



Mateusz Wiatrowski

Prognozowanie stężenia pyłów PM2.5 i
PM10 z wykorzystaniem danych
pogodowych

Prediction of PM2.5 and PM10 pollutant
concentration using weather data

Praca licencjacka

Promotor: dr Łukasz Wawrowski

Data przyjęcia:

Podpis promotora

Kierunek: Informatyka i ekonometria

Specjalizacja: Analityka gospodarcza

Poznań 2019

Spis treści

Wstęp	2
1 Zanieczyszczenia i ochrona powietrza atmosferycznego	3
1.1 Źródła i rodzaje zanieczyszczeń	3
1.2 Skutki zanieczyszczenia powietrza	6
1.2.1 Skład smogu i jego szkodliwy wpływ	6
1.2.2 Dziura ozonowa	9
1.2.3 Kwaśne deszcze	9
1.2.4 Globalne ocieplenie	9
1.3 Normy dotyczące maksymalnych wartości poszczególnych substancji w Europie i w Polsce	10
1.4 Technologie i procesy ochrony powietrza	12
1.5 Obecna sytuacja w Polsce	14
2 Metody prognozowania szeregów czasowych	16
2.1 Prognozowanie	16
2.2 Dekompozycja szeregów czasowych	17
2.3 Model Holta	18
2.4 Model Wintersa	19
2.5 Model k-najbliższych sąsiadów	22
2.6 Błędy prognoz	24
3 Prognozowanie stężenia pyłów PM2.5 i PM10	26
3.1 Przygotowanie danych do analizy	26
3.1.1 Źródła danych	26
3.1.2 Imputacja braków metodą knn	27

3.2	Eksploacyjna analiza danych	28
3.2.1	Wizualizacja zmiennych	28
3.2.2	Miary klasyczne i pozycyjne	30
3.3	Prognoza pyłu PM2.5	33
3.3.1	Prognoza Holta-Wintersa	33
3.3.2	Prognoza z wykorzystaniem regresji k najbliższych sąsiadów	36
3.3.3	Porównanie modeli	39
3.4	Prognoza pyłu PM10	39
3.4.1	Prognoza Holta-Wintersa	39
3.4.2	Prognoza z wykorzystaniem regresji k najbliższych sąsiadów	41
3.4.3	Porównanie modeli	44
	Podsumowanie	46
	Spis tabel	49
	Spis rysunków	51

Wstęp

Powietrze atmosferyczne jest jednym z elementów środowiska, które decyduje o jakości życia człowieka i jego otoczenia. Do powietrza trafia wiele substancji zanieczyszczających o bardzo zróżnicowanym potencjale wpływu na stan środowiska. Niekiedy dochodzi do sytuacji, w których stężenia zanieczyszczeń wielokrotnie przewyższają dopuszczalne normy, określone przez odpowiednie instytucje badawcze. Zbyt duże zanieczyszczenie powietrza powoduje szkodliwe oddziaływanie na zdrowie ludzi a także niekorzystne zmiany w jakości gleb, wód, szaty roślinnej oraz budynków.

W Polsce w ostatnich latach nastąpił wzmożony wzrost zainteresowania problemem zanieczyszczonego powietrza. Szczególnie w okresie jesienno-zimowym, czyli sezonie grzewczym rośnie ilość i częstotliwość informacji o wysokich stężeniach pyłów PM10 i PM2.5 oraz zjawisku tzw. smogu, który jest problemem nie tylko dużych, ale także i małych miast, a nawet wsi. Problemem są niekiedy bardzo wysokie stężenia zanieczyszczeń przekraczające wielokrotnie dopuszczalne normy, spowodowane głównie niską emisją.

Głównym celem niniejszej pracy jest zbudowanie modeli prognostycznych do tworzenia krótkookresowych prognoz dla pyłów PM2.5 i PM10. Następnie każdy z modeli został zbadany pod kątem precyzyjności obliczonych prognoz na podstawie miar ex post. Praca opiera się na danych dotyczących substancji w powietrzu zebranych przez GIOŚ (Główny Inspektorat Ochrony Środowiska) oraz na podstawie historycznych danych o pogodzie zgromadzonych przez IMGW (Instytut Meteorologii i Gospodarki Wodnej).

Praca składa się z trzech rozdziałów. Pierwszy rozdział wprowadza w tematykę zanieczyszczeń powietrza, opisuje substancje zawarte w powietrzu oraz ich wpływ

na środowisko i zdrowie ludzkie. Zawarte są w nim również informacje odnośnie ilości emitowanych zanieczyszczeń oraz ich źródeł. Opisano również technologie stosowane w związku z ograniczaniem emisji, dopuszczone normy i limity, a także w jaki sposób są nadzorowane i jak rozłożone są stacje w Polsce.

Rozdział drugi ma charakter teoretyczny i zostały w nim przedstawione podstawowe pojęcia związane z prognozowaniem szeregów czasowych. Szczególną uwagę poświęcono modelowi Holta-Wintersa oraz modelowi k-najbliższych sąsiadów. W ostatnim podrozdziale przedstawiono miary służące ocenie prognoz *ex post* i *ex ante*.

W rozdziale trzecim podjęto próbę przeprowadzenia prognozy pyłów PM_{2.5} i PM₁₀. Do realizacji tego celu niezbędne było prawidłowe przygotowanie danych, do czego został wykorzystany pakiet R. Do prognozowania wykorzystano metody opisane w rozdziale drugim. Uzyskane wyniki zostały ocenione z wykorzystaniem odpowiednich miar błędu prognoz.

Pracę kończy Podsumowanie, w którym zamieszczono najważniejsze wnioski z przeprowadzanych analiz.

Rozdział 1

Zanieczyszczenia i ochrona powietrza atmosferycznego

1.1 Źródła i rodzaje zanieczyszczeń

Powietrze atmosferyczne to jeden z elementów środowiska, który składa się z bezbarwnej i bezwonnej mieszaniny gazów, głównie z azotu (78%), tlenu (21%), dwutlenku węgla (0,03%), niewielkiej ilości gazów szlachetnych, pary wodnej i innych substancji określanych mianem zanieczyszczeń.

Zanieczyszczenie powietrza to wprowadzenie do powietrza substancji stałych ciekłych lub gazowych, w ilościach, które mogą wpłynąć szkodliwie na zdrowie człowieka, klimat, żywą przyrodę, glebę, wodę lub powodować inne szkody w środowisku (Pyłka-Gutowska, 2004, s. 93).

Zgodnie z obowiązującą ustawą Prawo Ochrony Środowiska (Ustawa z 27 kwietnia 2001) przez zanieczyszczenie rozumie się emisję, która może być szkodliwa dla zdrowia ludzi i stanu środowiska, może powodować szkodę w dobrach materialnych, pogarszać walory estetyczne środowiska. (Wielgosiński & Zarzycki, 2018, s. 25)

Na proces tworzenia zanieczyszczeń składa się wiele czynników. Wyróżniamy dwa podstawowe rodzaje zanieczyszczeń:

- pochodzenia naturalnego – powstałe na wskutek emisji do atmosfery szkodliwych substancji podczas pożarów lasów, wybuchów wulkanów, huraganów, procesów gnilnych na bagnach, burz piaskowych.
- pochodzenia antropogenicznego – to zanieczyszczenia wynikające z działalności człowieka – (pyły, gazy) powstałe w procesach przemysłowych, energetyce i ciepłownictwie, transporcie, przemyśle hutniczym, wydobywczym, chemicznym (Pyłka-Gutowska, 2004, s. 93).

W tabeli 1.1 przedstawiono najważniejsze zanieczyszczenia, które są emitowane do powietrza ze źródeł naturalnych i antropogenicznych.

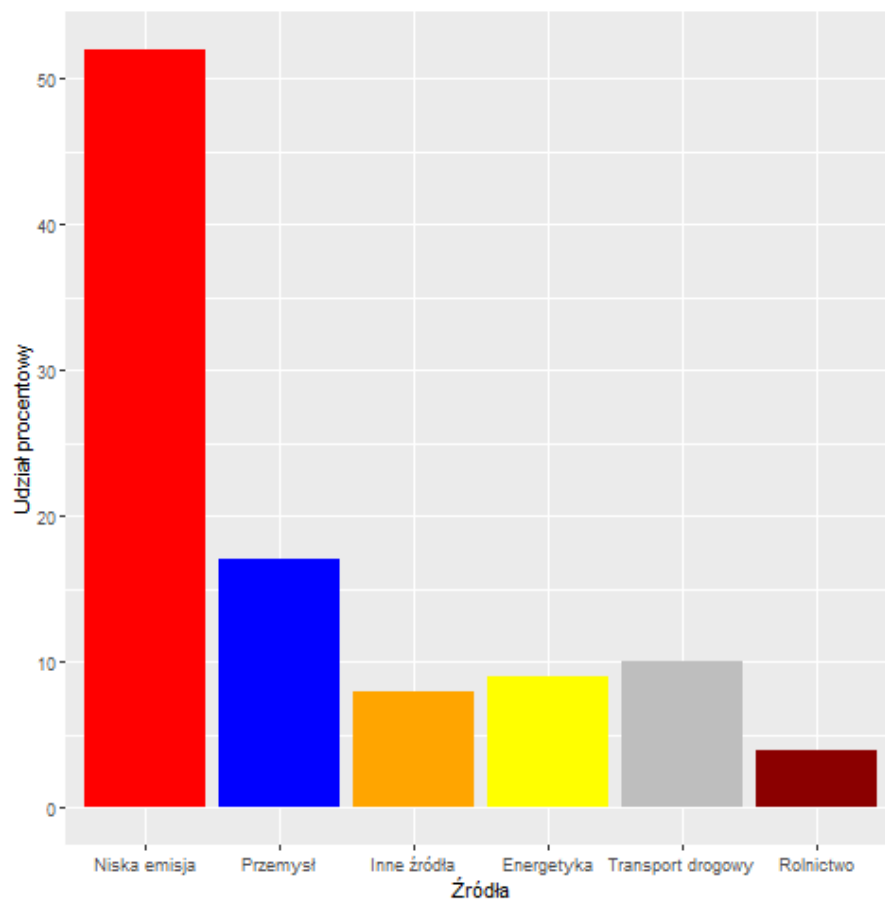
Tabela 1.1. Wielkości emisji ze względu na źródła

Składnik	Emisja ze źródeł naturalnych (Mg)	Emisja ze źródeł antropogenicznych (Mg)
Ditlenek węgla	$7,0 * 20^{10}$	$1,5 * 10^{10}$
Tlenek węgla	-	$2,0 * 10^8$
Ditlenek siarki	$1,42 * 10^8$	$1,3 * 10^8$
Tlenki azotu	$1,4 * 10^9$	$6,5 * 10^7$
Pył	$1,2 * 10^9$	$2,0 * 10^8$

Źródło: (Wielgosiński & Zarzycki, 2018, s. 56)

Z tabeli 1.1 wynika, że większość emitowanych przez człowieka substancji do powietrza jest mniejsza lub porównywalna z emisją pochodzenia naturalnego.

Źródła emisji pyłu w Polsce doskonale ilustruje Rysunek 1.1



Rysunek 1.1. Źródła emisji pyłu PM10 w Polsce

Źródło: opracowanie własne na podstawie (Mazurek, 2018, s. 13)

Głównym powodem występowania wysokiej zawartości pyłów w powietrzu są spaliny, które pochodzą z pieców i kotłów grzewczych zainstalowanych w gospodarstwach domowych (niska emisja). Problem ten jest spowodowany niskiej jakości węglem i paliwami kopalnianymi, ze zwiększoną ilością siarki w swoim składzie. Podczas spalania takich surowców uwalniane są do powietrza popioły i sadze, a także niebezpieczne dla zdrowia tlenki siarki, które wolno rozprzestrzeniają się w atmosferze z powodu dużego ciężaru właściwego (Polski Alarm smogowy, 2016).

1.2 Skutki zanieczyszczenia powietrza

1.2.1 Skład smogu i jego szkodliwy wpływ

Smog – to określenie, będące połączeniem dwóch wyrazów z języka angielskiego smoke (dym) i fog (mgła) – oznacza mieszaninę mgły z dymami i spalinami unoszącymi się w powietrzu (Mazurek, 2018, s. 8). Zjawisko to zachodzi gdy spalane paliwa kopalniane mieszają się z mgłą. Jest to szczególnie zauważalne, gdy nałożą się na siebie odpowiednie warunki atmosferyczne, bezwietrzna pogoda oraz brak deszczu. Zawieszone w ten sposób w powietrzu substancje nie mogą zostać rozproszone i skutecznie utrzymują się w miejscu w którym powstały.

Według (Mazurek, 2018, s.12) smog możemy podzielić na dwa rodzaje:

- Smog typu londyńskiego (zimowy) – najczęściej występuje w klimacie umiarkowanym w czasie sezonu grzewczego od października do marca, kiedy potrzebna jest energia do uzyskania ciepła. Najczęściej w tym celu wykorzystywane są paliwa kopalniane, które w procesie spalania emitują szkodliwe substancje. Na jego zawartość składają się pyły (PM10, PM2.5), sadze, tlenki węgla, tlenki siarki oraz tlenki azotu. Jest szczególnie niebezpieczne gdy panuje wysoki poziom wilgotności powietrza, zawarte w dymie tlenki siarki reagują, następuje wytworzenie się kwasu siarkowego, który jest parzący i może uszkodzić drogi oddechowe.
- Smog typu Los Angeles (fotochemiczny) – występuje w miastach w których panuje duży ruch uliczny i szczególne nasłonecznienie, z reguły na niskich szerokościach geograficznych. Pojawia się pomiędzy majem a wrześniem. Spaliny wydobywające się z pojazdów zawierają tlenki azotu, tlenki węgla i pyły. Zanieczyszczenia unoszą się nad ulicami, dodatkowo przy znacznym nasłonecznieniu powstaje reakcja fotochemiczna wyzwalamąca ozon.

