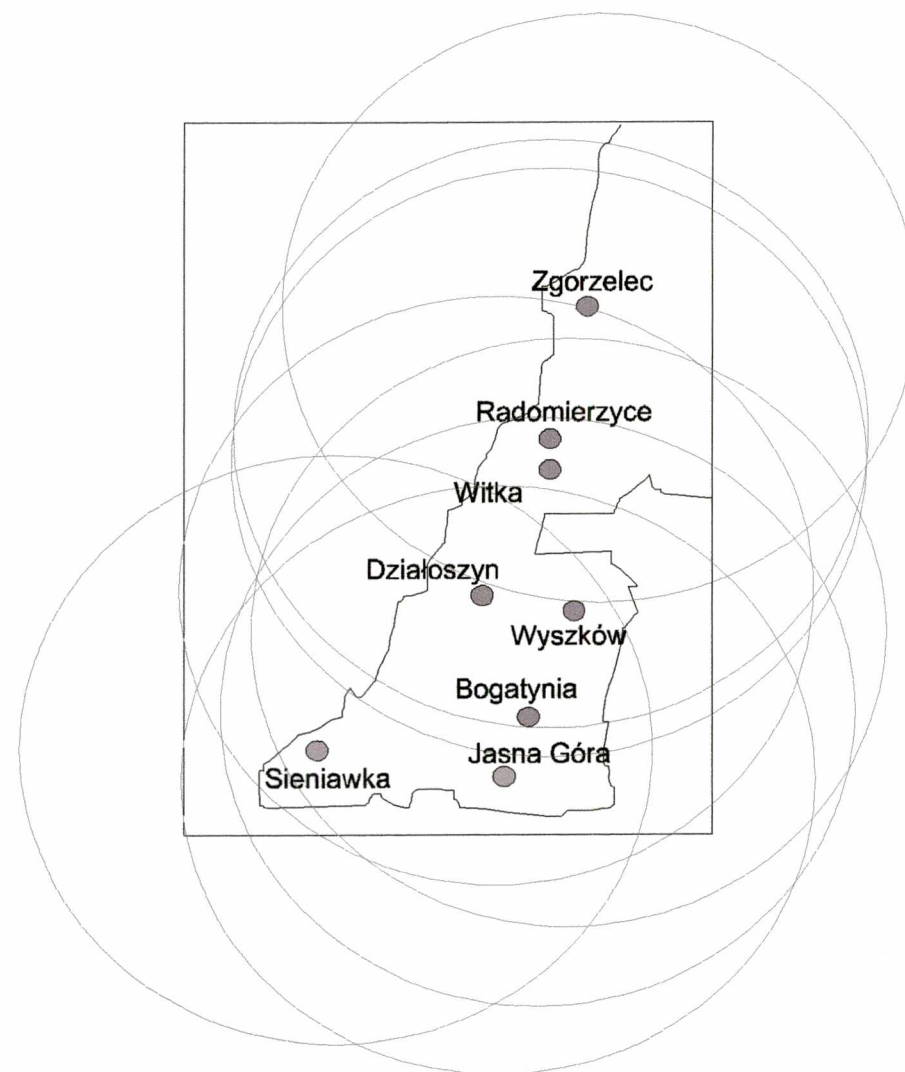
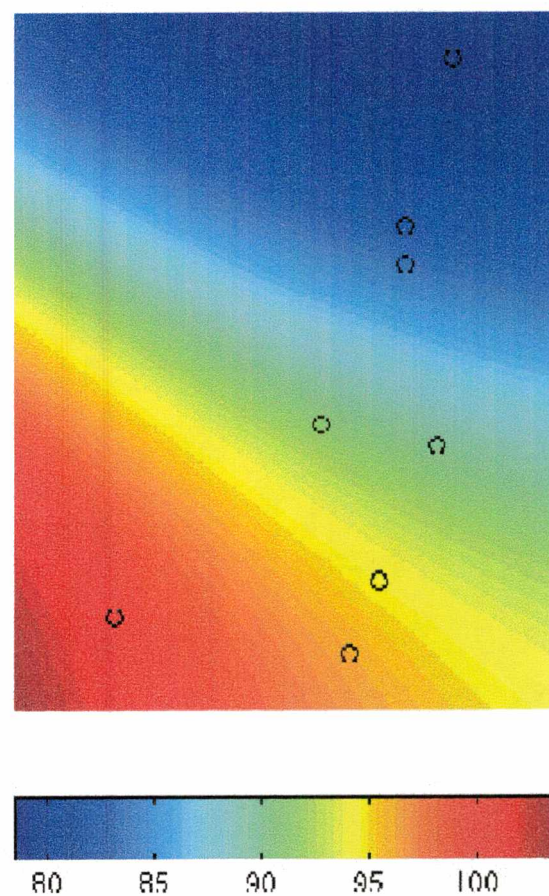
Rysunek 3.36. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 dla wartości spread nr1 i zakreślone okręgami obszary o promieniu równym spread wokół punktów pomiarowych



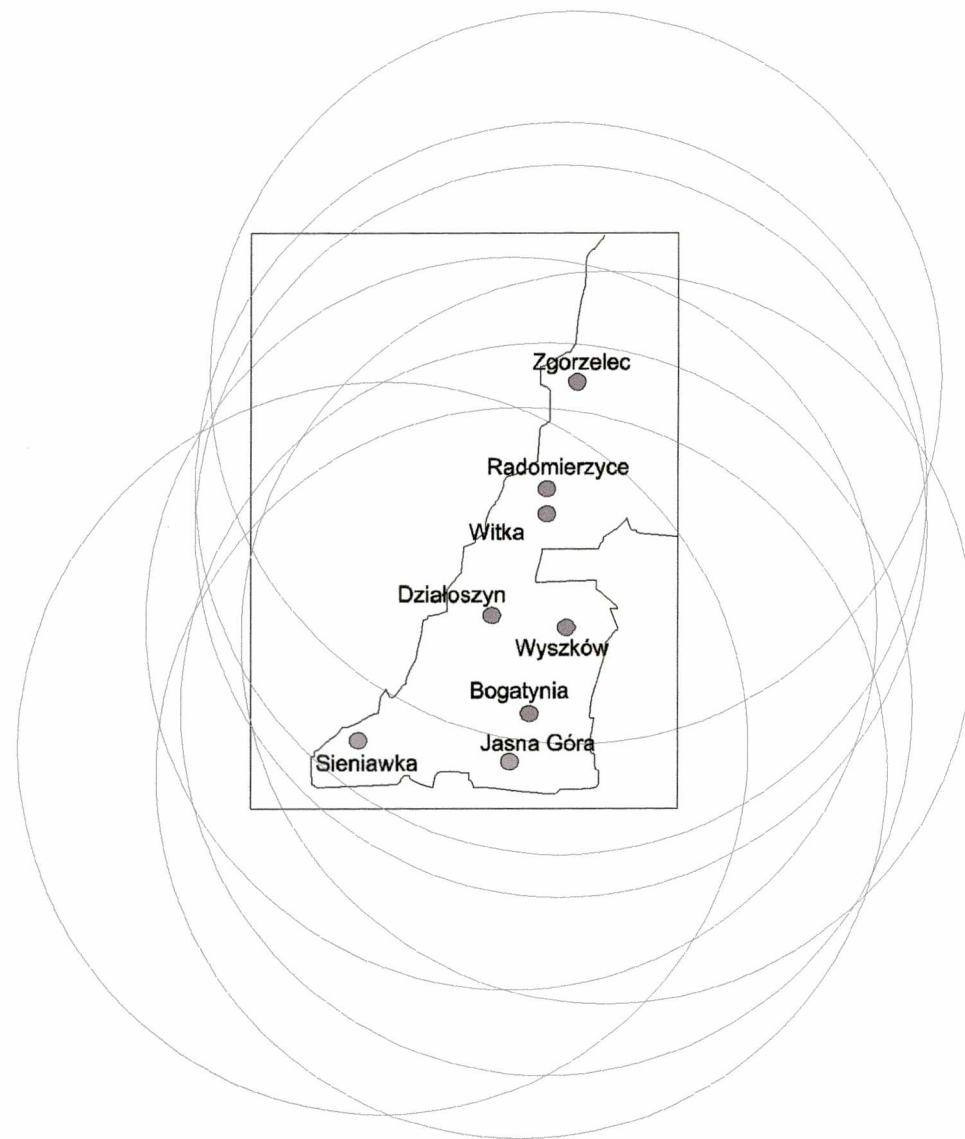
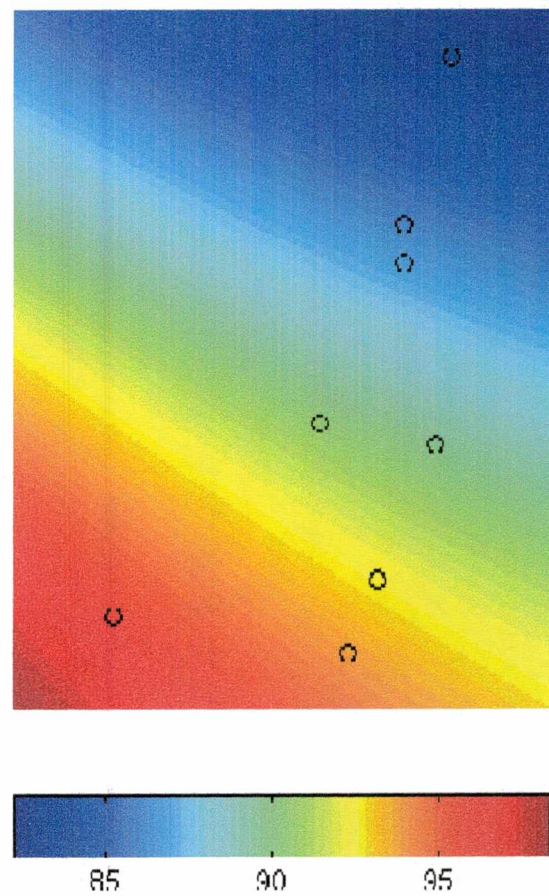
Rysunek 3.37. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 dla wartości spread nr2 i zakreślone okręgami obszary o promieniu równym spread wokół punktów pomiarowych



