

dr inż. Joanna Amelia Kamińska

Katedra Matematyki

Wydział Inżynierii Kształtowania Środowiska i Geodezji

Uniwersytet Przyrodniczy we Wrocławiu

Autoreferat

Wrocław, grudzień 2018 r.

Spis treści

1. Sylwetka habilitantki	1
1.1. Imię i nazwisko.....	1
1.2. Posiadane dyplomy i stopnie naukowe.....	1
1.3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych	2
2. Wskazanie osiągnięcia naukowego	3
2.1. Tytuł osiągnięcia naukowego	3
2.2. Wykaz publikacji stanowiących podstawę postępowania habilitacyjnego, o którym mowa w art. 16 ust. 2 ustawy	3
2.3. Cel naukowy pracy oraz osiągnięte wyniki	4
3. Pozostałe osiągnięcia naukowo–badawcze	26
4. Podsumowanie dorobku habilitantki	32

1. Sylwetka habilitantki

1.1. Imię i nazwisko

Joanna Amelia Kamińska

1.2. Posiadane dyplomy i stopnie naukowe

- 13.06.2007 r. **doktor nauk rolniczych** w zakresie kształtowania środowiska w dyscyplinie – agrometeorologia,
Wydział Inżynierii Kształtowania Środowiska i Geodezji,
Uniwersytet Przyrodniczy we Wrocławiu,
Tytuł rozprawy: *„Zmiany klimatycznego bilansu wodnego i bilansu cieplnego pszenicy jarej w wybranych rejonach Polski”*
Promotor: dr hab. Elżbieta Musiał, prof. UPWr
- 3.06.2002 r. **magister matematyki** w zakresie matematyki (z wyróżnieniem)
Wydział Matematyki i Fizyki, Uniwersytet w Białymstoku,
Temat pracy: *„Macierzowe reprezentacje pierścieni”*
Promotor: dr A. Andruszkiewicz
- 16.03.2012 r. **inżynier**, Mechanika i Budowa Maszyn,
Wydział Mechaniczny, Politechnika Wrocławska,
Temat pracy: *„Probabilistyczne ujęcie modelu zużycia układu pierścieni tłokowy – tuleja cylindrowa”*
Promotor: dr inż. K. Miksiewicz
- 25.06.2002 r. **licencjat** w zakresie matematyki finansowej,
Wydział Matematyki i Fizyki, Uniwersytet w Białymstoku,
Temat pracy: *„Kontrakty terminowe i ich zastosowanie w zarządzaniu ryzykiem inwestycyjnym”*
Promotor: dr J. Olbryś.

Inne formy edukacji

- 2013r. ukończenie studiów podyplomowych na kierunku *Menedżer projektów badawczych*,
Wyższa Szkoła Ekonomii i Innowacji w Lublinie.

1.3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- od 01.02.2009 adiunkt w Katedrze Matematyki Uniwersytetu Przyrodniczego we Wrocławiu
- 01.10.2007 – 31.01.2009 asystent w Katedrze Matematyki Uniwersytetu Przyrodniczego we Wrocławiu
- 01.10.2004 – 13.06.2007 studia doktoranckie na Wydziale Inżynierii Kształtowania Środowiska i Geodezji Uniwersytetu Przyrodniczego we Wrocławiu

2. Wskazanie osiągnięcia naukowego

Osiągnięciem naukowym wynikającym z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.) w związku z art.179 ust.2 ustawy przepisy wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz.1669) jest cykl publikacji powiązanych tematycznie.