Analiza literatury wskazuje, że autorzy wyróżniają następujące podstawowe składniki smogu (Mazurek, 2018, s. 9)

1. Pyły zawieszone (cząstki stałe)

- (a) Grube (pyły o średnicy aerodynamicznej 2.5-10 μm) – PM10

(b) Drobnie (pyły o średnicy aerodynamicznej $\leq 2.5 \mu\text{m}$) – PM_{2.5}

(c) Ultradrobne (pyły o średnicy aerodynamicznej $\leq 0,5 \mu\text{m}$) – PM_{0,1}

2. Substancje gazowe

(a) Tlenki siarki (SO_x)

(b) Tlenki azotu (NO_x)

(c) Ozon (O₃)

(d) Tlenki węgla (CO₂ i CO)

3. Półlotne związki organiczne

(a) Węglowodory aromatyczne (np. Benzo(a)piren)

(b) Dioksyny

(c) Benzen

(d) Aldehydy

Charakterystyka substancji:

PM₁₀ – Powstaje w procesie spalania, procesach produkcyjnych, podczas wyburzeń budynków. Dotyczy również pyłków rozpylanych przez niektóre rośliny i zarodniki grzybów. Czas utrzymywania się w powietrzu jest dość krótki, nie przekracza kilku godzin. Może zostać przeniesiony na krótkie dystanse.

PM_{2.5} – jest następstwem spalania paliw kopalnianych i drewna. Powstaje również przy pracach rolnych, transporcie, w zakładach przemysłowych oraz procesach technologicznych. Czas utrzymywania się w powietrzu jest dość długi, może przenosić się na spore odległości. Podczas przenoszenia ulega rozproszeniu, tym samym obniża się jego stężenie.

Wpływ szkodliwych substancji na zdrowie, znajdujących się w smogu nie został jeszcze w pełni zbadany, tzn. nie udało się określić wartości progowych, kiedy pojawiają się pierwsze skutki zdrowotne. Przez co należy uważać, że każda ilość występujących w powietrzu szkodliwych substancji jest narażeniem na zdrowie,

a im wyższe są wartości wskaźników tym jest to groźniejsze. Zanieczyszczenia zawarte w smogu najbardziej szkodzą na układ oddechowy. W górnej części dróg oddechowych najczęstszymi dolegliwościami wysokiej ekspozycji na zanieczyszczenia jest uczucie suchości błony śluzowej nosa i jamy ustnej, co prowadzi do podrażnienia gardła, krtani, chrypy. Dłuższe narażenie na wysokie stężenia szkodliwych substancji może prowadzić do wystąpień stanów zapalnych. Niestety po przestaniu bycia narażonym na zagrożenie, stan zapalny może utrzymywać się przez długi czas. W tabeli 1.2 przedstawiono wpływ pyłów poszczególnych frakcji cząstek na człowieka.

Tabela 1.2. Oddziaływanie pyłów na zdrowie człowieka

Wielkość cząstek	Oddziaływanie na człowieka
PM10	Indukują objawy ze strony układu oddechowego(kaszel, świsty, duszności)
	Wywołują zaostrzenia chorób przewlekłych (astmy, POChP)
	Powodują niższe wskaźniki czynnościowe płuc
PM2.5	Wywołują stany zapalne m.in. zatok przynosowych
	Powodują mniejszą masę urodzeniową noworodków
	Wpływają na zwiększoną częstość hospitalizacji z przyczyn oddechowych lub krążeniowych
	Zwiększają ryzyko zgonu
PM0.1	Przenikają do układu krążenia i stymulują zapalenie systemowe oraz pobudzenie układu współczulnego
	Indukują zapalenie i stres oksydacyjny, prowadząc do chorób układu krążenia

Źródło: (Mazurek, 2018, s. 27)

Wielopierścieniowe węglowodory aromatyczne(WWA) zawarte w pyłach są związkami chemicznymi, których budulec stanowi węgiel i wodór, zawierają w cząsteczce kilka pierścieni aromatycznych. Pojawiają się w wyniku spalania paliw kopalnych. W tabeli 1.3 przedstawione zostały własności wielopierścieniowych węglowodórów aromatycznych (Pyłka-Gutowska, 2004, s. 96).

Tabela 1.3. Pochodzenie i oddziaływanie Benzo(a)pirenu

Pochodzenie	Jest to produkt niepełnego spalania (węgla, paliw silnikowych, drewna, tytoniu, śmieci itd.)
	W Polsce powstaje głównie w kotłowniach przydomowych, ale także podczas grillowania czy wędzenia
Działanie	Ma silne właściwości rakotwórcze
	Rozwój nowotworu następuje najwcześniej po ok.15 latach od pierwszego kontaktu z istotnymi ilościami tej substancji
	Wzmacnia działanie innych substancji

Źródło: (Mazurek, 2018, s. 28)

1.2.2 Dziura ozonowa

Dziura ozonowa jest miejscem w atmosferze, gdzie warstwa ozonowa jest bardzo cienka lub nie występuje. Warstwa ozonowa jest częścią atmosfery ziemskiej, która ma w sobie najwięcej ozonu, chroni Ziemię przed szkodliwym promieniowaniem słonecznym. Braki w warstwach mogą być katastrofalne w skutkach dla Ziemi, ponieważ mogą powodować topnienia lodowców (Wielgosiński & Zarzycki, 2018).

1.2.3 Kwaśne deszcze

Kwaśne deszcze są to opady atmosferyczne, których odczyn pH jest mniejszy niż 5,6. Pojawiają się gdy związki siarki uwalniane podczas spalania zasilanych paliw kopalnych łączą się z wodą lub parą wodną w atmosferze. Kwaśne deszcze powodują znaczne zniszczenia w środowisku. Są przyczyną zakwaszeń w glebie, korozji w metalach, niszczeniem plonów. Obecnie w Europie dzięki rygorystycznym normom i ograniczeniom mającym na celu zmniejszenie emisji tlenków siarki do atmosfery udało się zminimalizować liczbę ich wystąpień. (Wielgosiński & Zarzycki, 2018, s. 61)

1.2.4 Globalne ocieplenie

Globalnym ociepleniem nazywamy długotrwały wzrost średniej temperatury na Ziemi, zmianami klimatycznymi wykazywanymi przez pomiary temperatur. Termin ten odnosi się do obserwowanego ocieplenia od czasów przedindustrialnych

i jego przewidywanej kontynuacji, chociaż było również dużo wcześniejszych okresów globalnego ocieplenia. Zmiany klimatyczne obejmują zarówno globalne ocieplenie, jak i jego skutki, takie jak zmiany opadów i wpływy, różniące się w zależności od regionu (Wielgosiński & Zarzycki, 2018, s. 72-85).

Globalne ocieplenie jest spowodowane zwiększoną obecnością gazów cieplarnianych w atmosferze do których należy 5 gazów: ditlenek węgla powstający we wszelkiego rodzaju procesach spalania, metan powstający w wyniku spalania gazów (głównie ziemnego) oraz podtlenek azotu wytwarzany przez rolnictwo, dodatkowo zaliczamy freony CFCs, ozon. Największe znaczenie ma także para wodna. W tabeli 1.4 przedstawiono udziały wyżej wymienionych gazów cieplarnianych. Ich wartości są zależne od metodyki i sposobu liczenia. Według różnych obliczeń mogą przyjmować inne wartości (Wielgosiński & Zarzycki, 2018, s. 72-85).

Tabela 1.4. Udział gazów w globalnym ociepleniu

Gaz	Przedział procentowy udziału w globalnym ociepleniu
Ditlenek węgla (CO ₂)	9-26
Metan (CH ₄)	4-18
Chlorowęglowodory (freony CFCs)	3-7
Ozon (O ₃)	
Podtlenek azotu (N ₂ O)	
Para wodna	33-66

Źródło: Opracowanie własne na podstawie (Wielgosiński & Zarzycki, 2018, s. 76-77)

1.3 Normy dotyczące maksymalnych wartości poszczególnych substancji w Europie i w Polsce

Unia Europejska nie wyznacza żadnych norm dotyczących maksymalnych wartości dotyczących pyłów PM₁₀ i PM_{2.5}. Jednakże według Światowej Organizacji Zdrowia (WHO) zalecane są wartości graniczne, które zostały przedstawione w tabeli 1.5

Tabela 1.5. Wartości norm dla pyłów PM10 i PM2.5 według WHO

Rodzaj pyłu	Okres uśredniania	Wartość ($\mu\text{g}/\text{m}^3$)
PM10	1 doba	50
	1 rok	20
PM2.5	1 doba	25
	1 rok	10

Źródło: (Mazurek & Badyda, 2018, s. 32)

W związku z brakiem norm ustalonych dla Unii Europejskiej każdy kraj może ustalić swoje własne. Wartości ustalone w Polsce są najmniej restrykcyjne spośród innych państw Unii. Wartości norm w Polsce zamieszczone są w tabeli 1.6.

Tabela 1.6. Normy prawne dla substancji występujących w powietrzu w Polsce

Substancja	Okres uśredniania wyników	Dopuszczalny poziom ($\mu\text{g}/\text{m}^3$)	Dopuszczalna liczba przekroczeń w ciągu roku(dni)
Pył PM10	Doba	50	35
	Rok	40	-
Pył PM2.5	Rok	25	-
Dwutlenek siarki(SO ₂)	1 godzina	350	24
	24 godziny	125	3
Tlenek węgla (CO)	8 godzin	10000	-
Dwutlenek azotu (NO ₂)	1 godzina	200	18
	Rok	40	-
Benzen (C ₆ H ₆)	Rok	5	-
Ołów (Pb)	Rok	0,5	-

Źródło: (ARMAAG, b.d.)

W tabeli 1.6 wyszczególnione są substancje dla różnych okresów uśredniania. Wyróżniamy w nich trzy, których częstość przekraczania odnosi się do poziomu dopuszczalnego wraz z marginesem tolerancji, wprowadzonych przez program ochrony powietrza, o których mowa w art. 91 (Ustawa z 27 kwietnia 2001).

Tabela 1.7. Poziom docelowy dla niektórych substancji występujących w powietrzu

Substancja	Okres uśredniania wyników pomiaru	Poziom docelowy substancji w powietrzu	Dopuszczalna liczba przekroczeń w ciągu roku(dni)
Arsen	Rok	6 ng/m ³	-
Benzo(a)piren	Rok	1 ng/m ³	-
Kadm	Rok	5 ng/m ³	-
Nikiel	Rok	20 ng/m ³	-
Pył zawieszony PM2.5	Rok	25 $\mu\text{g}/\text{m}^3$	25
Ozon	Osiem godzin	120 $\mu\text{g}/\text{m}^3$	-
	rok	18 000 $\mu\text{g}/\text{m}^3$	-

Źródło: (ARMAAG, b.d.)

W tabeli 1.7 zostały przedstawione docelowe poziomy substancji występujących w powietrzu w Polsce. Wprowadzono je w celu ochrony ludzkiego zdrowia i środowiska a także ograniczania ich szkodliwego działania, w określonym czasie.

Tabela 1.8. Poziomy pyłu PM10 ze względu na zagrożenie

Poziom	Dobowe stężenie PM10($\mu\text{g}/\text{m}^3$)
Dopuszczalny	50
Ostrzegania	200
Alarmowy	300

Źródło: (Mazurek, 2018, s. 15)

W tabeli 1.8 zostały określone poziomy zagrożeń ze względu na zawartość dobowych stężeń pyłu PM10 w powietrzu. Poniżej zostały szczegółowo opisane.

- poziom dopuszczalny - wartość ustalona na podstawie wiedzy naukowej, mający na celu ochronę ludzkiego zdrowia i środowiska, a także ograniczania ich szkodliwego działania, w określonym czasie.
- poziom ostrzegania - wartość powyżej której zdrowie jest narażone, nawet podczas krótkotrwałego przebywania w obszarze zanieczyszczonego powietrza, w szczególności osób wrażliwych. Konieczne jest przekazanie wiadomości możliwie jak największej liczbie osób.
- poziom alarmowy - wartość która wskazuje na realne zagrożenie zdrowia, w wyniku krótkotrwałego narażenia na zanieczyszczone powietrze. Wymaga natychmiastowego działania ze strony państwa.