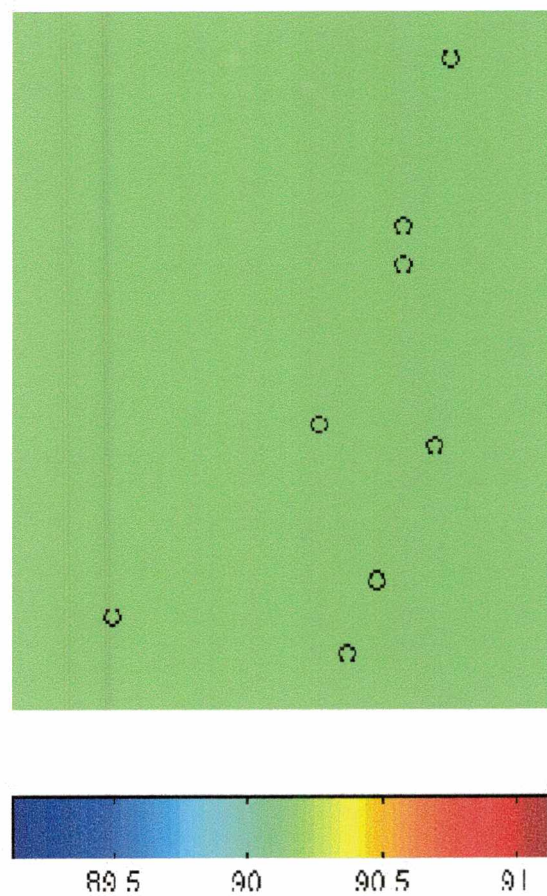
Rysunek 3.38. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 dla wartości spread nr3 i zakreślone okręgami obszary o promieniu równym spread wokół punktów pomiarowych



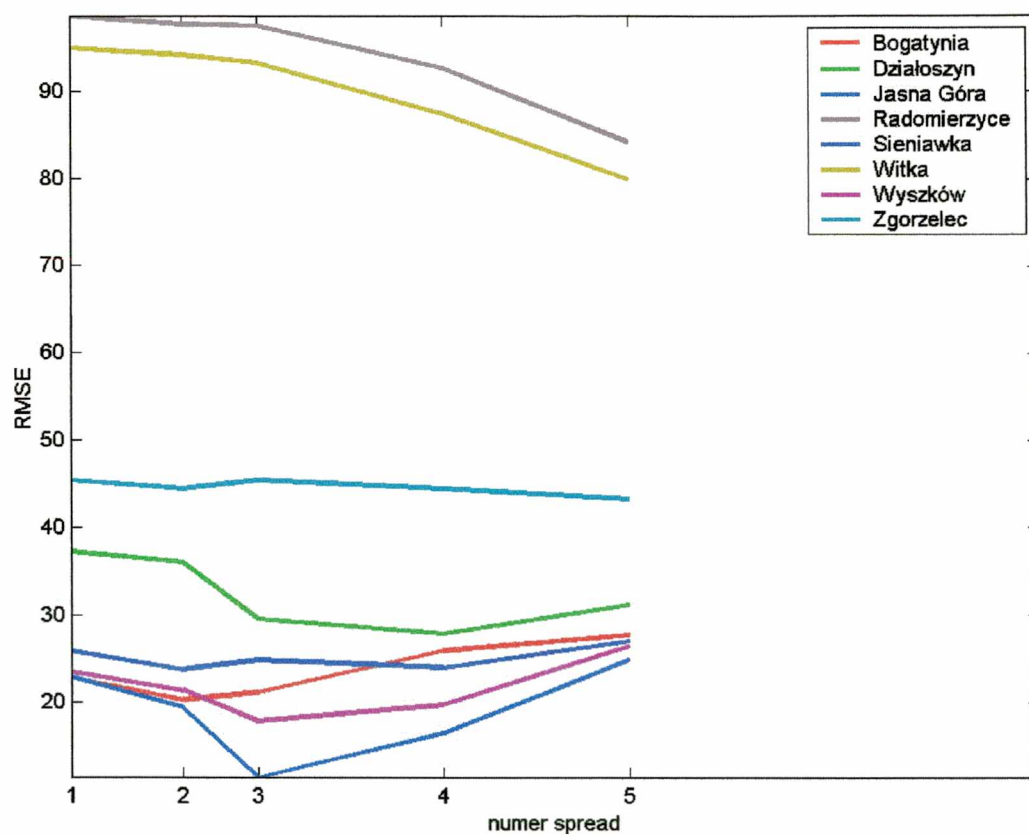
Rysunek 3.39. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 dla wartości spread nr4 i zakreślone okręgami obszary o promieniu równym spread wokół punktów pomiarowych



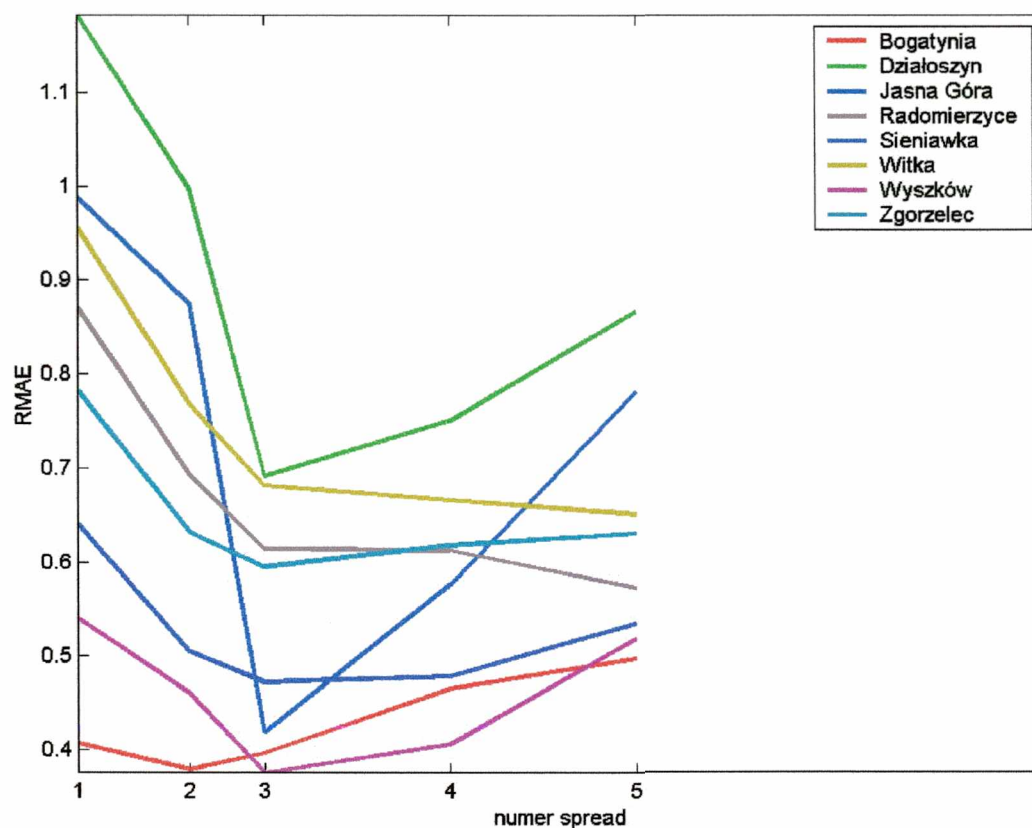
Rysunek 3.40. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 dla wartości spread nr5 i zakreślone okręgami obszary o promieniu równym spread wokół punktów pomiarowych



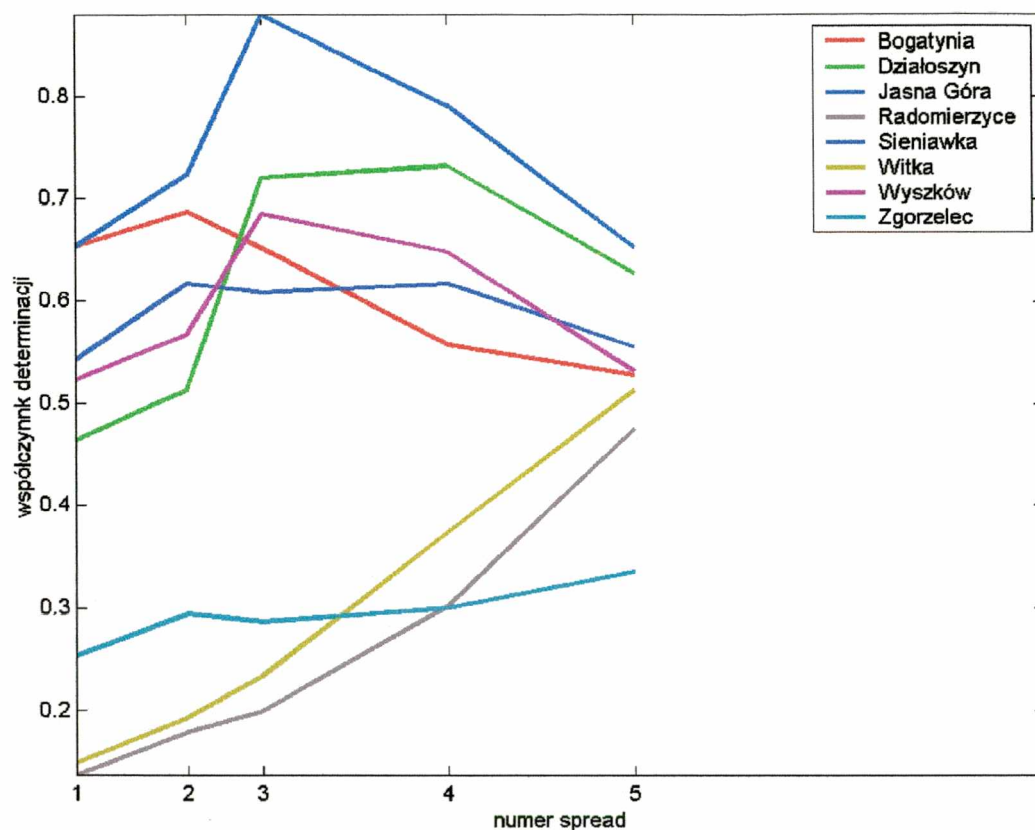
Rysunek 3.41. Przykładowy rozkład przestrzenny 30-minutowych stężeń SO_2 w formie płaszczyzny na poziomie średniego stężenia ze wszystkich stacji. Odpowiada to $\text{spread}=\infty$, czyli brakowi okręgów na rysunku po prawej stronie.