2.1. Tytuł osiągnięcia naukowego

Badanie zależności stężenia zanieczyszczeń w powietrzu od czynników meteorologicznych oraz natężenia ruchu w kanionie komunikacyjnym

2.2. Wykaz publikacji stanowiących podstawę postępowania habilitacyjnego

- [A1] **Kamińska J.A.**, 2019. A random forest partition model for predicting NO₂ concentrations from traffic flow and meteorological conditions. *Science of the Total Environment* 651, p.475-483
<https://doi.org/10.1016/j.scitotenv.2018.09.196> (MNiSW – 40 pkt., IF₂₀₁₇=4,610)
- [A2] **Kamińska J.A.**, 2018. The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław, *Journal of Environmental Management* 217C (2018) p. 164-174
<https://doi.org/10.1016/j.jenvman.2018.03.094> (MNiSW – 35 pkt., IF₂₀₁₇=4,005)
- [A3] **Kamińska J.A.**, 2018. Probabilistic forecasting of nitrogen dioxide concentrations at an urban road intersection. *Sustainability* 2018, 10, 4213
doi:10.3390/su10114213 (MNiSW – 20 pkt., IF₂₀₁₇=2,075)
- [A4] **Kamińska J.A.**, 2018. Residuals in the modelling of pollution concentration depending on meteorological conditions and traffic flow, employing decision trees. *XLVIII Seminar of Applied Mathematics, ITM Web Conferences* 23, 00016 (2018)
<https://doi.org/10.1051/itmconf/20182300016> (MNiSW – 15pkt., IF=0)
- [A5] **Kamińska J.A.**, 2017. Zjawisko smogu na tle jakości powietrza we Wrocławiu w latach 2012-2016. *Inżynieria Ekologiczna, Ecological Engineering vol.18 Iss.5* p.66-76
<https://doi.org/10.12912/23920629/76777> (MNiSW – 9 pkt, IF=0)
- [A6] **Kamińska J. A.**, Chalfen M., Szczucka-Lasota B., 2017. Wpływ ruchu samochodowego oraz warunków meteorologicznych na stężenia tlenków azotu, *Autobusy Technika, Eksploatacja, Systemy Transportowe* 12/2017 p.196-200 (MNiSW – 7 pkt, IF=0, udział 90%)

Mój udział w pracy [A6] polegał na opracowaniu koncepcji modelu, wykonaniu obliczeń, analizie statystycznej, interpretacji wyników, sformułowaniu wniosków i redakcji tekstu.

Łączna wartość bibliometryczna publikacji stanowiących przedstawione powyżej osiągnięcie naukowe wynosi **126** punktów wg listy Ministerstwa Nauki i Szkolnictwa Wyższego, a sumaryczny Impact Factor według listy JCR wynosi **10,690**.

Oświadczenia współautorów pracy [A6] wraz z określeniem ich indywidualnego wkładu pracy stanowi załącznik 6.

Kopie publikacji wchodzących w skład przedstawionego cyklu znajdują się w załączniku 5.

Oznaczenia porządkowe publikacji [A1] – [A6] w dalszej części referatu stanowią odnośniki bibliograficzne.

2.3. Cel naukowy pracy oraz osiągnięte wyniki

Wstęp

Zanieczyszczenie powietrza w aglomeracji miejskiej stanowi bardzo poważny problem zdrowotny, społeczny oraz ekonomiczny. Nienaturalnie wysokie stężenia tlenków azotu (głównie NO₂ i PM_{2,5}) w powietrzu mają negatywny wpływ na zdrowie ludzi wywołując głównie dysfunkcje ze strony układu oddechowego oraz krwionośnego [Hien i in., 2016; Hoek i in., 2013]. Badania wskazują, że zanieczyszczenie powietrza może być również przyczyną autyzmu u dzieci [Flores-Pajot i in., 2016], choroby Parkinsona [Pei Chen i in., 2016], a w efekcie ich następstw nawet śmierci [Tang i in., 2017]. Według Światowej Organizacji Zdrowia [WHO, 2012] w 2012 roku na całym świecie 3,7 miliona zgonów było spowodowanych zanieczyszczeniem powietrza, z czego 480 tys. w Europie.