1.4 Technologie i procesy ochrony powietrza

Mając na uwadze zdrowie i świadomość zagrożeń płynących ze stosowanych źródeł pozyskiwania energii, powinniśmy ograniczać emisję szkodliwych substancji do atmosfery. Do redukcji emisji zanieczyszczeń powinien przyczyniać się każdy człowiek, wybierając ekologiczne metody ogrzewania domu czy posługiwać się poprawnymi technikami palenia i stosować najwyższej jakości paliwa. Wielki wpływ powinny mieć także bardziej rygorystyczne regulacje prawne. Zmuszałyby one do polepszania się sytuacji na rynku paliw kopalnych i minimalizacji emitowania zanieczyszczeń. Jednocześnie państwo powinno stworzyć możliwości finansowania

dla bardziej ekologicznych rozwiązań, tak aby stworzyć szanse zmiany również dla najuboższych mieszkańców, zachęcając tym sposobem na zmianę sposobu ogrzewania mieszkań. W tabeli 1.9 przedstawione zostały przykładowe sposoby ograniczania emisji zanieczyszczeń

Tabela 1.9. Ograniczanie emisji zanieczyszczeń - rozwiązania systemowe

Gałąź Gospodarki	Sposoby ograniczania emisji zanieczyszczeń
Przemysł	Monitorowanie i ograniczanie emisji
	Budowanie zakładów poza dolinami i centrami miast
	Wysokie kominy
	Stosowanie skutecznych filtrów
Urbanizacja	Zmniejszenie gęstości zabudowy
	Dbłość o przewietrzanie miast (brak zabudowy w miejscach ciągów mas powietrza)
	Zachowanie i zwiększanie powierzchni obszarów zielonych
Transport drogowy	Wprowadzenie standardów emisji spalin
	Stosowanie paliw o ograniczonej zawartości siarki
	Zmiana paliw na bardziej ekologiczne
	Wymóg stosowania filtrów cząstek stałych (także w starszych samochodach)
	Badanie składu spalin jako element kontroli pojazdów
	Ograniczanie ruchu pojazdów w centrach miast, rozwój komunikacji zbiorowej
Ciepłownictwo	Ograniczenie używania paliw niskiej jakości wyłącznie do elektrociepłowni wyposażonych w skuteczne filtry
	Wymiana pieców na ekologiczne
	Stosowanie paliw o niskiej zawartości siarki
	Edukacja w zakresie poprawnych technik palenia
	Egzekwowanie zakazu spalania śmieci (wysokie kary)
	Docieplenie mieszkań

Źródło: (Mazurek, 2018, s. 45)

1.5 Obecna sytuacja w Polsce

W Polsce badaniem i oceną jakości powietrza zajmuje się Inspekcja Ochrony Środowiska, składająca się z 17 inspektoratów wojewódzkich. obejmujące swym obszarem wyznaczone strefy. Do analizy jakości powietrza wykorzystywane są substancje wyszczególnione w podrozdziale (1.2.1). W 2010 roku nastąpiło wydzielenie 46 stref, mających na celu ochronę zdrowia człowieka. Strefy dzielą się ze względu na liczbę mieszkańców, według następującego podziału:

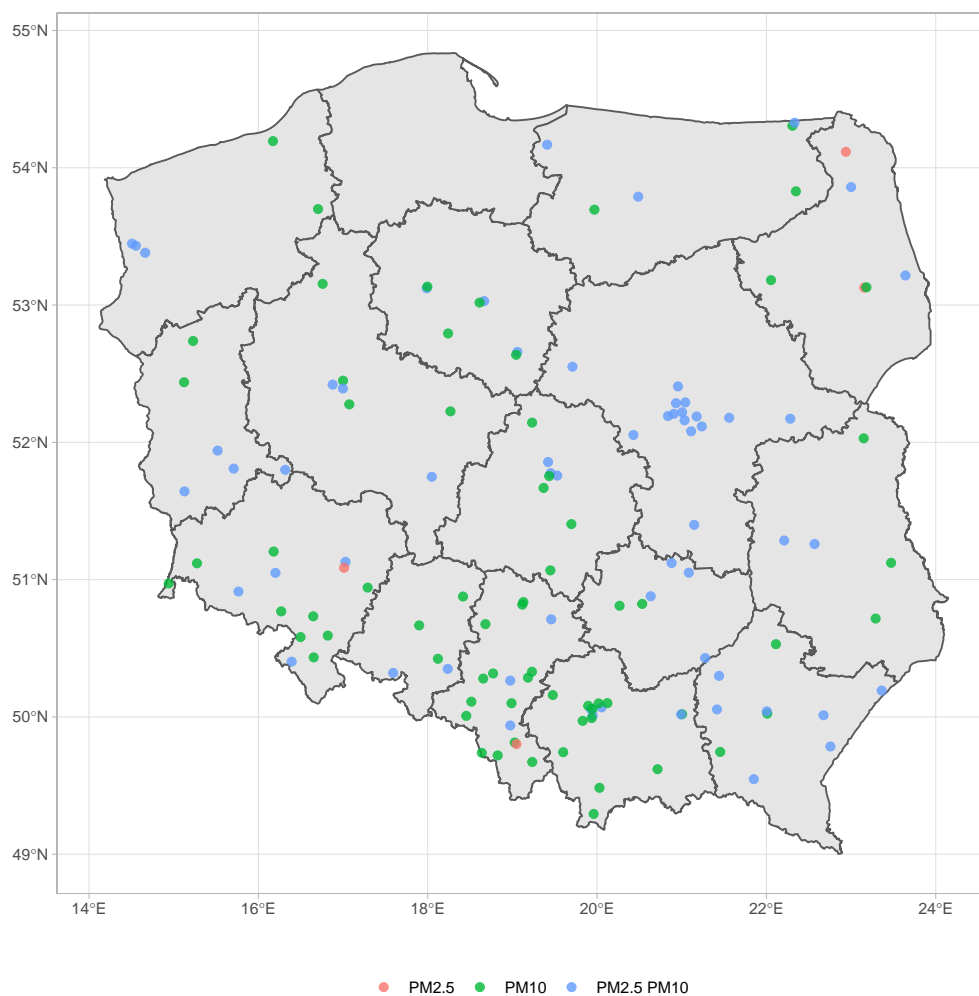
- aglomeracje miejskie (miasta o liczbie mieszkańców przekraczającej 250 tys. mieszkańców),
- miasta nie będące aglomeracjami (miasta o liczbie mieszkańców powyżej 100 tys. mieszkańców),
- pozostałe obszary województw.

Na podstawie analiz dokonywanych przez Inspektoraty przyporządkowuje się strefy do odpowiednich klas (GIOŚ, 2018):

- klasa A – poziom stężeń nie przekracza poziomu dopuszczalnego/docelowego,
- klasa B – poziom stężeń zanieczyszczenia przekracza poziom dopuszczalny, lecz nie przekracza poziomu dopuszczalnego powiększonego o margines tolerancji,
- klasa C - poziom stężeń zanieczyszczenia przekracza poziom dopuszczalny,
- klasa D1 - poziom stężeń zanieczyszczenia nie przekracza poziomu celu długoterminowego (dotyczy tylko ozonu),
- klasa D2 - poziom stężeń zanieczyszczenia przekracza poziom celu długoterminowego (dotyczy tylko ozonu),

Do pomiarów stężeń pyłów PM₁₀ i PM_{2.5} w Polsce wykorzystywane są specjalnie przeznaczone do tego stacje, których rozmieszczenie przedstawione jest na

rysunku 1.2. Zgodnie z rysunkiem wyróżniamy 68 stacji mierzących tylko zawartość pyłu PM10, 4 PM2.5, i 62 stacje na których mierzone są oba pyły jednocześnie. W województwie Pomorskim zauważa się braki punktów pomiarowych.



Rysunek 1.2. Rozmieszczenie stacji pomiarowych dla pyłów PM10 i PM2.5

Źródło: Opracowanie własne na podstawie danych GIOŚ

Rozdział 2

Metody prognozowania szeregów czasowych

2.1 Prognozowanie

Prognozowanie to przewidywanie przyszłych zdarzeń na podstawie zebranych wcześniej doświadczeń. Najczęściej dotyczą danych opartych na szeregach czasowych, wtedy prognozujemy przyszłe wartości zdarzeń, ale mogą być także oparte na danych przekrojowych (Gajda, 2017).

Powstałe prognozy możemy oceniać w kategoriach jako trafne i nietrafne. Pierwsze z nich są to przewidywania, które w niewielkim stopniu różnią się od wartości zrealizowanych zmiennych. Natomiast nietrafne dotyczą prognoz, które znacząco odbiegają od wielkości prognozowanej (Gajda, 2017).

Prognozowanie dotyczące zmiennych ilościowych można przeprowadzić na kilka sposobów:

- Prognozy punktowe - wyznaczona zostaje odpowiednia wartość, określona jako zmienna prognozowana w przyszłości.
- Prognoza przedziałowa, określona za pomocą przedziału liczbowego, w którym zawiera się wartość zmiennej prognozowanej. W zależności od budowy, mogą to być przedziały ufności lub przedziały predykcji.
- Prognoza wariantowa, określona jako zbiór pewnych wartości, które mogą wystąpić w okresie prognozy (Dittman, 2008, s.21).

2.2 Dekompozycja szeregów czasowych

Szeregiem czasowym nazywamy obserwacje mierzalnego zjawiska w kolejnych odstępach czasowych lub okresach. W pierwszej kolumnie przedstawione zostają momenty lub okresy czasu, a w drugiej wartości zjawisk zaobserwowanych w tych jednostkach czasu. Schemat szeregu czasowego został zobrazowany w tabeli 2.1, stanowi ona podstawę do bardziej szczegółowych analiz statystycznych (Starzyńska, 2012).

Tabela 2.1. Ogólna postać szeregu czasowego

Czas (t)	Zjawisko (y_t)
t_1	y_1
t_2	y_2
.	.
.	.
t_n	y_n

Źródło: (Starzyńska, 2012, s.95)

Zebrane zdarzenia mogą ulegać pewnym prawidłowościom, których odkrycie jest możliwe dzięki właściwej analizie statystycznej. Dekompozycja szeregu czasowego polega na wyodrębnieniu, identyfikacji i opisie dwóch składowych:

- składowej systematycznej na którą wpływ mają stałe czynniki na zmienną prognozowaną. Może występować w postaci stałego przeciętnego poziomu zmiennej prognozowanej, trendu, składowej okresowej (wahania sezonowe i cykliczne)
- składowej przypadkowej, będącej efektem oddziaływań przyczyn ubocznych.

Opis składowych systematycznych:

- Trend - występuje w postaci wzrastających lub malejących wartości zmiennej prognozowanej. Kształtowany jest przez stały zestaw czynników, aby można go było wyznaczyć potrzeba długiego czasu obserwacji.
- Stały przeciętny poziom - wartości zmiennej prognozowanej oscylują przez całą długość ciągu obserwacji na podobnym stałym poziomie.

- Wahania sezonowe - są to wahania wartości zmiennej prognozowanej w obrębie trendu lub stałego poziomu. Powtarzają się w przedziale czasu, który nie przekracza jednego roku. Najczęściej są efektem zmian pór roku lub związane z przyzwyczajeniami i wydarzeniami cyklicznymi.
- Wahania cykliczne - to powtarzające się rytmicznie, trwające dłużej niż rok wahania o charakterze długookresowym i nieregularnym, posiadające różną długość cyklu. W ekonomii najczęściej związane są z cyklem koniunkturalnym. Ich analiza nie należy do łatwych, przewidywania prognozowanych wartości i czas wystąpienia recesji czy ożywienia jest ciężkie do obliczenia, ich modele obarczone są większymi błędami niż przy wahaniami sezonowych (Cieślak, 2005, s.64).

2.3 Model Holta

Model Holta jest przykładem modelu z wykorzystaniem wygładzania wykładniczego polegającym na tym, że uwzględniamy w nim występujące tendencje rozwojowe i wahania przypadkowe. Przy wyznaczaniu opisu jego trendu korzysta się z wielomianu stopnia pierwszego. Opisywany jest przez następujące równania:

$$F_{t-1} = \alpha y_{t-1} + (1 - \alpha)(F_{t-2} + S_{t-2}) \quad (2.1)$$

$$S_{t-1} = \beta (F_{t-1} - F_{t-2}) + (1 - \beta)S_{t-2} \quad (2.2)$$

Gdzie:

F_{t-1} - wygładzona wartość zmiennej prognozowanej okres t-1,

S_{t-1} - wygładzona wartość przyrostu trendu na okres t-1,

α - parametr opisujący wpływ danych historycznych na prognozę, zawiera się w przedziale [0,1],

β - parametr opisujący wpływ w jakim najnowsze trendy powinny być zestawiane ze starszymi, zawiera się w przedziale [0,1].

Równanie prognozy w okresie $t > n$ przyjmuje postać:

$$y_t^* = F_n + (t - n)S_n, \quad t > n \quad (2.3)$$

gdzie:

y_t^* - prognoza zmiennej Y wyznaczona na okres t ;

F_n - wygładzona wartość zmiennej prognozowanej dla okresu n ;

S_n - wygładzona wartość przyrostu trendu na okres n ;

n - liczba wyrazów szeregu czasowego zmiennej prognozowanej.

Dobierając wartości początkowe dla wartości F i S , możemy skorzystać z wielu sposobów. Jedną z kilku dostępnych możliwości jest przyjęcie za F_1 pierwszej wartości zmiennej prognozowanej y_1 w przypadku S_1 - różnicy $y_2 - y_1$. Innym ze sposobów jest przyjęcie F_1 jako wyrazu wolnego i S_1 jako współczynnika kierunkowego liniowej funkcji trendu obliczonego na podstawie próbki wstępnej. Kolejnym napotkanym problemem jest wybranie wartości początkowych dla współczynników α i β . Ich miara jest najczęściej obliczana za pomocą metody prób i błędów, wybrane zostają te wartości które minimalizują średni błąd prognoz wygasłych, określone za pomocą wzoru 2.4 (Cieślak, 2005, s.73-74).

$$y_t^* = F_{t-1} + S_{t-1}, \quad 2 \leq t \leq n \quad (2.4)$$

2.4 Model Wintersa

Model Wintersa jest przykładem modelu z wykorzystaniem wygładzania wykładniczego polegającym na tym, że uwzględniamy w nim występujące tendencje rozwojowe i wahania przypadkowe a dodatkowo wahania sezonowe. Przy wyznaczaniu opisu jego trendu korzysta się z wielomianu stopnia pierwszego (Cieślak, 2005, s.74).

Dane, które zostają poddane dekompozycji mogą posiadać sezonowość addytywną lub multiplikatywną. Model Wintersa uwzględnia te dwie możliwości i posiada dwa odrębne mechanizmy umożliwiające dekompozycje ze względu na rodzaj sezonowości która występuje.