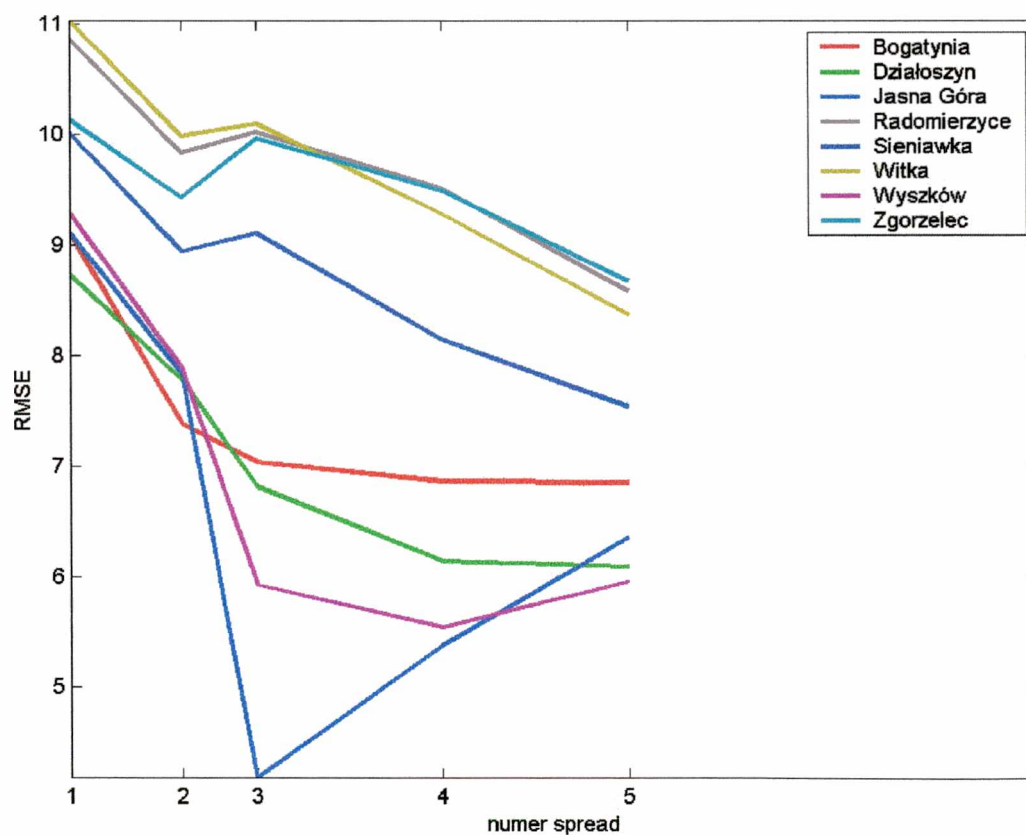
Rysunek 3.42. Zależność średniego błędu kwadratowego interpolacji stężeń 30-minutowych SO_2 [$\mu\text{g}/\text{m}^3$] od wartości spread



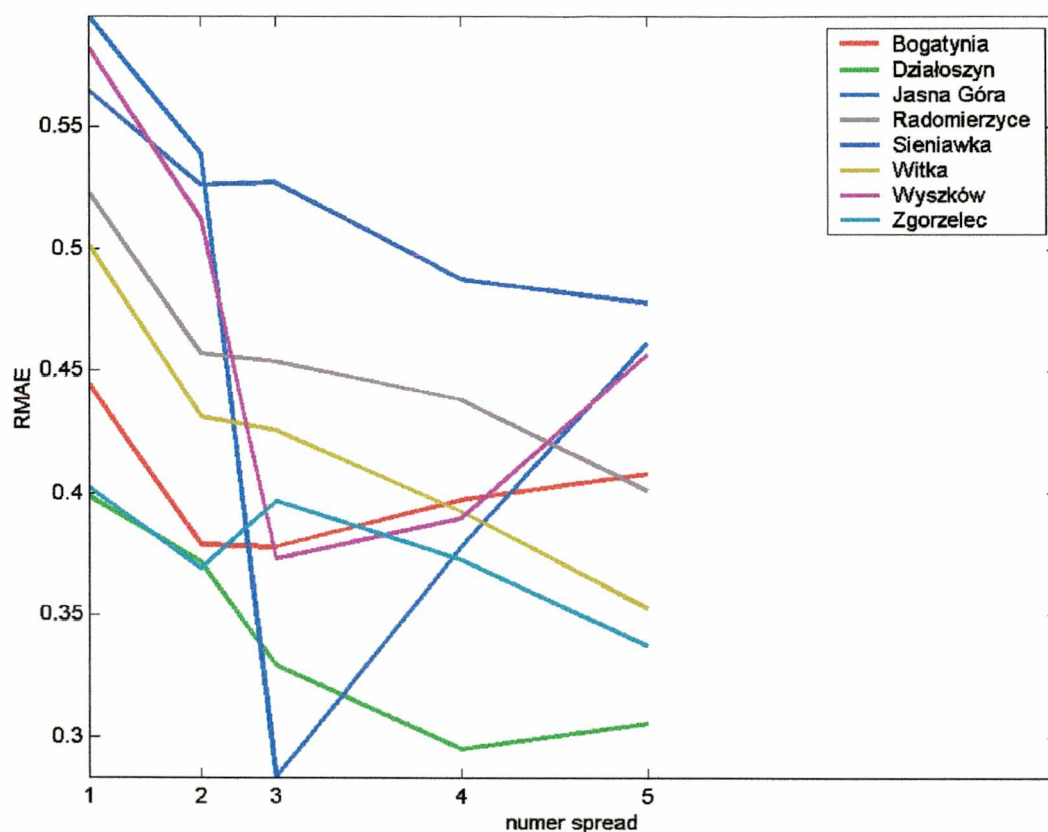
Rysunek 3.43. Zależność średniego błędu względnego interpolacji stężeń 30-minutowych SO_2 od wartości spread



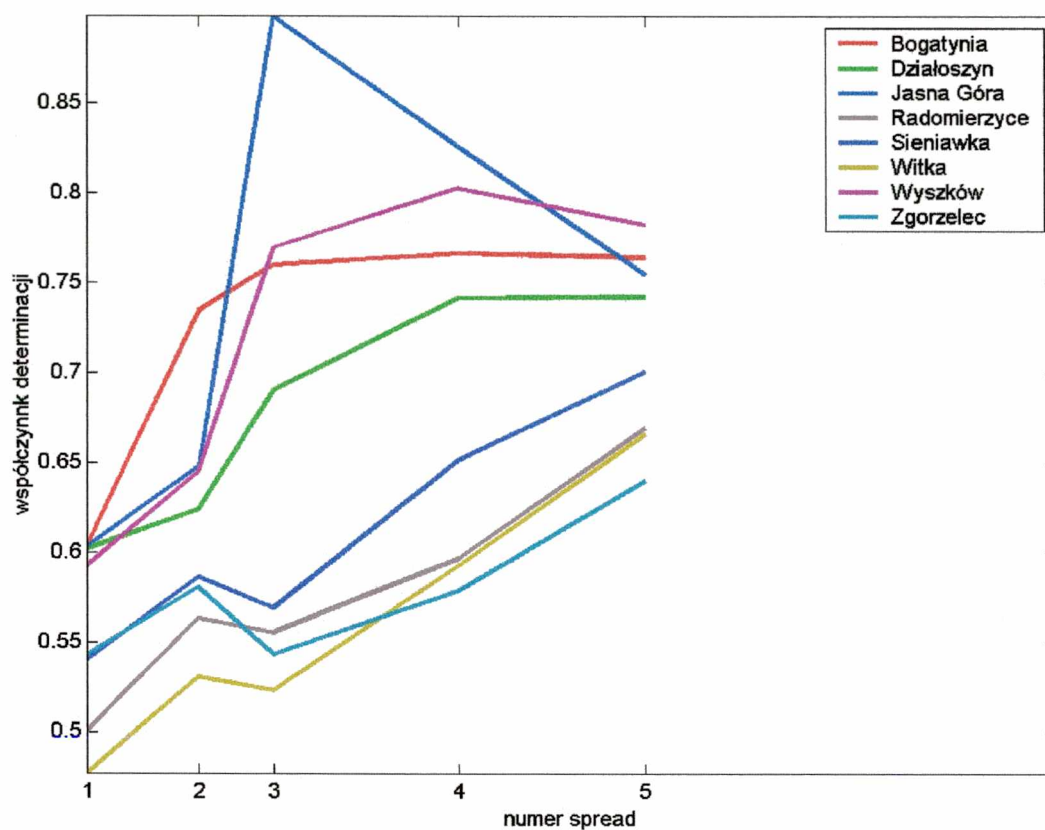
Rysunek 3.44. Zależność współczynnika determinacji dla interpolacji stężeń 30-minutowych SO_2 od wartości spread



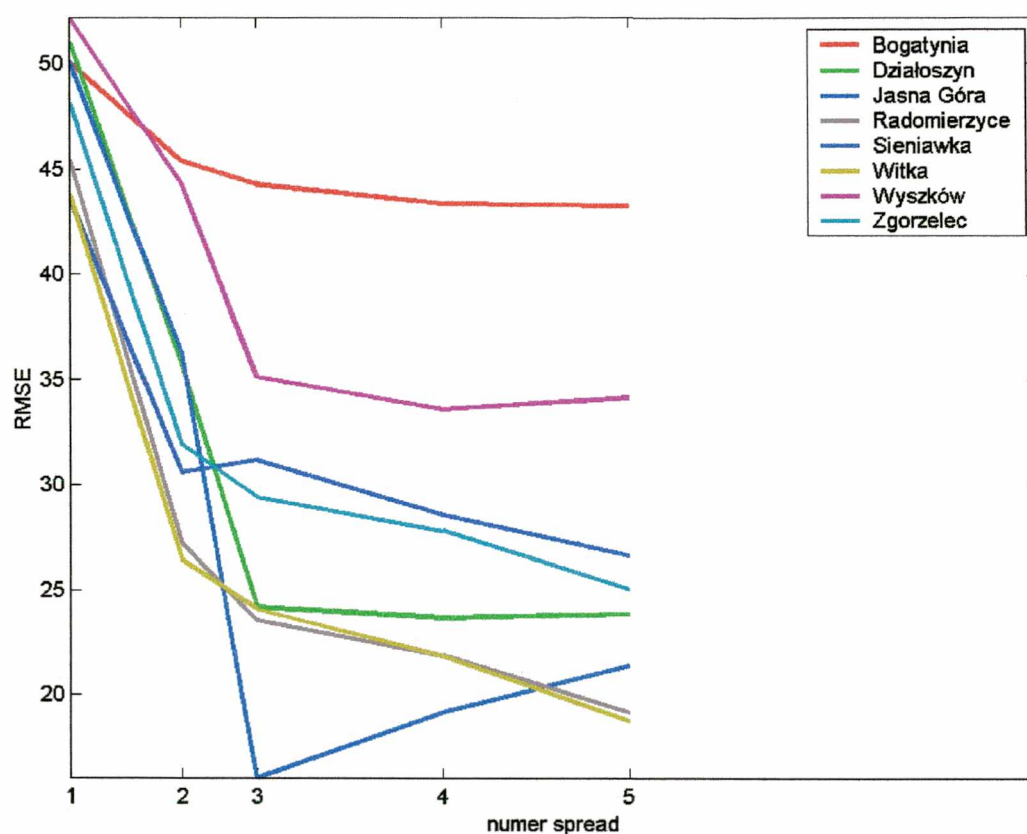
Rysunek 3.45. Zależność średniego błędu kwadratowego interpolacji stężeń 30-minutowych NO_2 [$\mu\text{g}/\text{m}^3$] od wartości spread



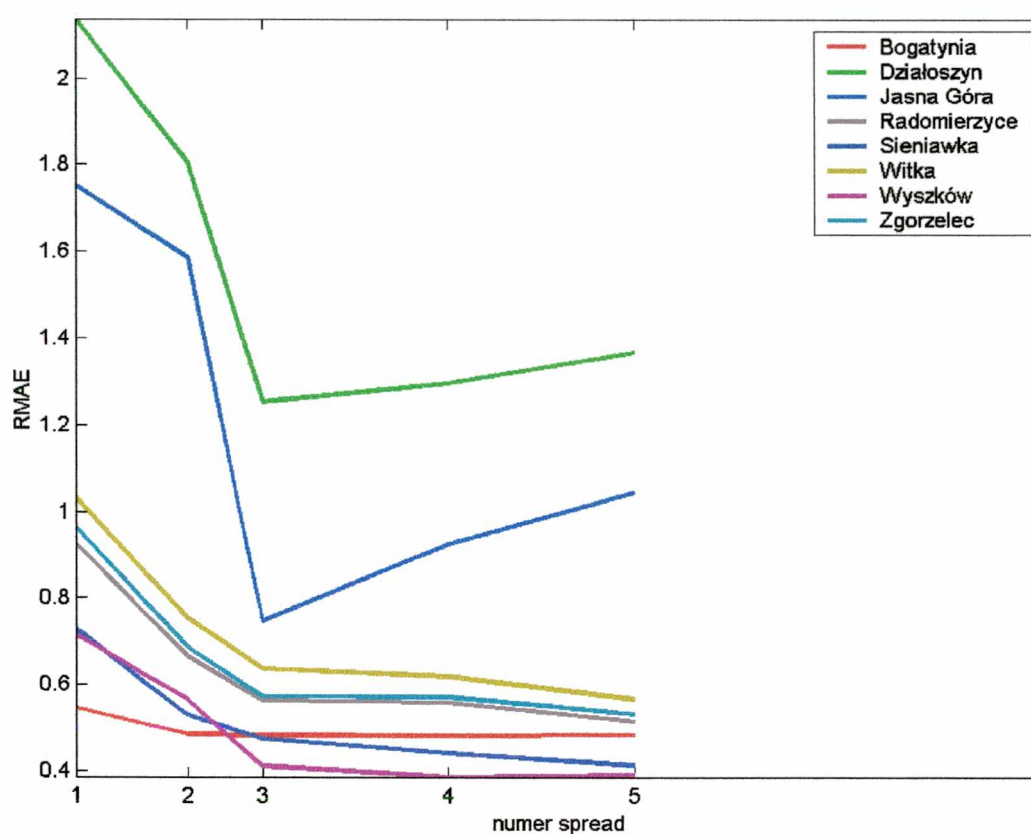
Rysunek 3.46. Zależność średniego błędu względnego interpolacji stężeń 30-minutowych NO_2 od wartości spread



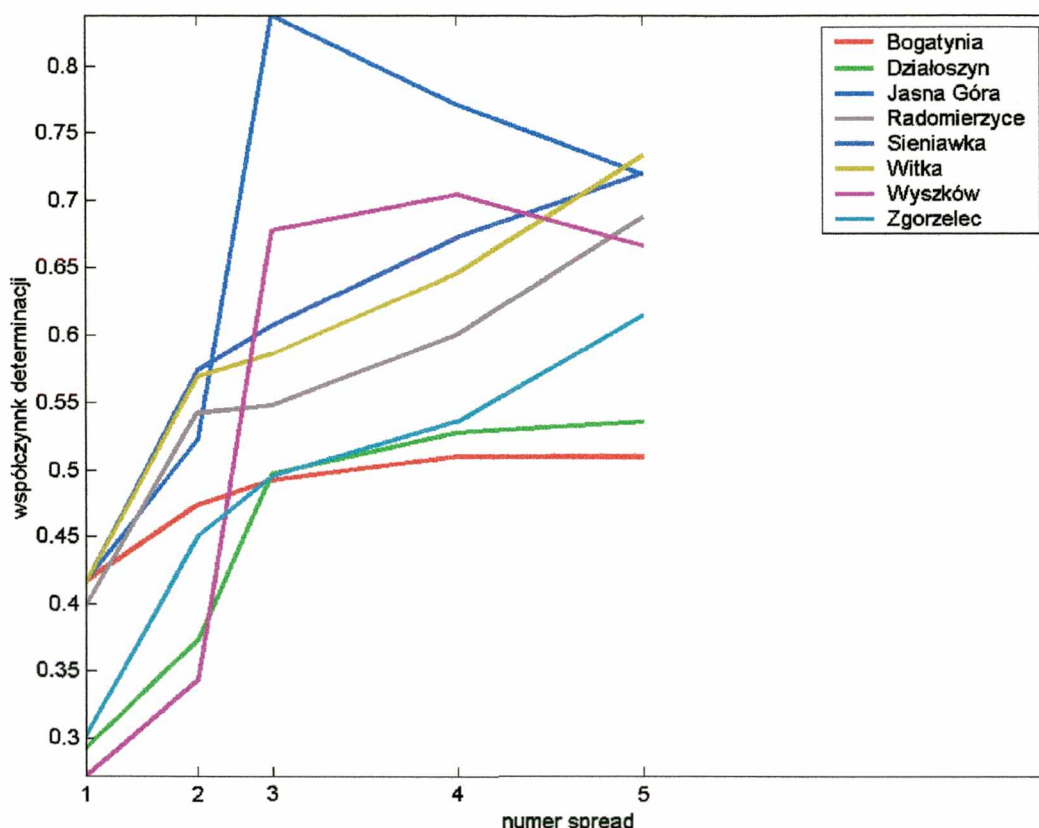
Rysunek 3.47. Zależność współczynnika determinacji dla interpolacji stężeń 30-minutowych NO_2 od wartości spread



Rysunek 3.48. Zależność średniego błędu kwadratowego interpolacji stężeń 30-minutowych pyłu [$\mu\text{g}/\text{m}^3$] od wartości spread



Rysunek 3.49. Zależność średniego błędu względnego interpolacji stężeń 30-minutowych pyłu od wartości spread



Rysunek 3.50. Zależność współczynnika determinacji dla interpolacji stężeń 30-minutowych pyłu od wartości spread

Rezultaty analiz wykonanych na pełnym zestawie danych wykazały, że najlepsze wyniki interpolacji stężeń pyłu i NO_2 otrzymano dla maksymalnej spośród wybranych wartości spread. Optimum interpolacji stężeń SO_2 wystąpiło natomiast dla pośredniej wartości spread = $\text{omin} + 0.25(\text{omax} - \text{omin})$. Najlepsza wartość spread odzwierciedla siłę powiązań stężenia w punkcie, gdzie jest wykonywana interpolacja ze stężeniami zarejestrowanymi w punktach pozostałych. Uzyskane wyniki mogą więc oznaczać, że w kształtowaniu 30-minutowych stężeń pyłu i NO_2 czynnik wspólny o zasięgu całego terenu objętego interpolacją dominuje nad czynnikami lokalnymi. Poprawność wyliczenia stężeń tych zanieczyszczeń w dowolnym punkcie rośnie w miarę coraz wyraźniejszego uwzględniania wpływu wszystkich pozostałych punktów.