W przypadku sezonowości addytywnej zakładamy, że szereg czasowy może być przedstawiony za pomocą modelu:

$$X_t = b_1 + b_2t + s_t + Y_t \quad (2.5)$$

gdzie:

b_1 - podstawowy sygnał (komponent stały),

b_2 - trend,

s_t - sezonowość (w tym przypadku addytywna),

Y_t - wektor reszt (szum).

Stosujemy go w przypadku, gdy amplituda sezonowości jest niezależna od średniego poziomu szeregu (Machura, b.d.).

Równania (2.6-2.7) opisują addytną wersję modelu Wintersa.

$$F_{t-1} = \alpha(y_{t-1} - C_{t-1-r}) + (1 - \alpha)(F_{t-2} - S_{t-2}) \quad (2.6)$$

$$S_{t-1} = \beta(F_{t-1} - F_{t-2}) + (1 - \beta)S_{t-2} \quad (2.7)$$

$$C_{t-1} = \gamma(y_{t-1} - F_{t-2}) + (1 - \gamma)C_{t-1-r} \quad (2.8)$$

gdzie:

F_{t-1} - wygładzona wartość zmiennej prognozowanej na okres t-1, po wykluczeniu wahań sezonowych,

S_{t-1} - wygładzona wartość przyrostu trendu na okres t-1,

C_{t-1} - ocena wskaźnika sezonowości na okres t-1,

α - parametr opisujący wpływ danych historycznych na prognozę, zawiera się w przedziale [0,1],

β - parametr opisujący wpływ w jakim najnowsze trendy powinny być zestawiane ze starszymi, zawiera się w przedziale [0,1];

γ - parametr opisujący wpływ sezonowości na prognozy, jego wartość zawiera się w przedziale [0,1].

Równanie 2.9 opisuje prognozę na okres $t > n$ (Cieślak, 2005, s.74-75)

$$y_t^* = F_n + S_n(t - n) + C_{t-r} \quad (2.9)$$

Natomiast w przypadku sezonowości multiplikatywnej zakładamy że szereg czasowy może być przedstawiony za pomocą modelu 2.10:

$$X_t = (b_1 + b_2 t)s_t + Y_t \quad (2.10)$$

gdzie:

b_1 - podstawowy sygnał,

b_2 - trend,

s_t - sezonowość (w tym przypadku multiplikatywna),

Y_t - wektor reszt.

Stosujemy go w przypadkach, w których amplituda sezonowości jest proporcjonalna do średniego poziomu szeregu (Machura, b.d.).

Równania (2.11-2.13) opisują multiplikatywną wersję modelu Wintersa:

$$F_{t-1} = \alpha \frac{y_{t-1}}{C_{t-1-r}} + (1 - \alpha)(F_{t-2} + S_{t-2}) \quad (2.11)$$

$$S_{t-1} = \beta(F_{t-1} - F_{t-2}) + (1 - \beta)S_{t-2} \quad (2.12)$$

$$C_{t-1} = \gamma \frac{y_{t-1}}{F_{t-1}} + (1 - \gamma)C_{t-1-r} \quad (2.13)$$

gdzie:

F_{t-1} - wygładzona wartość zmiennej prognozowanej na okres $t-1$, po dokonaniu redukcji wahań sezonowych;

S_{t-1} - wygładzona wartość przyrostu trendu na okres $t-1$;

C_{t-1} - ocena wskaźnika sezonowości na okres $t-1$;

α, β, γ - parametry modelu o wartościach z przedziału $[0,1]$.

Równanie prognozy w chwili t lub okresie $t > n$ przyjmuje postać:

$$y_t^* = [F_n + S_n(t - n)]C_{t-r} \quad (2.14)$$

Wartości początkowe dla modeli Wintersa w obu wersjach kształtują się następująco:

- F_1 - pierwsza wartość zmiennej prognozowanej, inną możliwością jest średnia z wartości zmiennej w pierwszym cyklu,
- S_1 - jako różnica drugiej i pierwszej wartości zmiennej prognozowanej, ewentualnie różnica średnich wartości zmiennej wyznaczonych dla drugiego i pierwszego cyklu,
- C_1 - wyznaczoną na podstawie szeregu czasowego średnią różnic (w przypadku modelu z wersją addytywną) lub ilorazów (w przypadku modelu z wersją multiplikatywną),
- α, β, γ - Obliczane za pomocą metody prób i błędów, wybrane zostają te wartości, które minimalizują średni błąd prognoz wygasłych, dla prognoz z wyprzedzeniem jednookresowym wyprzedzeniem (Cieślak, 2005, s.75).

2.5 Model k-najbliższych sąsiadów

Metoda k-najbliższych sąsiadów jest prostym algorytmem, polegającym na klasyfikacji nowego obiektu na podstawie miary podobieństwa (np. funkcji odległości). Model ten został użyty w statystycznych szacowaniach i rozpoznawaniu wzorców już na początku lat 70. jako technika nieparametryczna, niewymagająca założeń (Sayad, 2010).

Istnieje kilka metod pomiaru odległości pomiędzy dwoma punktami. Zostały przedstawione we wzorach (2.15 - 2.18).

1. Miara Euklidesowa w postaci

$$D(x, p) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2.15)$$

2. Kwadrat miary Euklidesowej

$$D(x, p) = (x - p)^2 \quad (2.16)$$

3. Manhattan(miejska)

$$D(x, p) = |x - p| \quad (2.17)$$

4. Miara Czebyszewa

$$D(x, p) = \text{Max}(|x - p|) \quad (2.18)$$

Wyznaczanie predykcji sprowadza się do określenia K(liczby najbliższych sąsiadów) W przypadku regresji obliczamy średnią dla wybranego K:

$$y = \frac{1}{K} \sum_{i=1}^K y_i \quad (2.19)$$

Metoda K-najbliższych sąsiadów stosuje model pamięciowy na podstawie danych treningowych, przechwytuje informacje o wszystkich przypadkach szkoleniowych i klasyfikuje nowe przypadki na podstawie podobieństwa. W zależności od rodzaju zmiennych możemy przeprowadzić regresję dla danych ciągłych, i klasyfikację dla danych skategoryzowanych (StatSoft, 2019).

Walidacja krzyżowa jest sposobem umożliwiającym znalezienie optymalnej wartości K. Metoda ta polega na losowym podziale oryginalnego zestawu danych na v podzbiorów o równych lub zbliżonych rozmiarach. Następnie każdy z tych podzbiorów v jest traktowany jako testowy, a reszta czyli v-1 należy do zbioru uczącego się, wykonującego analizę. Obliczamy sumę kwadratów reszt dla każdej możliwości, uśredniamy wyniki i otrzymujemy miarę jakości modelu przy wybranym K. Szukając optymalnej wartości K musimy opierać się na metodzie eksploracyjnej podstawiając pod nią różne wartości aż do uzyskania najlepszej jakości modelu (Schutt & O'Neil, 2013, s. 62-63).

Innym sposobem obliczeń jest wprowadzenie wag. Metoda K-najbliższych sąsiadów zakłada, że obiekty znajdujące się w pobliżu są podobne. Powinniśmy więc wprowadzić każdej ze zmiennej k najbliższych sąsiadów odpowiednie wagi. Im bliżej wybranego punktu się znajdują tym większy będą miały wpływ. Wprowadzono więc wagę W, aby móc określić wpływ odległości (Samworth i in., 2012).

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^K \exp(-D(x, p_i))} \quad (2.20)$$

gdzie:

$D(x, p_i)$ jest miarą odległości pomiędzy nowym punktem x , a punktem p_i

Określone wagi w równaniu 2.20 spełniają warunek 2.21

$$\sum_{i=1}^K W(x_0 - x_i) = 1 \quad (2.21)$$

2.6 Błędy prognoz

Wszystkie prognozy zbudowanych modeli predykcji powinny zakładać mierniki rzędu dokładności, dodatkowo każdy z nich powinien dążyć do osiągnięcia jak najwyższej efektywności precyzji. Istnieją dwa mierniki służące do określania predykcji (Zeliaś, Pawełek & Wanat, 2008):

- ex ante - określająca dokładność prognozy, przed upływem czasu, na który prognoza była ustalona,
- ex post - określająca trafność prognozy, po czasie na który prognoza była ustalona (w przypadku znanej zmiennej realizowanej).

Do przeprowadzenia interpretacji błędów prognoz posłużą wskaźniki absolutne i względne. Do grupy najczęściej stosowanych wchodzących w skład miar absolutnych należą przedstawione we wzorach (2.22-2.26) (Szmuksta-Zawadzka & Zawadzki, 2012)

- ME (ang. Mean Error) - średni błąd predykcji, jej wartość powinna być równa zero lub blisko zero,

$$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - Y_{tp}) \quad (2.22)$$

- MAE (ang. Mean Absolute Error) - średni błąd absolutny, sprawdza jak rzeczywiste realizacje zmiennej prognozowanej będą się odchyłać co do wartości bezwzględnej od prognoz,

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - Y_{tp}| \quad (2.23)$$

- MSE (ang. Mean Squared Error) - błąd średniokwadratowy, informuje o ile średnio rzeczywiste wartości, będą odchyłać się co do wartości bezwzględnej

wartości od prognoz.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - Y_{tp})^2 \quad (2.24)$$

- RMSE (ang. Root Mean Square Error) - pierwiastek błędu średniokwadratowego, bada, o ile średnio odchylają się rzeczywiste wartości zmiennej prognozowanej od obliczonych prognoz,

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - Y_{tp})^2} \quad (2.25)$$

- MAPE (ang. Mean Absolute Percent Error) średni absolutny błąd procentowy jego postać jest przedstawiona we wzorze (2.26)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - Y_{tp}|}{Y_t} \quad (2.26)$$

gdzie:

n - liczba okresów,

Y_t - rzeczywista wartość zmiennej prognozowanej Y w okresie t ,

Y_{tp} - prognoza zmiennej Y na okres t .

Rozdział 3

Prognozowanie stężenia pyłów PM2.5 i PM10

3.1 Przygotowanie danych do analizy

3.1.1 Źródła danych

Do zbudowania precyzyjnego modelu predykcyjnego potrzebujemy dokładnych i ciągłych danych zebranych w danym okresie. Dane do przeprowadzanych analiz zostały uzyskane za pośrednictwem GIOŚ (Główny Inspektorat ochrony środowiska) i IMGW (Instytut Meteorologii i Gospodarki Wodnej). Ze strony Internetowej GIOŚ udało się pobrać dane dotyczące zawartości pyłów znajdujących się w powietrzu dla jednodniowego czasu uśredniania.

We wstępnej analizie odrzucone zostały stacje, w których występowały za duże braki lub stacje zaprzestawały pomiarów z powodu awarii lub zamknięcia.

W związku z tym wykorzystane zostały informacje z okresu 01.01.2013 r. - 31.12.2017r. Dane punktów na których mierzone były zawartości pyłów PM10 i PM2.5, przedstawione zostały w tabeli 3.1

Tabela 3.1. Dane stacji mierzących zawartości pyłów PM10 i PM2.5 w powietrzu

Kod stacji	Kod międzynarodowy	Nazwa stacji	Badany pył	Rodzaj stacji	Ulica	Kompletność danych
MzWarTolstoj	PL0308A	Warszawa-Tołstoja	PM10	kontenerowa stacjonarna	ul. Tołstoja 2	97,1%
MzWarKondrat	PL0143A	Warszawa-Targówek	PM2.5	kontenerowa stacjonarna	ul. Kondratowicza 8	96,3%

Źródło: Opracowanie własne na podstawie danych GIOŚ

Do zwiększania lub zmniejszania się zawartości pyłów PM10 i PM2.5 sprzyjają odpowiednie warunki atmosferyczne. W związku z tym zostały one uwzględnione w badaniach prognostycznych. Na podstawie plików umieszczonych na stronie internetowej IMGW udało się zebrać informacje dotyczące warunków atmosferycznych. Z otrzymanych danych o pogodzie, wykorzystałem te, które mierzone były najbliżej stacji pomiarowych dla pyłów. Wybrane zostało stanowisko Warszawa-bielany mieszczące się na ul. Podleśnej 61 w Warszawie. W zbiorze znajdują się informacje dotyczące jednodniowego poziomu uśredniania wartości dla temperatury, wilgotności, prędkości wiatru i zachmurzenia. Pomiary zostały dokonywane nieprzerwanie w okresie 01.01.2013 - 31.12.2017 r.

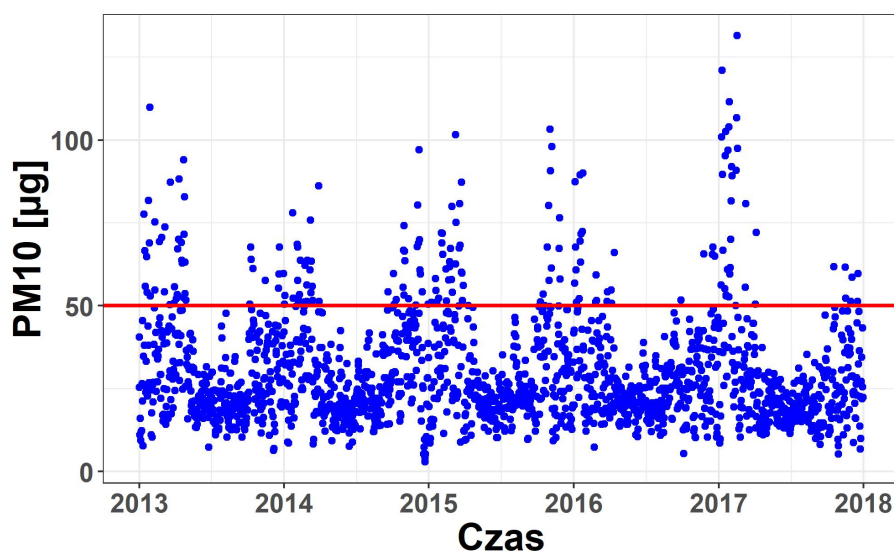
3.1.2 Imputacja braków metodą knn

W rozdziale 2.1 została przedstawiona budowa szeregu czasowego, z której wynika, że do stworzenia modelu predykcyjnego konieczna jest zachowana ciągłość przedstawionych obserwacji w badanym okresie. W związku z występującą niekompletnością niektórych obserwacji dotyczących pyłów w latach 2013-2017 została przeprowadzona imputacja danych metodą k najbliższych sąsiadów z wykorzystaniem funkcji kNN z pakietu VIM (Kowarik & Templ, 2016). Dzięki tej metodzie stworzona została możliwość przeprowadzenia imputacji do danych 53(2,9%) obserwacje dla stacji MzWarKondrat mierzącej zawartość pyłów PM2.5, natomiast dla stacji MzWarTolstoj mierzącej zawartość pyłów PM10 67(3,7%) obserwacji.