Poszukując przyczyny różnic między prawidłowościami dotyczącymi poszczególnych zanieczyszczeń, oszacowano powtórnie wyniki interpolacji stężeń rozdzielając dane na okres letni i zimowy. Powyżej zaś rys.3.42 do rys.3.50 przedstawiono wyniki dotyczące okresu zimowego.

Analiza miar jakości interpolacji w funkcji wartości spread prowadzi do wniosku, że w przypadku interpolacji stężeń NO_2 oraz pyłu ta sama prawidłowość utrzymuje się zarówno w okresie letnim jak i zimowym. Natomiast w przypadku SO_2 optymalną wielkością spread w okresie letnim jest spread maksymalny natomiast w okresie zimowym najlepsze wyniki wiążą się ze spread = $\text{omin} + 0.25(\text{omax} - \text{omin})$.

Wspólne źródło wszystkich rozważanych zanieczyszczeń na terenie Worka Żytawskiego stanowi Elektrownia Turów, której oddziaływanie nakłada się na tło adwekcyjne pochodzące od elektrowni pracujących w Dolnej Saksonii i po stronie słowackiej. Można postawić tezę, że oddziaływanie emisji z tych źródeł ma charakter regionalny i jest przyczyną silnego związku stężeń między punktami pomiarowymi. Przejawem tego są optymalne wyniki interpolacji stężeń zanieczyszczeń dla maksymalnej wartości spread. Najwyraźniej obserwujemy to zjawisko dla stężeń NO_2 .

W przypadku pyłu na regionalny wpływ elektrowni nakłada się oddziaływanie zwałowiska zewnętrznego, wyrobiska oraz pomniejszego wyrobiska położonego po stronie niemieckiej poniżej Zgorzelca. Obecność tych źródeł ze względu na ich powierzchniowy charakter i rozległość wzmacnia regionalny charakter uwarunkowań zmienności stężeń pyłu. Inaczej przedstawia się obraz stężeń SO_2 . Regionalny wpływ wysokich źródeł emisji podlega ekranowaniu przez lokalne źródła emisji. Jest to wyraźnie zauważalne w okresie zimowym. Optymalna wartość spread dla interpolacji stężeń SO_2 wynosi wtedy $\text{spread} = \text{omin} + 0.25(\text{omax} - \text{omin})$, podczas gdy w okresie letnim, gdy wpływ źródeł lokalnych zanika, najlepsze wyniki uzyskuje się dla maksymalnej wartości spread.

Na możliwość uzyskania najlepszych rezultatów w wyniku interpolacji stężeń SO_2 dla mniejszych wartości spread, niż w przypadku NO_2 i pyłu wskazywały wartości wskaźników jednorodności przestrzennej stężeń przedstawione w rozdz.3.4.1, tab.3.12. Maksymalne różnice stężeń między stacjami wystąpiły w przypadku SO_2 , a ich mniejsze wartości stwierdzono dla NO_2 oraz pyłu.

Interesujące, że wspólny czynnik warunkujący zmienność stężeń wybranego zanieczyszczenia na analizowanym obszarze, według PCA, nie ma przełożenia na wyniki interpolacji. Według rezultatów analizy czynnikowej wspólny czynnik najsilniej ujawnia się w przypadku SO_2 , a najsłabiej w przypadku pyłu, co oznacza, że zmienność stężeń SO_2 jest najbardziej uwarunkowana wspólnym czynnikiem, a zmienność stężeń pyłu najsłabiej (rozd.3.4.1 (tab.3.15 do tab.3.17)). Warto jednak przypomnieć, że PCA szuka podobieństwa między stacjami w perspektywie podobieństwa zmienności czasowej stężeń. Metoda interpolacji natomiast bazuje na pojedynczym przekroju czasowym. Stąd bardziej adekwatnymi wskaźnikami kierunku doboru metody interpolacji wydają się współczynniki korelacji (rozd. 3.4.1 (tab.3.18)) i wzajemne proporcje stężeń na stacjach (rozd. 3.4.1 (tab.3.12)).

Liczbowe wartości błędów średniego kwadratowego, względnego i współczynnika determinacji dla optymalnych wartości $\text{spread} = \text{omin} + 0.25(\text{omax} - \text{omin})$ w przypadku SO_2 oraz $\text{spread} = \text{omin} + 0.75(\text{omax} - \text{omin})$ dla NO_2 i pyłu dla kompletu danych zestawiono w tabeli (tab.3.20).

W granicy powiększania spread, a więc i w granicy możliwości każdej metody interpolacji znajduje się płaszczyzna odpowiadająca średniemu poziomowi stężeń. W związku z tym w tab.3.21. zamieszczono miary jakości interpolacji przestrzennej stężeń za pomocą takiej płaszczyzny.

Tabela 3.20. Zestawienie miar jakości interpolacji dla optymalnych wartości spread w odniesieniu do poszczególnych zanieczyszczeń wg procedury cross-validation

Stacja	średni błąd kwadratowy [µg/m³]			średni błąd względny			Współczynnik determinacji		
	SO ₂	NO ₂	pył	SO ₂	NO ₂	pył	SO ₂	NO ₂	pył
Bogatynia	23	7	42	0.43	0.51	0.50	0.55	0.67	0.44
Działoszyn	33	7	26	1.34	0.43	1.33	0.39	0.59	0.43
Jasna Góra	21	8	26	0.90	0.73	1.22	0.54	0.60	0.51
Radomierzyce	77	10	23	0.90	0.69	0.69	0.12	0.52	0.51
Sieniawka	25	8	31	0.65	0.72	0.51	0.40	0.60	0.51
Witka	74	8	20	0.87	0.56	0.70	0.14	0.55	0.60
Wyszków	22	7	35	0.58	0.60	0.48	0.46	0.66	0.42
Zgorzelec	38	8	25	0.76	0.39	0.61	0.21	0.60	0.51
średnia	39	8	29	0.81	0.58	0.75	0.35	0.60	0.49

Tabela 3.21. Zestawienie miar jakości interpolacji za pomocą płaszczyzny odpowiadających średniemu poziomowi stężeń wg procedury cross-validation

Stacja	średni błąd kwadratowy [µg/m³]			średni błąd względny			współczynnik determinacji		
	SO ₂	NO ₂	pył	SO ₂	NO ₂	pył	SO ₂	NO ₂	pył
Bogatynia	29	7	42	0.53	0.53	0.50	0.44	0.67	0.43
Działoszyn	35	7	25	1.36	0.45	1.31	0.41	0.59	0.44
Jasna Góra	28	8	25	0.98	0.74	1.19	0.50	0.60	0.51
Radomierzyce	74	9	23	0.78	0.69	0.68	0.27	0.53	0.53
Sieniawka	30	8	31	0.72	0.74	0.51	0.43	0.59	0.50
Witka	72	8	20	0.85	0.56	0.67	0.26	0.56	0.63
Wyszków	28	7	35	0.68	0.61	0.47	0.42	0.66	0.42
Zgorzelec	38	8	24	0.75	0.39	0.59	0.32	0.60	0.52
średnia	42	8	28	0.83	0.59	0.74	0.38	0.60	0.50

Z tab.3.20 i tab.3.21 można wywnioskować, że najlepsze wyniki interpolacji stężeń NO₂ i pyłu uzyskuje się przybliżając pole stężeń tych zanieczyszczeń płaszczyzną na poziomie średniego stężenia ze wszystkich punktów pomiarowych. W przypadku SO₂ lepszym rozwiązaniem jest bardziej złożona reprezentacja pola stężeń, którą umożliwia radialna sieć neuronowa z wartością $spread = o_{min} + 0.25(o_{max} - o_{min})$.

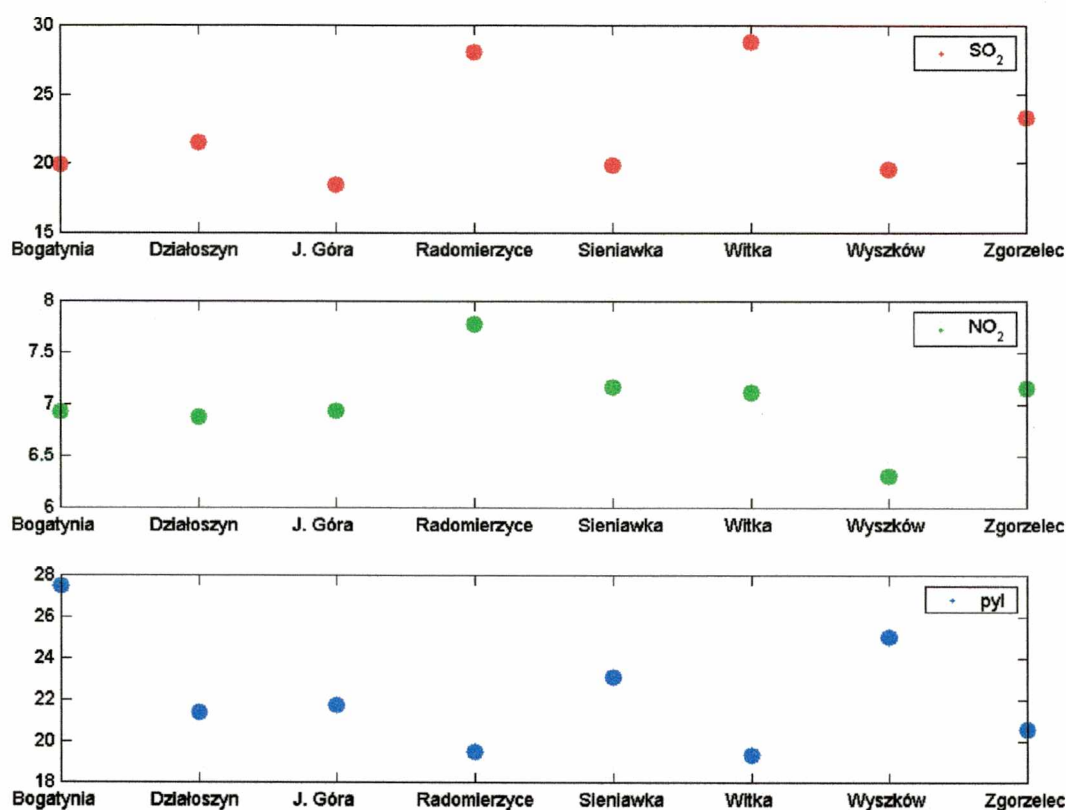
Zestawienie wykresów zmienności błędu względnego i współczynnika determinacji (rys.3.43, 3.44, 3.46, 3.47, 3.49, 3.50) oraz dane zawarte w tabeli (tab.3.20 i tab.3.21) pozwalają stwierdzić, że spośród wszystkich zanieczyszczeń najlepsze wyniki uzyskano interpolując stężenia dwutlenku azotu. Drugi w kolejności wynik daje interpolacja stężeń pyłu. Najtrudniej natomiast wygenerować przestrzenny rozkład stężeń dwutlenku siarki. Wnioski te są w zgodzie z informacją o wartościach współczynników korelacji między szeregami czasowymi

stężeń 30-minutowych SO_2 , NO_2 i pyłu na wszystkich stacjach. Wynika z nich (rozdz.3.4.1 (tab.3.18)), że najsłabiej są ze sobą skorelowane szeregi czasowe stężeń 30-minutowych SO_2 (średni współczynnik korelacji 0.57). Najsilniejsze związki występują między stężeniami NO_2 rejestrowanymi równolegle na różnych stacjach (średni współczynnik korelacji 0.69). Pośrednim natężeniem charakteryzują się zależności dla pyłu (średni współczynnik korelacji 0.61).

Ze względu na zależność jakości interpolacji od wartości spread można wyróżnić dwie grupy stacji. Przebieg krzywych na rys.3.42 do rys.3.50 w przypadku Sieniawki, Zgorzelca, Witki i Radomierzyc, czyli punktów położonych w pewnym oddaleniu od głównego zgrupowania stacji monitoringowych, jakość interpolacji wzrasta systematycznie wraz ze wzrostem spread. Dotyczy to interpolacji stężeń wszystkich trzech rozważanych zanieczyszczeń. O wzroście jakości interpolacji świadczy spadek wartości błędów średniego kwadratowego i względnego oraz wzrost współczynnika determinacji. Parametry jakości interpolacji stężeń na pozostałych stacjach poprawiają się ze wzrostem wartości spread do poziomu $\text{spread} = \text{omin} + 0.25(\text{omax} - \text{omin})$. Dalej następuje ich dramatyczne pogorszenie, jak w przypadku Jasnej Góry lub utrzymują się praktycznie na stałym poziomie, co dotyczy pozostałych stacji. Jak wynika stąd, bez względu na zauważone tendencje ogólne, istnieją przypadki, w których interpolacja stężeń zanieczyszczenia wokół punktu pomiarowego daje najlepsze rezultaty dla wartości spread różnej od obowiązującej dla większości stacji. Klasycznym tego przykładem jest punkt pomiarowy w Jasnej Górze. Optimum interpolacji stężeń wszystkich zanieczyszczeń w tym punkcie wiąże się z wartością $\text{spread} = \text{omin} + 0.25(\text{omax} - \text{omin})$. W tej sytuacji, jeżeli rozkład stężeń wokół takiego miejsca jest przedmiotem szczególnego zainteresowania, wykonując interpolację należy wybrać spread najlepszy dla tej stacji i zrezygnować ze spread optymalnego dla większości stacji. Pożądanym rozwiązaniem wydaje się sieć radialna o zróżnicowanych wartościach spread funkcji transferu neuronów radialnych sieci.

Metoda cross-validation wykazuje, że dla stałych wartości spread punkty różnią się co do wartości błędów popełnianych w wyniku wyliczania tam stężeń metodą interpolacji (tab.3.20). Średni błąd kwadratowy interpolacji stężeń w punkcie wydaje się być uwarunkowany wielkością różnic między stężeniem zanieczyszczenia w tym punkcie a stężeniami w pozostałych punktach. Średnie absolutne różnice stężeń między poszczególnymi stacjami a pozostałymi przedstawiono na rysunku (rys.3.51). Wynika z niego, że największe różnice występują w przypadku stężeń SO_2 dla stacji Radomierzyc i Witka, w przypadku stężeń NO_2 dla stacji Radomierzyc, a w przypadku pyłu dla Bogatyni, Sieniawki i Wyszkowa. W tab.3.20 można zauważyć, że wartości stężeń SO_2 , NO_2 i pyłu otrzymane w ramach procedury cross-validation

odpowiednio dla wymienionych stacji są obarczone największym średnim błędem kwadratowym. Wniosek stąd, że im bardziej stężenie w wybranym punkcie odbiega od stężeń w pozostałych punktach, tym większy błąd jest popełniany podczas wyliczania tego stężenia metodą interpolacji. Dotyczy to w równym stopniu rozwiązań uwzględniających zmienność lokalną stężeń (SO_2) jak regionalną (NO_2 , pył).



Rysunek 3.51. Średnie z różnic 30-minutowych stężeń zanieczyszczeń między daną stacją a pozostałymi [$\mu\text{g}/\text{m}^3$]

Błędy interpolacji, oszacowane na podstawie wyników uzyskanych metodą cross-validation wydają się duże (tab.3.20). Dzięki metodzie interpolacji wykorzystującej sieci neuronowe sprawdzono pełne spektrum możliwych rozwiązań problemu. Począwszy od rozważenia dominującej roli czynników lokalnych w kształtowaniu pola stężeń, a kończąc na maksymalnym uproszczeniu zakładającym średni poziom stężeń nad całym terenem. Wydaje się więc, że w wyniku wyboru optymalnego rozwiązania błąd metody został zminimalizowany. Poziom błędu, którego nie potrafimy pokonać wynika z niedostatecznej pojemności informacyjnej wykorzystanych danych pomiarowych.

4. Podsumowanie i wnioski

Przedmiotem pracy jest prognozowanie oraz interpolacja 30-minutowych stężeń imisyjnych w sąsiedztwie dominującego punktowego źródła emisji z wykorzystaniem neuronowych modeli prognostycznych i neuronowej metody interpolacji.

Praca obejmuje:

- przegląd statystycznych modeli prognostycznych, analitycznych metod interpolacji oraz sieci neuronowych pod kątem ich możliwości i ograniczeń,
- opracowanie neuronowych modeli prognostycznych i neuronowej metody interpolacji stężeń imisyjnych,
- dyskusję wyników zastosowania opracowanych modeli prognostycznych i metody interpolacji do modelowania stężeń zanieczyszczeń rejestrowanych przez sieć monitoringu powietrza Elektrowni Turów w Worku Żytawskim.

Szereg wniosków, które nasuwają się po przeanalizowaniu materiału przedstawionego w pracy sformułowano poniżej.

Określono poziom wpływu parametrów meteorologicznych rejestrowanych standardowo przez stacje monitoringu zanieczyszczeń powietrza na zmienność 30-minutowych stężeń zanieczyszczeń. Jak wykazały rezultaty najlepszych spośród neuronowych modeli prognostycznych opartych na podejściu deterministycznym prędkość wiatru, kierunek wiatru, wilgotność, temperatura i ciśnienie atmosferyczne wspólnie wyjaśniają ok. 50% zmienności stężeń NO_2 , 40% dla SO_2 i 30% w przypadku pyłu. Opis zmienności stężeń 30-minutowych wyłącznie na podstawie parametrów meteorologicznych jest więc mało efektywny i niewystarczający dla celów praktycznych.

Opracowano metodę identyfikacji dodatkowego źródła emisji na obszarze zdominowanym wpływem jednego źródła. Metoda wykorzystywała fakt, że zmiany jakości neuronowych modeli deterministycznych w funkcji grup warunków meteorologicznych są jednakowe dla zanieczyszczeń pochodzących ze wspólnego źródła. Wynika to z podobnego wpływu warunków meteorologicznych na zmienność stężeń imisyjnych zanieczyszczeń emitowanych z jednego źródła, np. SO_2 , NO_2 i pyłu. Model opracowany dla zanieczyszczenia, które podlega wpływowi dodatkowego źródła reaguje na zmianę warunków meteorologicznych inaczej od pozostałych. Korzystając z tej prawidłowości postawiono tezę, że w Bogatyni NO_2 i pył są tego samego pochodzenia a w przypadku SO_2 ujawnia się wpływ emisji lokalnej.