3.2 Eksploracyjna analiza danych

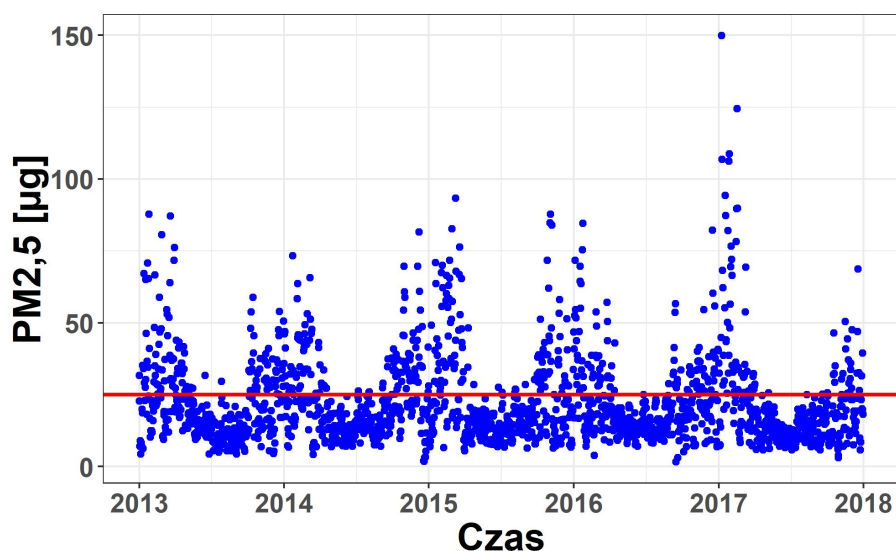
3.2.1 Wizualizacja zmiennych

W tej części pracy zostanie przeprowadzona szczegółowa analiza pozyskanych danych za pomocą wizualizacji oraz statystyk opisowych.



Rysunek 3.1. Wykres przedstawiający wartości pyłu PM10 w latach 2013-2017 wraz z linią wyznaczającą dopuszczalne normy według Światowej Organizacji Zdrowia dla dziennego okresu uśredniania (WHO, 2005, s. 8)

Źródło: Opracowanie własne na podstawie (WHO, 2005) oraz danych GIOŚ

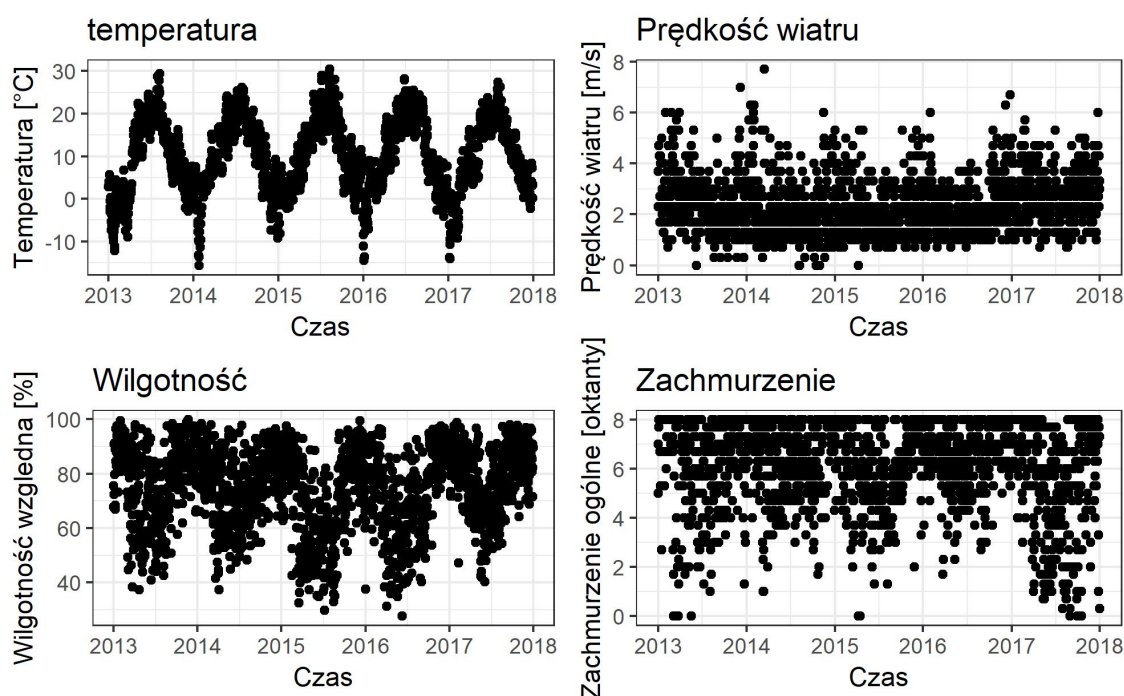


Rysunek 3.2. Wykres przedstawiający wartości pyłu PM_{2.5} w latach 2013-2017 wraz z linią wyznaczającą dopuszczalne normy według Światowej Organizacji Zdrowia dla dziennego okresu uśredniania (WHO, 2005, s. 8)

Źródło: Opracowanie własne na podstawie danych GIOŚ i WHO

Z przedstawionych rysunków 3.1 i 3.2 wynika, że najwyższe wartości pyłu PM₁₀ i PM_{2.5} występują w okresach zimowych, co spowodowane jest przez wzmożoną potrzebę ogrzewania mieszkań. Najczęstszą metodą dostarczania ciepła jest spalanie paliw kopalnych, dlatego wartości w tych okresach są wyższe niż w czasie pozostałych pór roku. W okresie lat 2013-2017 dopuszczalne normy wartości stężeń wyznaczone według (WHO, 2005) wynoszące 50 µg dla pyłów PM₁₀ zostały przekroczone 194 razy. Natomiast dla pyłów PM_{2.5} dopuszczalna wartość wynosząca 25 µg, została przekroczona 538 razy.

Przebiegi zmiennych dotyczących pogody w badanym okresie zostały przedstawione na rysunku 3.3.



Rysunek 3.3. Wykresy przedstawiające przebieg zmiennych dotyczących pogody w czasie

Źródło: Opracowanie własne na podstawie danych IMGW

Temperatura charakteryzuje się wyraźną sezonowością ze względu na występowanie pór roku, najwyższe wartości są osiągane w miesiącach letnich. W przypadku wilgotności, nie jest tak wyraźna - niższe wartości występują w miesiącach letnich. Najwyższa temperatura została odnotowana 09.08.2015 r. jej wartość wyniosła 30,4°C, najniższa wyniosła -15,6°C dnia 25.01.2014 r. W przypadku wilgotności najwyższa wartość została osiągnięta 22.11.2013 r. z wynikiem 99%, najniższa została zaobserwowana 08.06.2016 r. i wyniosła 27,8%.

3.2.2 Miary klasyczne i pozycyjne

Do przedstawienia opisu zbioru danych dla pyłu PM10 i PM25 posłużą miary klasyczne oraz miary pozycyjne.

Tabela 3.2. Porównanie klasycznych miar pyłów PM10 i PM2.5

Rodzaj pyłu	PM10	PM2.5
Liczba obserwacji	1826	1826
Minimum	2,90	1,63
Maksimum	131,53	149,89
Zakres	128,63	148,26
Średnia	29,68	22,29
Odchylenie Standardowe	16,49	15,56
Klasyczny współczynnik zmienności (Vx)	55,57%	69,83%
Typowy obszar zmienności (Xtyp)	13,19<Xtyp<46,17	6,73<Xtyp<37,86
Kurtoza	7,35	10,32
Skośność	1,77	2,20

Źródło: opracowanie własne na podstawie danych GIOŚ

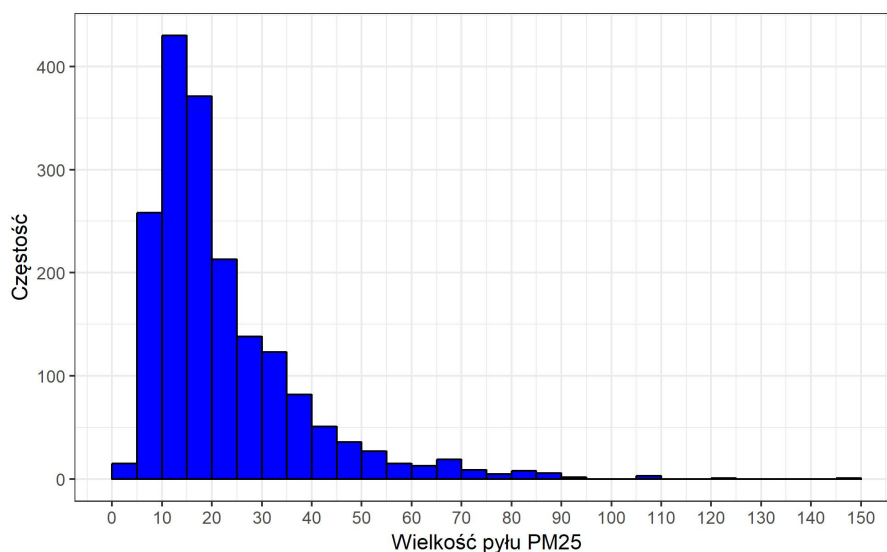
Na podstawie umieszczonych miar w tabeli 3.2 można zauważyć, że najniższa zawartość pyłu PM10 wynosiła $2,9\mu\text{g}$, zaobserwowana została 16.02.2017 r. W przypadku pyłu PM2.5 wartość ta wynosiła $1,63\mu\text{g}$ w dniu 16.02.2017 r. Maksymalna wielkość pyłu PM10 została zaobserwowana 13.12.2014 r. z wartością $131,53\mu\text{g}$, a dla pyłu PM2.5 16.02.2017r. z wynikiem $149,89\mu\text{g}$. W obu przypadkach wartości współczynników zmienności informują o silnym zróżnicowaniu. Średnie wartości dla pyłów PM10 i PM2.5 wynoszą odpowiednio $29,68\mu\text{g}$ i $22,29\mu\text{g}$, średnio odchylając się od średnich o $16,49\mu\text{g}$ i $15,56\mu\text{g}$. Dodatkowo wartość skośności w obu pyłach wskazują na prawostronną asymetrię.

Tabela 3.3. Porównanie pozycyjnych miar pyłów PM10 i PM2.5

Rodzaj pyłu	PM10	PM2.5
Q1	18,69	12,00
Mediana	25,01	17,60
Q3	37,01	27,80
Odchylenie ćwiartkowe(Q)	9,160	7,90
Pozycyjny współczynnik zmienności(Vq)	36,63%	44,89%
Pozycyjny współczynnik asymetrii (Aq)	62%	58,23%

Źródło: opracowanie własne na podstawie danych GIOŚ

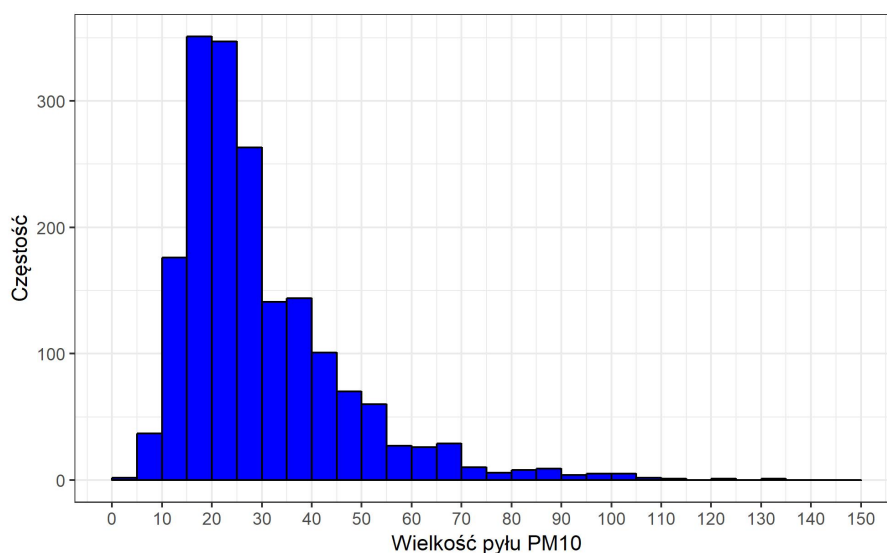
Na podstawie umieszczonych miar w tabeli 3.3 wynika, że mediana dla pyłu PM10 wynosi $25,01\mu\text{g}$ i dla pyłu PM2.5 wynosi $17,6\mu\text{g}$. W obu przypadkach można zauważyć, że mediana znajduje się bliżej wartości kwartyła górnego.