Wykazano możliwość neuronowego prognozowania 30-minutowych stężeń zanieczyszczeń w układzie niedeterministycznym, czyli na podstawie stężeń występujących w szeregu czasowym wcześniej niż prognozowane. Ze względu na fakt, że najsilniejszy związek występuje między stężeniami sąsiadującymi ze sobą w serii czasowej, modele nieliniowe nie wykazują przewagi nad liniowymi a optymalny horyzont prognozy wynosi tylko 0.5 godziny. Absolutny błąd względny prognozy wykonanej za pomocą modelu neuronowego dla takiego horyzontu wynosi 12% w przypadku stężeń NO_2 , 20% dla SO_2 i 29% w przypadku pyłu. Horyzont prognozy o długości 0.5 godziny jest jednak zbyt krótki dla zastosowań praktycznych. Istnieje możliwość wydłużenie horyzontu prognozy do 6 godz. w przypadku NO_2 , 3 godz. w przypadku SO_2 i tylko 1 godz. dla pyłu. Wiąże się to jednak z koniecznością zaakceptowania błędu prognozy na poziomie 50%.

Stwierdzono większą przydatność modeli pracujących w układzie niedeterministycznym do prognozowania szeregów czasowych charakteryzujących się łagodną zmiennością stężeń niż serii czasowych w których występują gwałtowne skoki stężeń. Świadczą o tym najlepsze rezultaty prognoz w przypadku NO_2 , a najgorsze w przypadku pyłu.

Opracowano neuronową metodę interpolacji przestrzennej na bazie sieci radialnej. Metoda opiera się na założeniu, że rozkład wielkości mierzonej - w tym przypadku stężeń zanieczyszczeń - wokół punktu pomiarowego ma postać powierzchni Gaussa. Dzięki temu, metoda umożliwia wnioskowanie o regionalnych lub lokalnych uwarunkowaniach przestrzennego rozkładu stężeń zanieczyszczeń na podstawie optymalnej rozpiętości powierzchni Gaussa. Metoda pozwala na adaptację rozkładu przestrzennego do cech rzeczywistego pola stężeń i z tego względu jest bardzo użyteczna w praktyce.

W oparciu o neuronową metodę interpolacji wykazano, że przestrzenny rozkład stężeń NO_2 i pyłu na terenie Worka Żytawskiego jest zdeterminowany czynnikami o regionalnej skali oddziaływania. Należą tu prawdopodobnie tło adwekcyjne pochodzące od niemieckich i słowackich wysokich źródeł emisji, na które nakłada się oddziaływanie Elektrowni Turów S.A. i Kopalni KWB Turów S.A. wraz z wyrobiskiem i zwałowiskiem zewnętrznym. Przestrzenny rozkład stężeń NO_2 i pyłu najlepiej reprezentuje płaszczyzna na poziomie średniego stężenia dla obszaru. Taki wynik uzyskuje się przy interpolacji z wykorzystaniem powierzchni Gaussa o nieskończonej rozpiętości.

Posługując się neuronową metodą interpolacji stwierdzono istnienie czynników silnie lokalnie oddziałujących na przestrzenny rozkład stężeń SO_2 w okresie grzewczym. Ten fakt przejawiał się uzyskaniem najlepszych wyników interpolacji stężeń SO_2 w sezonie grzewczym przy wykorzystaniu powierzchni Gaussa o małej rozpiętości.

Podsumowując należy stwierdzić, że opracowane neuronowe modele prognostyczne i neuronowa metoda interpolacji umożliwiają pozyskiwanie informacji o uwarunkowaniach czasowej i przestrzennej zmienności stężeń zanieczyszczeń powietrza. Wydaje się, że dalsze badania powinny pójść w kierunku integracji opisu imisji w czasie i przestrzeni.

Spis literatury

- Asbaugh L., Myrup L., Flocchini R., 1984, A principal component analysis of sulfur concentrations in the Western United States. *Atmospheric Environment*, Vol.18, No.4, 783-91.
- Baran J., Świąć A., Hoffman S., 1999. Zastosowanie sztucznych sieci neuronowych do modelowania stężenia ozonu przy powierzchni ziemi. *Inżynieria Ochrony Środowiska*, t.2, nr.3-4, 305-21.
- Bower S.J., Broughton F.J., Stedman J.R., Williams M.L., 1994. A winter NO₂ smog episode in the U.K. *Atmospheric Environment*, Vol.28, No.3, 461-75.
- Boznar M., Lesjak M., Primož M., 1993. A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment*, Vol. 27B, No.2, 221-30.
- Boznar M., Mlakar P., 1995. Neural networks - a new mathematical tool for air pollution modelling. *Air Pollution Theory and Simulation, Proc. of III International Conference on Air Pollution, Computational Mechanics Publications*, 259-66.
- Boznar M., Mlakar P., 1996. Analysis of Ambient SO₂ concentrations in the surrounding of the Sostanj thermal power plant. *Air Pollution IV Monitoring, Simulation and Control, Proc. of IV International Conference on Air Pollution, Computational Mechanics Publications*, 727-36.
- Brzózka J., Dobroczyński L., Programowanie w Matlabie, MIKOM, Warszawa 1998.
- Carter R., Chaffin C., Thomas M., A method of predicting point and path-averaged ambient air VOC concentrations, using meteorological data. *J. Air Waste Manage. Assoc.*, 43, 480-88.
- Chen J.L., Islam S., Biswas P., 1998, Nonlinear dynamics of hourly ozone concentrations: nonparametric short term prediction. *Atmospheric Environment*, Vol.32, 1839-48.
- Chock D. P., Terrell T.R., Levitt S.B., 1975. Time-series analysis of Riverside, California air quality data. *Atmospheric Environment*, Vol.9, 978-89.
- Christakos G., Virkam M. V., 1998. A composite space/time approach to studying ozone distribution over eastern United States, *Atmospheric Environment*, Vol.32, No.16, 2845-57.
- Christensen R., Linear models for multivariate, time series and spatial data, Springer-Verlag, New York 1991.
- Chu D., Kriging software for Matlab 5.2-EasyKrig2.1, 2000.
- Cieślak M., Dittmann P., Kuropka I., Krupowicz J., Radzikowska B., Załuska U., Kwiatkowska-Ciotucha D., Prognozowanie gospodarcze, metody i zastosowanie, PWN, Warszawa 1997.
- Comrie A., 1997. Comparing neural networks and regression models for ozone forecasting, *J. of Air and Waste Manage. Assoc.*, Vol.47, 653-63.
- Davis J. M., Eder B.K., Nychka D., Yang Q., 1998. Modelling the effect of meteorology on ozone in Houston using cluster analysis and generalized additive models. *Atmospheric Environment*, Vol. 32, No.14/15, 2505-20.

- Derwent R.G., Middleton D.R., 1995. Analysis and interpretation of air quality data from urban roadside location in Central London over the period from July 1991 to July 1992. *Atmospheric Environment*, Vol. 29, No.8, 923-46.
- Eder B.K., Davis J.M., Bloomfield P., 1993. A characterization of the spatiotemporal variability of non-urban ozone concentrations over the Eastern United States. *Atmospheric Environment*, Vol.27A, No.16, 2645-68.
- Egmond N.D., Onderdelinden D., 1981. Objective analysis of air pollution monitoring network data; spatial interpolation and network density. *Atmospheric Environment*, Vol.15, 1035-46.
- Elsom D.M., 1978. Spatial correlation analysis of air pollution data in an urban area. *Atmospheric Environment*, Vol.12, 1103-07.
- Fedorov V.V., 1989. Kriging and other estimators of spatial field characteristics (with special reference to environmental studies). *Atmospheric Environment*, Vol.23, 175-84.
- Foszcz D., Gawenda T., Siewior J., Tumidajski T., 2000., Modele prognoz średniego dobowego stężenia SO₂ dla wybranych miast Górnośląskiego Okręgu przemysłowego. *Ochrona powietrza i problemy odpadów*, vol. 34, No. 5, 174-179.
- Gajek L., Kałuszka M., Wnioskowanie statystyczne, WNT, Warszawa 1996.
- Gardner M.W., Dorling S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment*, Vol.34, 21-34.
- Gardner M.W., Dorling S.R., 1999. Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmospheric Environment*, Vol.33, 709-19.
- Gardner M.W., Dorling S.R., 1998. Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. *Atmospheric Environment*, Vol. 32, No.14/15, 2627-36.
- Greń J., Statystyka matematyczna modele i zadania, PWN, Warszawa 1975.
- Guardani R., Nascimento C., Guardani M., Martins M., Romano J., 1999. study of atmospheric ozone formation by means of neural network-based model. *J. of Air and Waste Manage. Assoc.*, Vol.49, 316-323.
- Hernandez E., Martin F., Valero F., 1992. Statistical forecast models for daily air particulate iron and lead concentrations for Madrid, Spain. *Atmospheric Environment*, Vol.26B, No.1, 107-16.
- Hertz J., Krogh A., Palmer R.G., Wstęp do teorii obliczeń neuronowych, WNT, Warszawa 1995.
- Hopkins L.P., Ensor K.B., Rifai H.S., 1999. Empirical evaluation of ambient ozone interpolation procedure to support exposure models. *J. of Air&Waste Manage. Assoc.* 49, 839-46.
- Horowitz J., Barakat S., 1979. Statistical analysis of the maximum concentration of an air pollutant effects of autocorrelation and non-stationarity. *Atmospheric Environment*, Vol.13, 811-18.
- Hwang J.S., Chan C.C., 1997. Redundant measurements of urban air monitoring networks in air quality reporting. *J. of Air&Waste Manage. Assoc.* 47, 614-19.
- Jorquera H., Palma W., Tapia J., 2000. An intervention analysis of air quality data at Santiago, Chile. *Atmospheric Environment*, Vol.34, 4073-84.
- Jorquera H., Perez R., Cipriano A., Espejo A., Letelier V., Acuna G., 1998. Forecasting ozone daily maximum levels at Antiago, Chile. *Atmospheric Environment*, Vol.32, No.20, 3415-3424.

- Juang K.W. , Lee D.Y., 1998. Simple indicator kriging for estimating the probability of incorrectly delineating hazardous areas in a contaminated site. *Environ. Sci. Technol.*, 32 2487-93.
- Juda J., Chróściel S., Ochrona Powietrza Atmosferycznego, PWN, Warszawa 1974.
- Kamiński W., Skrzypski J., Strumiłło P., 1999. Zastosowanie sieci neuronowych do lokalizacji stacji monitoringu zanieczyszczeń powietrza. *Archiwum Ochrony Środowiska*, vol.25, No.3, 83-90.
- Kao J., Huang S., 2000. Forecasts using neural network versus Box-Jenkins methodology for ambient air quality monitoring data. *J. of Air and Waste Manage. Assoc.*, Vol.50, 219-26.
- Kartal S., Ozer U., 1998. Determination and parametrization of some air pollutants as a function of meteorological parameters in Kayseri, Turkey. *J. Air&Waste Manage. Assoc.* 48, 853-59.
- Kolehmainen M., Martikainen H., Russkanen J., 2001. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, Vol.35, 815-25.
- Kurzyński M., Rozpoznawanie obiektów - metody statystyczne, Oficyna wyd. PWR, Wrocław 1997.
- Langstaff J., Seigneur C., Liu M.K., 1987. Design of an optimum air monitoring network for exposure assessment. *Atmospheric Environment*, Vol. 21, No.6, 1393-1410.
- Liu X., Gao N., Hopke P.K., Cohen D., Baileys G., Crisp P., 1996. Evaluaton of spatial patterns of fine particle sulfur and lead concentrations in New South Wales, Australia. *Atmospheric Environment*, Vol.30, No.1, 9-24.
- Luszniewicz A., Słaby T., Statystyka stosowana, PWE, Warszawa 1997.
- Malinowski P., Maciejewska M., Zwoździak J., 1999. Klasyfikacja przestrzennych rozkładów 30-min stężeń SO₂ z zastosowaniem sieci neuronowej, POLIMIS'99
- Mantri S., Bullock D., Garrett J., 1997. Vehicle detection using a hardware-implemented neural-net. *AI in Civil Engineering*, 15-21.
- Masters T., Sieci neuronowe w praktyce, WNT, Warszawa 1996.
- McCollister G. M., Wilson K., 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmospheric Environment*, Vol.9, 417-23.
- Melas D., Kioutsioukis I., Ziomas I.C., 2000. Neural network model for predicting peak photochemical pollutant levels. *J. of Air and Waste Manage. Assoc.*, Vol.50, 495-501.
- Milionis A.E., Davies T.D. 1994. Regression and stochastic models for air pollution - I. review, comments and suggestions. *Atmospheric Environment*, Vol.28, No.17, 2801-2809.
- Milionis A.E., Davies T.D. 1994. Regression and stochastic models for air pollution - II. Application of stochastic models to examine the links between ground-level smoke concentrations and temperature inversions. *Atmospheric Environment*, Vol.28, No.17, 2810-22.
- Miller A.A., sager T.W., 1994. Site redundancy in urban ozone monitoring. *J. of Air&Waste Manage. Assoc.* 44, 1097-1102.
- Mlakar M., Boznar M., 1994. Short term pollution prediction on the basis of artificial neural networks. *Air Pollution II Vol. 1: Computer Simulation, Proc. of II International Conference on Air Pollution, Computational Mechanics Publications*, 545-52.

- Mlakar P., Boznar M., 1996. Analysis of winds and SO₂ concentrations in complex terrain. *Air Pollution IV Monitoring, Simulation and Control, Proc. of IV International Conference on Air Pollution, Computational Mechanics Publications*, 455-64.
- Mlakar P., Boznar M., Lesjak M., 1994. Neural networks predict pollution. *Air Pollution Modelling and its Application X*, Plenum Press, Vol. 18, 659-60.
- Mrożek B., Mrożek Z., Matlab5.x Simulink 2.x, PLJ, Warszawa 1998.
- Mulholland J.A., Tolbert P.E., Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications. *J. of Air&Waste Manage. Assoc.* 48, 418-26.
- Murlikiewicz K., Stawczyk J., Zbiciński I., 1999. Modelowanie neuronowe emisji gazów w procesie spalania pulsacyjnego.
- Myrup L.O., Flocchini R.G., 1984. A principal components analysis of sulfur concentrations in the western United States. *Atmospheric Environment*, Vol.18, No.4, 783-91.
- Nakamori Y., Sawaragi Y., 1984. Interactive design of urban level air quality monitoring network. *Atmospheric Environment*, Vol.18, No.4, 793-99.
- Nakamori Y., Ikeda S., Sawaragi Y., 1979. Design of air pollutant monitoring system by spatial sample stratification. *Atmospheric Environment*, Vol.15, 97-103.
- Noll K., Miller T., Norco J., Raufer R.K., 1977. An objective air monitoring site selection methodology for large point sources. *Atmospheric Environment*, Vol.13, 1051-59.
- Ossowski S., Sieci neuronowe w ujęciu algorytmicznym, WNT, Warszawa 1996.
- Ośródk K., Ośródk L., Wojtylak M., 1995. Zastosowanie metody sztucznych sieci neuronowych do prognozy zanieczyszczeń powietrza w aglomeracji miejsko-przemysłowej. *Wiadomości IMGW*, Tom XVIII, zesz. 3-4, 5-18.
- Perez P., Trier A., Reyes J., Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment*, Vol.34, 1189-96.
- Poissant L., Bottenhaim J.W., Roussel P., Redis N., Niki H., 1996. Multivariate analysis of a 1992 SONTOS data subset. *Atmospheric Environment*, Vol.30, No.12, 2133-44.
- Pyta H., Trzpiot G., 1999. Prognozowanie zmian stężenia zanieczyszczeń komunikacyjnych w krótkich przedziałach czasowych. *Archiwum Ochrony Środowiska*, vol. 25, No.4, 7-15.
- Reich S.L., Gomez D.R., Dawidowski L.E., 1999. Artificial neural network for the identification of unknown air pollution sources. *Atmospheric Environment*, Vol.33, 3045-52
- Reifaman J., Feldman E., 1998. Identification and control of NO_x emissions using neural networks. *J. of Air and Waste Manage. Assoc.*, Vol.48, 408-17.
- Rzewuski M., Karbowski A., Malinowski K., 1998, Rozproszone asynchroniczne uczenie dużych sieci neuronowych złożonych zjawisk fizycznych na przykładzie zadania estymacji wielkości strat powodziowych.
- Shi J. P., Harrison R. M., 1997. Regression modelling of hourly NO_x and NO₂ concentrations in urban air in London. *Atmospheric Environment*, Vol.31, No.24, 4081-94.
- Shlink U., Herbarth O., Tetzlaff G., 1997. A component time-series model for SO₂ data: forecasting, interpretation and modification. *Atmospheric Environment*, Vol.31, No.9, 1285-95.

- Soja G., Soja A-M., 1999. ozone indices based on simple meteorologiczal parameters: potential and limitations of regression and neural network models. *Atmospheric Environment*, Vol.33, 4299-4307.
- Sonkin L.P., Nikołajew W.D., 1993. Sinopticzeskij analiz i prognoz zagrażenia atmosfery, *Meteorologija i gidroloija*, Nr 5, 14-19.
- Statheropoulos M., Vassiliadis N., Pappa A., 1997. Principal components and canonical correlation analysis for examining air pollution and meteorological data. *Atmospheric Environment*, Vol.32, No.6, 1087-95.
- Straszko J., 2000. Statystyczna analiza ciągłych pomiarów imisji. *Ochrona powietrza i problemy odpadów*, vol. 34., No.1, 3-10.
- Straszko J., Dziubakiewicz D., 2000. Sieci neuronowe w ocenie jakości powietrza. *Ochrona powietrza i problemy odpadów*, vol. 34., No.3, 87-93.
- Straszko J., Dziubakiewicz D., 2000. Dynamika skażenia powietrza atmosferycznego. *Ochrona powietrza i problemy odpadów*, vol. 35., No.1, 3-9.
- Tadeusiewicz R., Sieci neuronowe, Akademicka Oficyna Wydawnicza, Warszawa 1993.
- Unal Y., Incecik S., Borhan Y., Montes S., 2000. Factors influencing the variability of SO₂ concentrations in Istanbul. *J. of Air and Waste Manage. Assoc.*, Vol.50, 75-84.
- Waibel A., Hanazawa T., Hinton G., Shikano K., Lang K., 1989. Phoneme recognition using time-delay neural networks. *IEEE Transactions on acoustisc, speech and signal processsing*, Vol.37, No.3.
- Wal J.T., Janssen L.H.J.M, 2000. Analysis of spatial and temporal variations of PM₁₀ concentrations in the Netherlands using Kalman filtering. *Atmospheric Environment*, Vol.34, 3675-87.
- Watson D.F., Contouring; A guide to the analysis and display of spatial data, Pergamon Press, Oxford 1992.
- Woś A., Meteorologia dla Geografów, PWN, Warszawa 1996.
- Young P., 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Monitoring and Software*, Vol.13, 105-22.
- Yu T., Chang L., 2000. Selection of the scenarios of ozone pollution at southern Taiwan area utilizing principal component analysis. *Atmospheric Environment*, Vol.34, 4499-4509.
- Zalewski A., Cegiela R., Matlab-obliczenia numeryczne i ich zastosowanie, Nakom, Poznań 1996.
- Ziomas I., Melas D., Zerefos C., Bais A., 1995. Forecasting peak pollutant levels from meteorological variables. *Atmospheric Environment*, Vol. 29, No.24, 3703-11.
- Zwoździak J., Metody identyfikacji źródeł emisji pyłów i oceny ich oddziaływania, monogr., Praca naukowa Instytutu Inżynierii Ochrony Środowiska Politechniki Wrocławskiej, Wrocław 1986.
- Zwoździak J., Zwoździak A., Szczurek A., Meteorologia w ochronie atmosfery, Oficyna wyd. PWR, Wrocław 1998.