Rysunek 3.4. Histogram dla pyłu PM2.5

Źródło: opracowanie własne na podstawie danych GIOŚ

Na rysunku 3.4 przedstawiony został histogram dla wielkości pyłu PM2.5 charakteryzujący się prawostronną asymetrią. Z wykresu można zauważyć również, że najczęściej występującymi wartościami są wartości z przedziału 10 - 15.



Rysunek 3.5. Histogram dla pyłu PM10

Źródło: opracowanie własne na podstawie danych GIOŚ

Z kolei na rysunku 3.5 przedstawiony został histogram dla wielkości pyłu PM10

z występująca prawostronna asymetrią. Z wykresu można zauważyć również, że najczęściej występującymi wartościami są wartości z przedziału 15 - 25.

3.3 Prognoza pyłu PM2.5

Do zbudowania modeli prognostycznych i ich testowania potrzebne są dwa zbiory jeden na którym model jest w stanie nauczyć się pewnych zachowań i drugi na którym możemy sprawdzić wcześniej nauczony model. W związku z tym dane zostały podzielone na dwie grupy. Jedna z nich odpowiada za model treningowy, który zawarty jest w przedziale czasowym 01.01.2013 r. - 03.12.2017 r. Druga grupa natomiast była odpowiedzialna za testowanie modeli, jej przedział czasowy zawierał się w okresie 04.12.2017 r. - 10.12.2017 r. Taki podział przyjęto ze względu na możliwość wdrożenia proponowanego rozwiązania do przewidywania stężenia pyłów PM2.5 i PM10 na podstawie historycznych pomiarów oraz prognoz pogody.

3.3.1 Prognoza Holta-Wintersa

Wartości początkowe parametrów zostały wyznaczone za pomocą funkcji *Holtwinters* znajdującej się w pakiecie *stats* (R Core Team, 2018). Funkcja ta dąży do znalezienia optymalnych wartości α, β, γ minimalizując błąd średnich prognoz wygasłych. Wartości poszczególnych parametrów umieszczone zostały w tabeli 3.4.

Tabela 3.4. Wartości współczynników występujących w modelu Wintersa w wersji addytywnej dla pyłu PM10

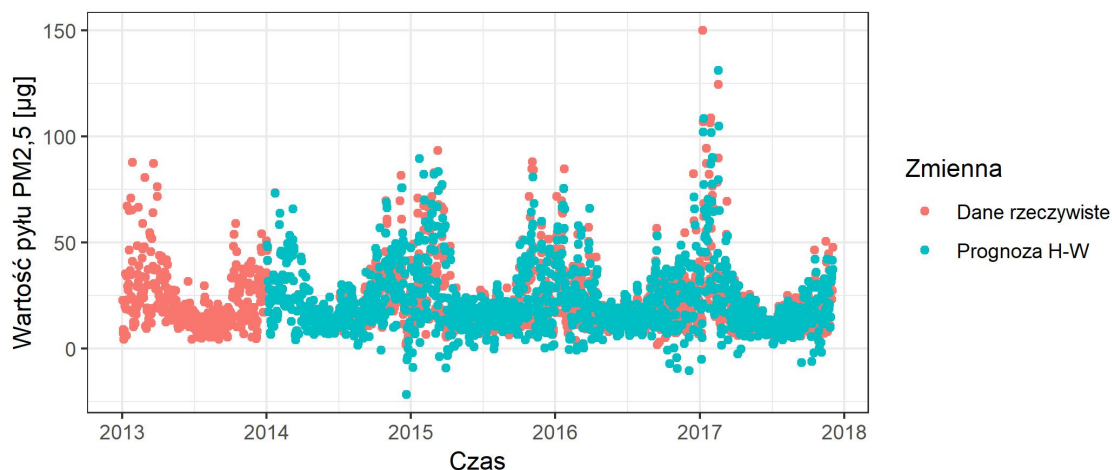
Współczynnik	Wartość współczynnika
α	0,66
β	0
γ	0,65

Źródło: opracowanie własne na podstawie danych GIOŚ

Na podstawie wielkości oszacowanych parametrów z tabeli 3.4, można stwierdzić, że wartość parametru α wskazuje, na to że poprzednie wartości znacząco

wpływają na prognozy. Z kolei współczynnik β wskazuje na brak zmian w trendzie, a z wysokiej wartości γ wynika występowanie wahań sezonowych.

Wartości modelu w badanym okresie również zostały obliczone z pomocą funkcji HoltWinters. Przebieg szacowanych w ten sposób wartości w czasie został pokazany na rysunku 3.7.



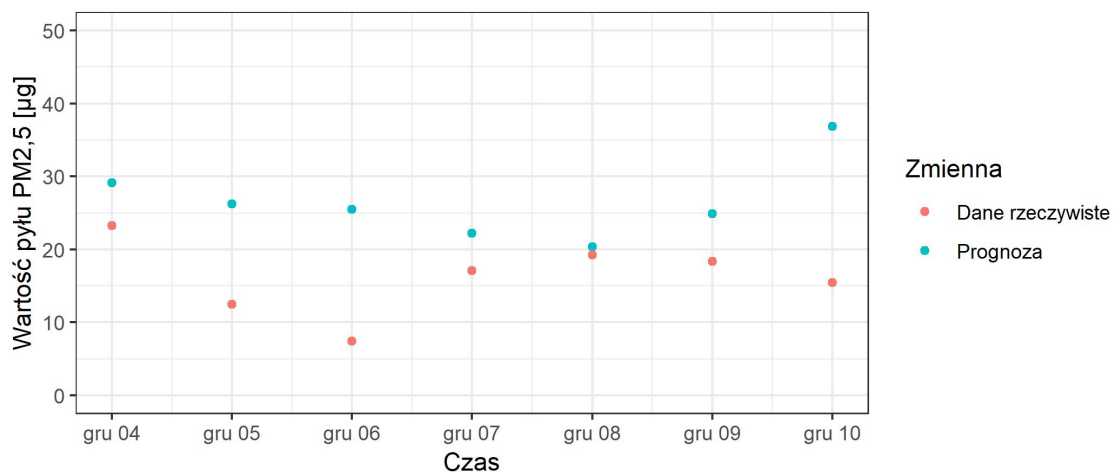
Rysunek 3.6. Wykres przedstawiający przebiegi rzeczywistych wartości oraz prognoz ex ante modelu Holta-Wintersa PM2.5 w okresie 01.01.2013 r. - 3.12.2017 r.

Źródło: opracowanie własne na podstawie danych GIOŚ

Sprawdzona została dokładność prognozy Holta-Wintersa dla pyłu PM2.5 w okresie ex ante — 01.01.2013 r. - 03.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wyniosła -0,01 świadczy o tym, że rzeczywista wartość pyłu PM2.5 była niższa od prognozy średnio o $0,01\mu\text{g}$,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 47,3%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 47,3%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący $13,44\mu\text{g}$, zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi $13,44\mu\text{g/dzień}$.

Do wyznaczenia wartości przewidywań w okresie prognozy (04.12.2017 r. - 10.12.2017 r.) wykorzystana została funkcja predict znajdująca się w pakiecie stats (R Core Team, 2018). Na rysunku 3.7 przedstawione zostały rzeczywiste oraz prognozowane wartości pyłu PM2.5.



Rysunek 3.7. Wykres przedstawiający wartości rzeczywiste i prognozowane ex post za pomocą modelu Holta-Wintersa w okresie 04.12.2017r. - 10.12.2017r.

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Ponieważ znana jest rzeczywista wartość pyłu PM2.5 w okresie 04.12.2017 r. - 10.12.2017 r., dzięki określonej wcześniej podziałowi danych na dwie części, możemy posłużyć się miarami ex ante do obliczenia miar błędów prognoz.

Sprawdzona została trafność prognozy Holta-Wintersa dla pyłu PM2.5 w okresie 04.12.2017 r. - 10.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

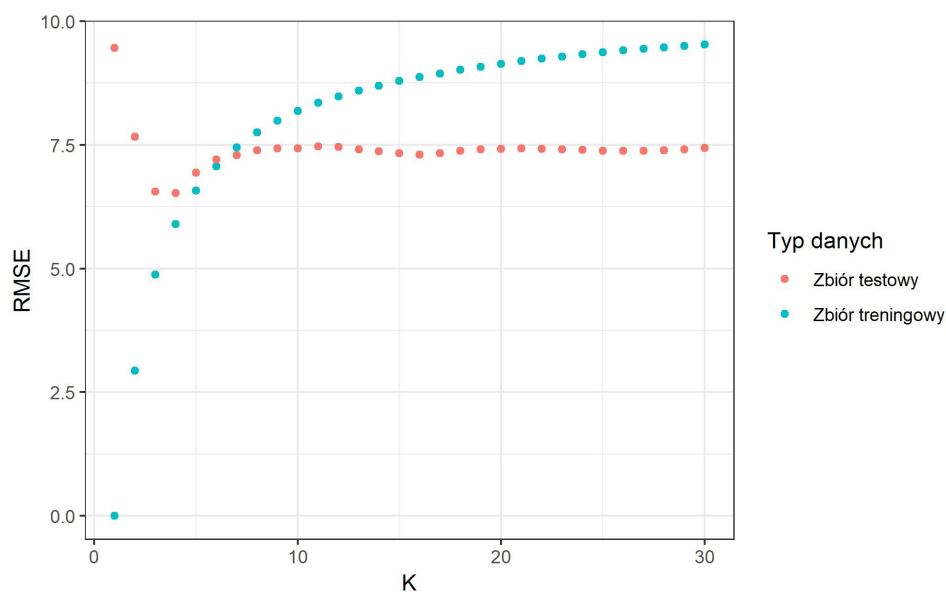
- Obliczona wartość średniego błędu prognozy (ME) wyniosła 10,26 świadczy to tym, że rzeczywista wartość pyłu PM2.5 była wyższa od prognozy średnio o 10,26 μg ,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 83,97%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 83,97%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego

(RMSE) wynoszący $13,44\mu\text{g}$, zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi $13,44\mu\text{g}/\text{dzień}$.

3.3.2 Prognoza z wykorzystaniem regresji k najbliższych sąsiadów

Model z wykorzystaniem regresji k najbliższych sąsiadów zbudowany został w oparciu o dane GIOŚ z których wykorzystane zostały pomiary pyłu PM2.5, dodatkowo uwzględnione zostały panujące w tym czasie warunki atmosferyczne dostarczone przez IMGW. W celu optymalizacji błędu prognozy, wybrane zostały cechy mające największy wpływ na najmniejsze wyniki odchyień. Wobec tego w modelu wykorzystano dane dotyczące pyłu PM2.5, temperatury, wilgotności, zachmurzenia i zmiennej PM2.5 opóźnionej o 1 okres. Następnie na podstawie wybranych czynników, określony został model przy użyciu funkcji `model.frame` z pakietu `stats` (R Core Team, 2018). Kolejnym krokiem było sformułowanie zadania regresji z pomocą metody `makeRegrTask` znajdującej się w pakiecie `mlr` (Bischl i in., 2016). Jednocześnie dla wybranego wcześniej wartości k (liczby najbliższych sąsiadów) została określona metoda nauki dla modelu uczącego się jako regresja k najbliższych sąsiadów. Na tej podstawie zbudowany został model, służący do określenia prognozy z pomocą funkcji `train` występującej w pakiecie `mlr` (Bischl i in., 2016).

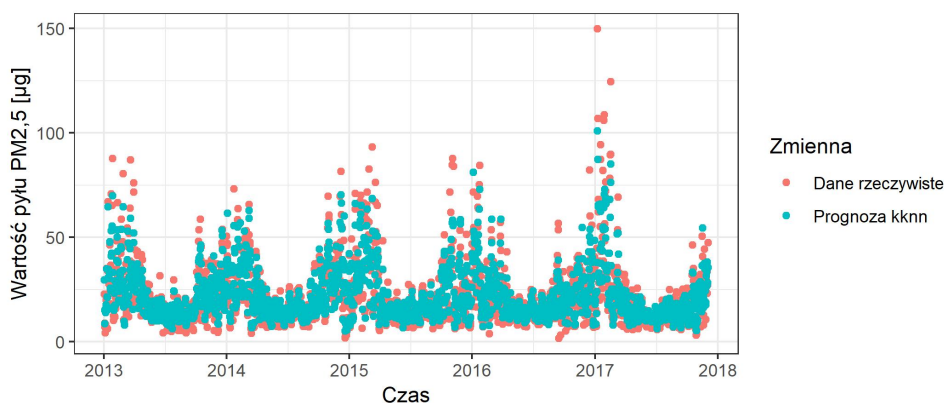
Metodą walidacji krzyżowej określona została wartość liczby najbliższych sąsiadów, wynosząca w przypadku pyłu PM2.5 6, aby zminimalizować różnicę pierwiastka błędu średniokwadratowego pomiędzy okresem *ex ante* a okresem *ex post*. Przebieg wartości RMSE ze względu na liczbę sąsiadów został przedstawiony na rysunku 3.8



Rysunek 3.8. Przebieg procesu poszukiwania najlepszej wartości K dla pyłu PM2.5

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Do wyznaczenia wartości dla okresu ex ante — 01.01.2013 r. - 03.12.2017 r. posłużono się funkcją predict znajdującą się w pakiecie stats (R Core Team, 2018).



Rysunek 3.9. Porównanie wartości rzeczywistych PM2.5 z wartościami z modeli z wykorzystaniem regresji k najbliższych sąsiadów w czasie ex ante

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

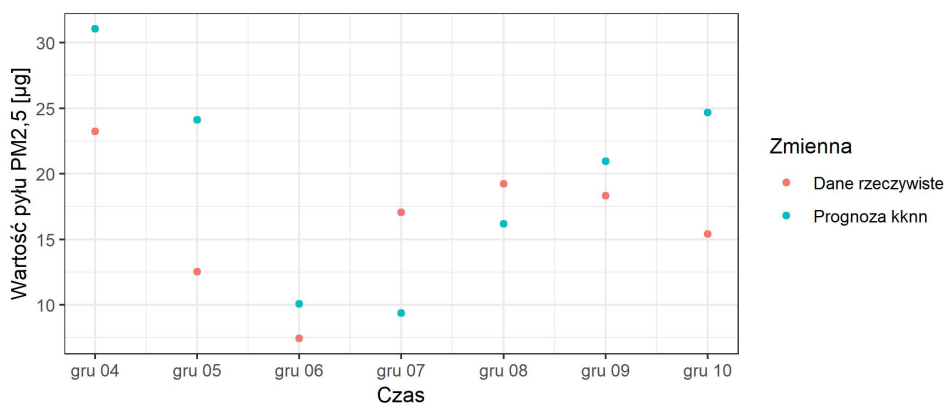
Sprawdzona została dokładność prognozy regresji k najbliższych sąsiadów dla pyłów PM2.5 w okresie 01.01.2013 r. - 03.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wyniosła -0,24 świadczy o tym, że rzeczywista wartość pyłu PM2.5 była niższa od prognozy średnio o

0,24 μg ,

- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 25,6%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 25,6%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący 7,07 μg , zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi 7,07 μg /dzień.

Na podstawie zebranych wcześniej danych przez model uczący się można było wyznaczyć prognozy na okres ex post — 04.12.2017 r. - 10.12.2017 r., których wartości wraz z rzeczywistymi, zostały ukazane na rysunku 3.15.



Rysunek 3.10. Porównanie wartości rzeczywistych PM2.5 z wartościami z modeli z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex post

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Sprawdzona została trafność prognozy regresji k najbliższych sąsiadów dla pyłów PM2.5 w okresie ex post 04.12.2017 r. - 10.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wyniosła 3,31 świadczy to tym, że rzeczywista wartość pyłu PM2.5 była wyższa od prognozy średnio o 3,31 μg ,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 42,43%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 42,43%,

- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący $7,21\mu\text{g}$ zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi $7,21\mu\text{g}/\text{dzień}$.

3.3.3 Porównanie modeli

Tabela 3.5. Porównanie miar błędów zbudowanych modeli dla pyłu PM2.5

PM2.5	Okres ex ante		Okres ex post	
Miara błędu	Model Holta-Wintersa	Model kkn	Model Holta-Wintersa	Model kkn
ME	-0,01	-0,24	10,27	3,31
MAPE	47,3%	25,59%	83,97%	42,43%
RMSE	13,45	7,07	12,41	7,21

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Na podstawie tabeli 3.5 wynika, że w okresie ex post modeli dla pyłu PM2.5 lepsze wartości miar błędów osiąga model z wykorzystaniem regresji k najbliższych sąsiadów.

3.4 Prognoza pyłu PM10

3.4.1 Prognoza Holta-Wintersa

Tak samo jak w przypadku pyłów PM2.5, wartości początkowe parametrów modelu zostały wyznaczone za pomocą funkcji Holtwinters znajdującej się w pakiecie stats (R Core Team, 2018). Funkcja ta dąży do znalezienia optymalnych wartości α, β, γ minimalizując błąd średnich prognoz wygasłych. Wartości poszczególnych parametrów umieszczone zostały w tabeli 3.6

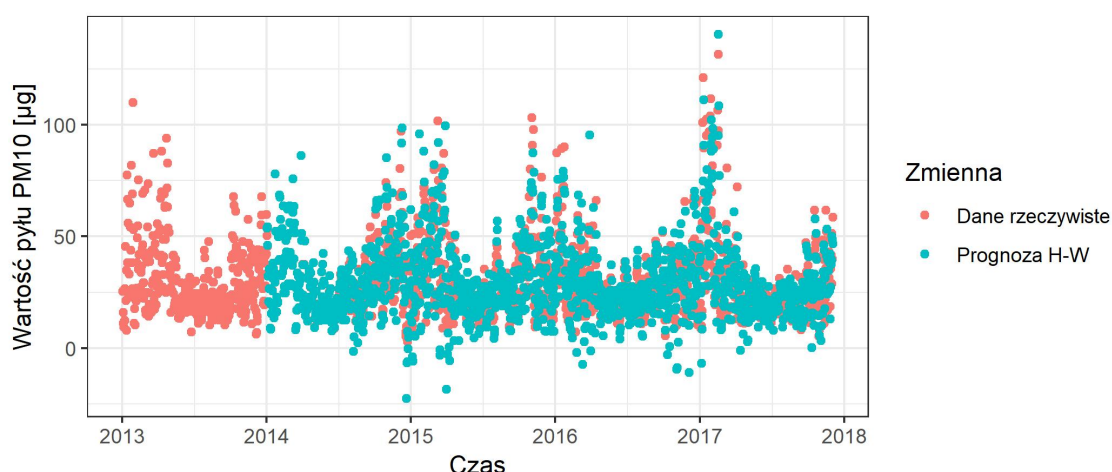
Tabela 3.6. Wartości wyznaczonych współczynników dla modelu Holta-Wintersa dla pyłu PM10

Nazwa współczynnika	Wartość współczynnika
α	0,68
β	0
γ	0,84

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Na podstawie wielkości oszacowanych parametrów zawartych w tabeli 3.6, można stwierdzić, że wartość parametru α wskazuje, że poprzednie wartości znacząco wpływają na prognozy. Z kolei współczynnik β wskazuje na brak zmian w trendzie, a z wysokiej wartości γ wynika występowanie wahań sezonowych.

Wartości modelu w badanym okresie również zostały obliczone za pomocą funkcji HoltWinters. Przebieg szacowanych w ten sposób wartości w czasie został pokazany na rysunku 3.11.

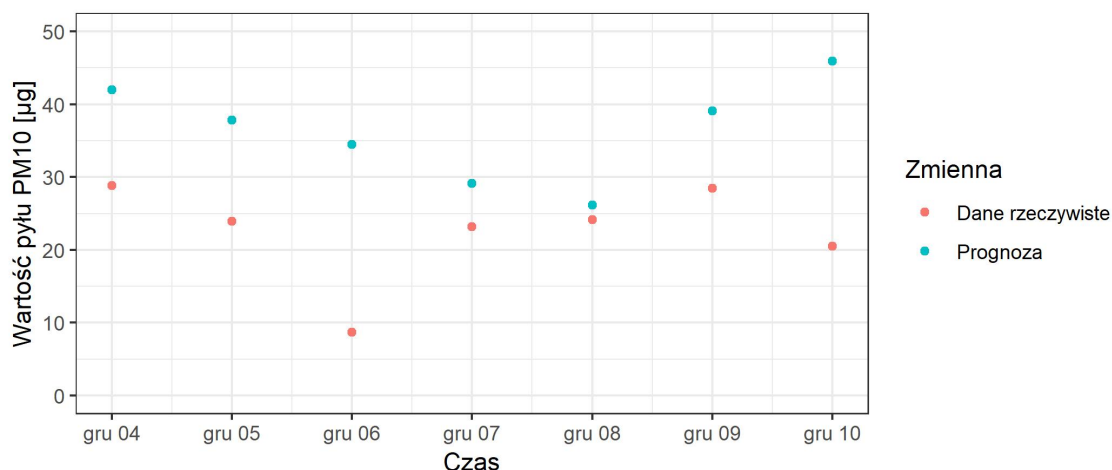


Rysunek 3.11. Wykres przedstawiający przebiegi rzeczywistych wartości PM10 oraz prognoz ex ante modelu Holta-Wintersa PM10 w okresie 01.01.2013 r. - 03.12.2017 r.

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Sprawdzona teraz zostanie trafność prognozy w okresie 01.01.2013 r. - 03.12.2017 r. z wykorzystaniem miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wynosząca -0,02 świadczy o tym, że rzeczywista wartość pyłu PM10 była niższa od prognozy średnio o 0,02 μg ,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 38,3%, dzięki niej wiadomo, że prognozy różniły się od wartości rzeczywistych średnio dziennie o 38,3%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący 14,6 μg , zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi 14,6 $\mu\text{g}/\text{dzień}$.



Rysunek 3.12. Wykres przedstawiający przebiegi rzeczywistych wartości oraz prognoz ex post modelu Holta-Wintersa PM10 w okresie 03.12.2017 r. - 10.12.2017 r.

Źródło: opracowanie własne na podstawie danych GIOŚ

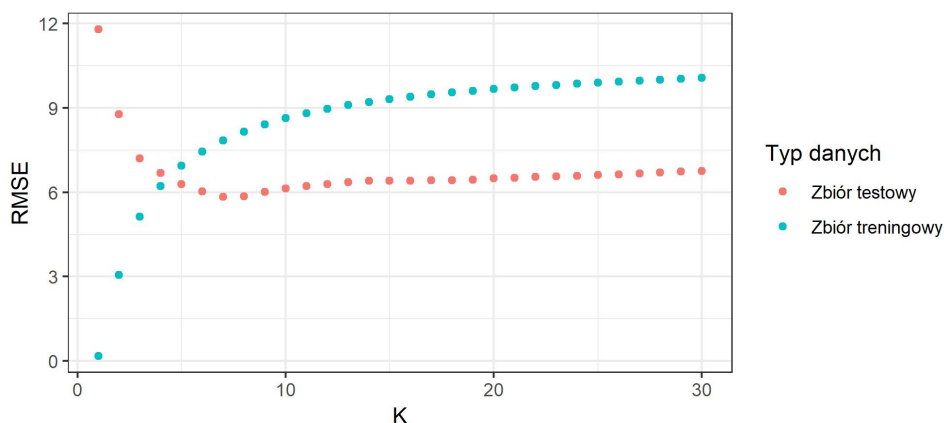
Sprawdzona została trafność prognozy Holta-Wintersa dla pyłu PM10 w okresie 04.12.2017 r. - 10.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wyniosła 13,82 świadczy to tym, że rzeczywista wartość pyłu PM10 była wyższa od prognozy średnio o 13,82µg,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 84,95%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 84,95%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący 16,15µg, zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi 16,15µg/dzień.

3.4.2 Prognoza z wykorzystaniem regresji k najbliższych sąsiadów

W sposób analogiczny jak w przypadku pyłu PM2.5 został zbudowany model dla pyłu PM10.

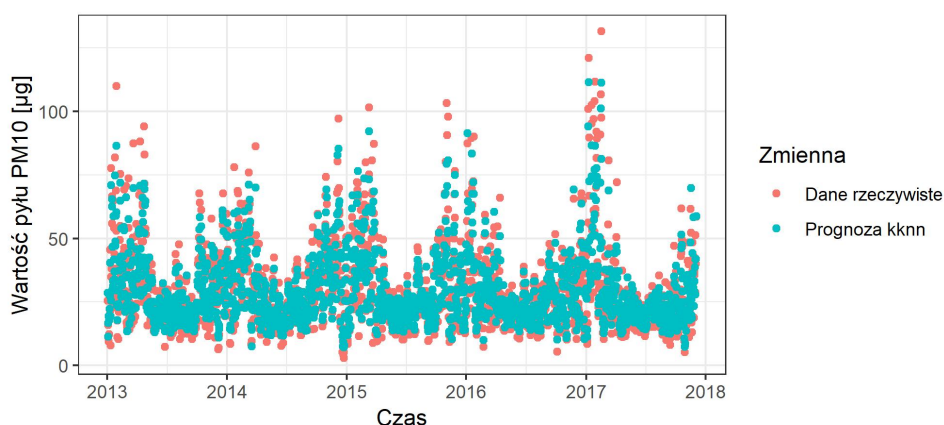
Metodą walidacji krzyżowej określona została wartość liczby najbliższych sąsiadów, wynosząca w przypadku pyłu PM10 4, aby zminimalizować różnicę pierwiastek błędu średniokwadratowego pomiędzy okresem ex ante a okresem ex post. Przebieg wartości RMSE ze względu na liczbę sąsiadów został przedstawiony na rysunku 3.13



Rysunek 3.13. Przebieg wartości RMSE w poszukiwaniu najlepszego K dla modelu z pyłem PM10

Źródło: Opracowanie własne na podstawie danych GIOŚ i IMGW

Do wyznaczenia wartości dla okresu ex ante — 01.01.2013 r. - 03.12.2017 r. posłużono się funkcją predict znajdującą się w pakiecie stats (R Core Team, 2018).



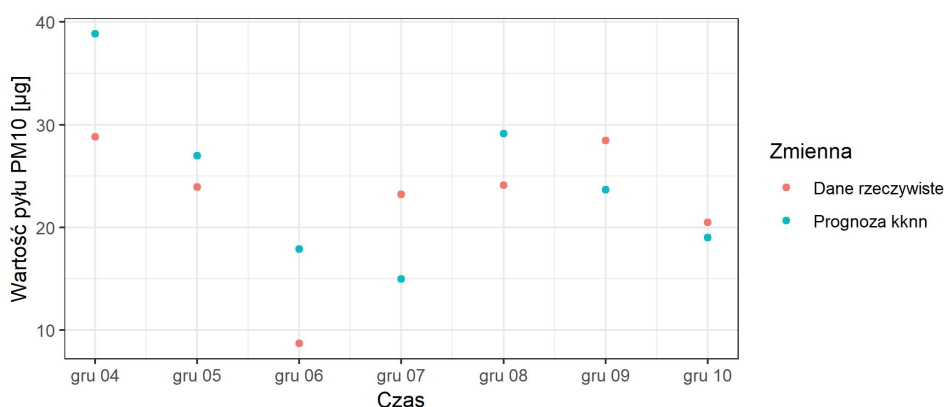
Rysunek 3.14. Porównanie wartości rzeczywistych z wartościami z modelu PM10 z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex post

Źródło: Opracowanie własne na podstawie danych GIOŚ i IMGW

Sprawdzona teraz zostanie trafność prognozy pyłu PM10 w okresie ex ante — 01.01.2013 r. - 03.12.2017 r. z wykorzystaniem miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wynosząca $-0,17$ świadczy o tym, że rzeczywista wartość pyłu PM10 była niższa od prognozy średnio o $0,17 \mu\text{g}$,
- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła $16,76\%$, dzięki niej wiadomo, że prognozy różniły się od wartości rzeczywistych średnio dziennie o $16,76\%$,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący $6,23 \mu\text{g}$, zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi $6,23 \mu\text{g}/\text{dzień}$.

Na podstawie zebranych wcześniej danych przez model uczący się można było wyznaczyć prognozy na okres ex post — 04.12.2017 r. - 10.12.2017 r., których wartości wraz z rzeczywistymi, zostały ukazane na rysunku 3.15.



Rysunek 3.15. Porównanie wartości rzeczywistych PM10 z wartościami z modelu PM10 z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex ante

Źródło: Opracowanie własne na podstawie danych GIOŚ i IMGW

Sprawdzona została trafność prognozy regresji k najbliższych sąsiadów dla pyłu PM10 w okresie 04.12.2017 r. - 10.12.2017 r. z wykorzystaniem niektórych miar błędów prognoz ex ante.

- Obliczona wartość średniego błędu prognozy (ME) wyniosła $1,82$ świadczy to o tym, że rzeczywista wartość pyłu PM10 była wyższa od prognozy średnio o $1,82 \mu\text{g}$,

- Drugą wykorzystaną miarę stanowił średni absolutny błąd procentowy (MAPE), którego wartość wyniosła 33,34%, dzięki niej wiadomo że prognozy różniły się od wartości rzeczywistych średnio dziennie o 33,34%,
- Kolejną wyróżnioną miarą jest pierwiastek błędu średniokwadratowego (RMSE) wynoszący $6,68\mu\text{g}$, zgodnie z wartością tego błędu przeciętna różnica między wartością rzeczywistą i prognozowaną wynosi $6,68\mu\text{g}/\text{dzień}$.

3.4.3 Porównanie modeli

Tabela 3.7. Porównanie miar błędów modeli dla pyłu PM10

Pył PM10	Okres ex ante		Okres ex post	
	Model Holta-Wintersa	Model kkn	Model Holta-Wintersa	Model kkn
ME	-0,02	-0,17	13,82	1,82
MAPE	38,3%	16,76%	84,95%	33,35%
RMSE	14,61	6,23	16,15	6,68

Źródło: opracowanie własne na podstawie danych GIOŚ i IMGW

Na podstawie tabeli 3.7 wynika, że w okresie ex post modeli dla pyłu PM10 lepsze wartości miar błędów osiąga model z wykorzystaniem regresji k najbliższych sąsiadów.

Podsumowanie

W niniejszej pracy podjęto bardzo ważną i aktualną tematykę prognozowania zanieczyszczenia powietrza pyłami PM2.5 i PM10. W przeprowadzonym badaniu wykorzystano dane pochodzące ze stacji mierzących zawartość pyłów w powietrzu oraz stacji pogodowych. Wszystkie punkty pomiarowe znajdowały się na terenie Warszawy.

Dane, które zostały poddane analizom, dostępne są na stronach internetowych GIOŚ i IMGW. Dane pochodzące od IMGW odnośnie panujących warunków atmosferycznych były kompletne. W przypadku danych z IMGW, zawierających informacje o pyłach, występowały pewne braki, wobec tego przeprowadzona została imputacja. Wynikało to z konieczności zachowania ciągłości danych, aby móc przeprowadzić badania dotyczące szeregów czasowych.

Do oszacowania wartości związanych z szeregami czasowymi wskazującymi na występowanie wahań sezonowych, użyte zostały dwie metody jedna z nich Holta-Wintersa zakładała, że na wartości pyłu PM10 i PM2.5 mają wpływ tylko wartości historyczne. Wyniki przeprowadzonych w ten sposób badań wykazały, że w okresie ex post przeciętna różnica między wartością rzeczywistą i prognozowaną dla pyłu PM10 wynosiła $16,15\mu\text{g}$, a dla pyłu PM2.5 była równa $12,41\mu\text{g}$. Z kolei drugim sposobem na wyznaczenie wartości prognoz została wykorzystana regresja k najbliższych sąsiadów, w oparciu o model zbudowany na podstawie informacji o pyle, temperaturze, zachmurzeniu, oraz sile wiatru. Na tej podstawie wyliczono przeciętną różnicę pomiędzy wartością rzeczywistą i prognozowaną. Dla pyłu PM10 w okresie ex post wartość ta wynosiła $6,23\mu\text{g}$ a dla PM2.5 $7,07\mu\text{g}$. Wyznaczone w ten sposób wartości metodą regresji k najbliższych sąsiadów znacząco różnią się wartości prognozowanych wyznaczonych metodą Holta-Wintersa. Lepszy więc okazał się model zbudowany na podstawie wielkości pyłów i panują-

cych w tym czasie warunków atmosferycznych i których wartości prognoz wyznaczone były z pomocą regresji k najbliższych sąsiadów. Wszystkie przeprowadzone obliczenia zostały wykonane z wykorzystaniem języka R.

Zaproponowane w pracy rozwiązanie może zostać wykorzystane w praktyce do prognozowania stężenia pyłów PM_{2.5} i PM₁₀. Bazuje ono na przeszłych wartościach miar zanieczyszczenia powietrza oraz danych pogodowych, a do predykcji można wykorzystać prognozy pogody, które aktualnie są bardzo dokładne.

Proponowane w niniejszej pracy rozwiązania mogą być dalej rozwijane przez użycie sieci neuronowych, które mogłyby poprawić wyniki odchyłeń prognoz od zmiennej prognozowanej. Dodatkowo do modeli można wprowadzić wpływ położenia geograficznego i kierunki wiatru. Zawarte w pracy metody budowania modeli prognostycznych mogą zostać także rozwinięte do prognozowania pyłów np. dla jednogodzinnego okresu uśredniania.

Bibliografia

- ARMAAG, (b.d.), https://armaag.gda.pl/powietrze_02.htm[dostęp 14.02.2018].
- Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., ... Jones, Z. M., (2016), mlr: Machine Learning in R, *Journal of Machine Learning Research*, 17(170), 1–5.
- Cieślak, M., (2005), *Prognozowanie gospodarcze metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa.
- Dittman, P., (2008), *Prognozowanie w przedsiębiorstwie*, Oficyna a Wolters Kluwer business.
- Dobrzańska, B., Dobrzański, G. & Kiełczewski, D., (2012), *Ochrona środowiska przyrodniczego*, Wydawnictwo Naukowe PWN.
- Gajda, J. B., (2017), *Prognozowanie i symulacje w ekonomii i zarządzaniu*, Wydawnictwo CH Beck.
- GIOŚ, (2018), www.powietrze.gios.gov.pl/pjp/content/measuring_air_assessment_rating_info, [dostęp 28.01.2019].
- Janka, R. M., (2014), *Zanieczyszczenia pyłowe i gazowe. Podstawy obliczania i sterowania poziomem emisji*, Wydawnictwo Naukowe PWN.
- Kowarik, A. & Templ, M., (2016), Imputation with the R Package VIM, *Journal of Statistical Software*, 74(7), 1–16, doi:10.18637/jss.v074.i07
- Machura, Ł., (b.d.), Analiza szeregów czasowych, https://el.us.edu.pl/ekonofizyka/index.php/Analiza_Szeregów_Czasowych/Dekompozycja_szeregu_czasowego [dostęp 18.05.2019].
- Mazurek, H., (2018), *Smog zagrożenie dla zdrowia czy moda na ekologię?*, ITEM Publishing, Warszawa.
- Mazurek, H. & Badyda, A., (2018), *Smog konsekwencje zdrowotne zanieczyszczeń powietrza*, Wydawnictwo Naukowe PZWL.

- Polski Alarm smogowy, (2016), <https://polskialarmsmogowy.pl/polski-alarm-smogowy/smog/szczegoly,skad-sie-bierze-smog,18.html>.
- Pyłka-Gutowska, E., (2004), *Ekologia z ochroną środowiska*, Wydawnictwo Oświata.
- R Core Team, (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Rup, K., (2017), *Procesy przenoszenia zanieczyszczeń w środowisku naturalnym*, Wydawnictwo WNT.
- Samworth, R. i in., (2012), Optimal weighted nearest neighbour classifiers, *The Annals of Statistics*, 40(5), 2733–2763.
- Sayad, S., (2010), *An Introduction to Data Science*, Copyright.
- Schutt, R. & O’Neil, C., (2013), *Doing data science: Straight talk from the frontline*, O’Reilly Media, Inc.
- Starzyńska, W., (2012), *Statystyka praktyczna*, PWN.
- StatSoft, (2019), https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstknn.html [dostęp 22.05.2019r.]
- Szeliga, M., (2017), *Data science i uczenie maszynowe*, Wydawnictwo Naukowe PWN SA.
- Szmuksta-Zawadzka, M. & Zawadzki, J., (2012), *O miernikach dokładności prognoz ex post w prognozowaniu zmiennych o silnym natężeniu sezonowości*, Katedra Ekonometrii i Statystyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie.
- WHO, (2005), WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide, *Global update, 2006*.
- Wielgosiński, G. & Zarzycki, R., (2018), *Technologie i procesy ochrony powietrza*, Warszawa: Wydawnictwo Naukowe PWN SA.
- Zeliaś, A., Pawełek, B. & Wanat, S., (2008), *Prognozowanie ekonomiczne: teoria, przykłady, zadania*, Wydawnictwo Naukowe PWN.

Spis tabel

1.1	Wielkości emisji ze względu na źródła	4
1.2	Oddziaływanie pyłów na zdrowie człowieka	8
1.3	Pochodzenie i oddziaływanie Benzo(a)pirenu	9
1.4	Udział gazów w globalnym ociepleniu	10
1.5	Wartości norm dla pyłów PM10 i PM2.5 według WHO	11
1.6	Normy prawne dla substancji występujących w powietrzu w Polsce	11
1.7	Poziom docelowy dla niektórych substancji występujących w powie- trzu	11
1.8	Poziomy pyłu PM10 ze względu na zagrożenie	12
1.9	Ograniczanie emisji zanieczyszczeń - rozwiązania systemowe . . .	13
2.1	Ogólna postać szeregu czasowego	17
3.1	Dane stacji mierzących zawartości pyłów PM10 i PM2.5 w powietrzu	27
3.2	Porównanie klasycznych miar pyłów PM10 i PM2.5	31
3.3	Porównanie pozycyjnych miar pyłów PM10 i PM2.5	31
3.4	Wartości współczynników występujących w modelu Wintersa w wer- sji addytywnej dla pyłu PM10	33
3.5	Porównanie miar błędów zbudowanych modeli dla pyłu PM2.5 . .	39
3.6	Wartości wyznaczonych współczynników dla modelu Holta-Wintersa dla pyłu PM10	39
3.7	Porównanie miar błędów modeli dla pyłu PM10	44

Spis rysunków

1.1	Źródła emisji pyłu PM10 w Polsce	5
1.2	Rozmieszczenie stacji pomiarowych dla pyłów PM10 i PM2.5 . . .	15
3.1	Wykres przedstawiający wartości pyłu PM10 w latach 2013-2017 wraz z linią wyznaczającą dopuszczalne normy według Światowej Organizacji Zdrowia dla dziennego okresu uśredniania (WHO, 2005, s. 8)	28
3.2	Wykres przedstawiający wartości pyłu PM2.5 w latach 2013-2017 wraz z linią wyznaczającą dopuszczalne normy według Światowej Organizacji Zdrowia dla dziennego okresu uśredniania (WHO, 2005, s. 8)	29
3.3	Wykresy przedstawiające przebieg zmiennych dotyczących pogody w czasie	30
3.4	Histogram dla pyłu PM2.5	32
3.5	Histogram dla pyłu PM10	32
3.6	Wykres przedstawiający przebiegi rzeczywistych wartości oraz pro- gnoz ex ante modelu Holta-Wintersa PM2.5 w okresie 01.01.2013 r. - 3.12.2017 r.	34
3.7	Wykres przedstawiający wartości rzeczywiste i prognozowane ex post za pomocą modelu Holta-Wintersa w okresie 04.12.2017r. - 10.12.2017r.	35
3.8	Przebieg procesu poszukiwania najlepszej wartości K dla pyłu PM2.5	37
3.9	Porównanie wartości rzeczywistych PM2.5 z wartościami z modeli z wykorzystaniem regresji k najbliższych sąsiadów w czasie ex ante	37
3.10	Porównanie wartości rzeczywistych PM2.5 z wartościami z modeli z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex post	38

3.11 Wykres przedstawiający przebiegi rzeczywistych wartości PM10 oraz prognoz ex ante modelu Holta-Wintersa PM10 w okresie 01.01.2013 r. - 03.12.2017 r.	40
3.12 Wykres przedstawiający przebiegi rzeczywistych wartości oraz pro- gnoz ex post modelu Holta-Wintersa PM10 w okresie 03.12.2017 r. - 10.12.2017 r.	41
3.13 Przebieg wartości RMSE w poszukiwaniu najlepszego K dla modelu z pyłem PM10	42
3.14 Porównanie wartości rzeczywistych z wartościami z modelu PM10 z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex post	42
3.15 Porównanie wartości rzeczywistych PM10 z wartościami z modelu PM10 z wykorzystaniem regresji k najbliższych sąsiadów w okresie ex ante	43