Źródła zanieczyszczeń powietrza tlenkami azotu są różne. Wśród najbardziej szkodliwych pochodzenia antropogenicznego wymienia się emisję gazów, spalinowych głównie z wysokotemperaturowego spalania w silnikach pojazdów samochodowych, jak również ze spalania paliw w procesie produkcji energii. W transporcie drogowym największymi emiterami dwutlenków azotu oraz cząstek stałych PM_{2,5} są silniki z zapłonem samoczynnym (Diesla). Według Europejskiej Agencji Środowiska 39% tlenków azotu pochodzi z transportu drogowego [EEA Report 9/2017]. Mimo spadku emisji tego zanieczyszczenia o połowę w porównaniu z rokiem 1990, w 2015 roku dla NO₂, na większości komunikacyjnych stacji pomiarowych w Europie (89 %) przekroczona została roczna dopuszczalna wartość średnia wynosząca 40 µg/m³ [EEA, 2017]. Gęsta i wysoka zabudowa miast modyfikuje temperaturę oraz wilgotność powietrza, kierunek wiatru, jak również jego siłę, utrudniając ewakuację zanieczyszczeń. Kumulacja tych niekorzystnych zjawisk pociąga za sobą, oprócz konsekwencji zdrowotnych, obniżenie komfortu życia w mieście. Władze, zarówno na szczeblu lokalnym jak i państwowym oraz międzynarodowym, dokładają starań aby zredukować i to zminimalizować zjawisko.

Rozważane są różne scenariusze zmian oraz ich konsekwencje społeczno-ekonomiczne. Modelowanie stężeń zanieczyszczeń w długim horyzoncie czasowym umożliwia określenie udziału poszczególnych rodzajów emisji w zanieczyszczeniu oraz prognozowanie zmian wynikających z realizacji rozważanych scenariuszy.

Modele opisujące zależności pomiędzy stężeniami zanieczyszczeń w powietrzu a czynnikami meteorologicznymi, czasowymi oraz natężeniem ruchu pojazdów dzielą się w literaturze na dwa zasadnicze nurty ze względu na położenie w czasie: aktualne (punktowe) oraz wsteczne (przedziałowe). Modele przedziałowe wymagają i wykorzystują wartości stężeń zanieczyszczeń z pewnego wstecznego przedziału czasu ($t - 1$, $t - 2$, $t - 24$, $t - 48$) do określenia wartości przewidywanej w chwili t . Jest to dość dobrze rozpoznane zagadnienie ze skutecznością przewidywania ponad 90%. Modele punktowe natomiast mają za zadanie określić wartość stężenia zanieczyszczenia na podstawie danych z danej chwili t . Zadaniem modeli punktowych jest przewidywanie wartości stężeń zanieczyszczeń przy zaistnieniu z góry określonych warunków (wartości zmiennych niezależnych). Pozwalają one ocenić wpływ zmiany wartości zmiennej/zmiennych niezależnych na stężenie zanieczyszczeń, a tym samym umożliwiają przewidywanie skutków modyfikacji wartości predyktorów. Wśród punktowych modeli można wyróżnić kilka podstawowych metod modelowania matematycznego. Wciąż popularne są modele regresyjne, które znacznie ewoluowały dzięki rozwojowi technik obliczeniowych [Ping Shi i Harrison, 1997; Aldrin i Haff, 2005; Zhang i in., 2015; Battista i de Lieto Vollaro, 2017]. Bardziej zaawansowane obliczeniowo są modele oparte na uczeniu maszynowym. Zastosowanie znalazły między innymi sztuczne sieci neuronowe [Elangasinghe i in., 2014; Nejadkoorki i Baroutian, 2012] również w połączeniu z modelem regresji wielokrotnej [Papanastasiou i in., 2007]. Coraz częściej w literaturze modelowania stężeń zanieczyszczeń spotyka się zastosowanie metod opartych na drzewach losowych. Stosowane są tu zarówno pojedyncze drzewa losowe [Singh i in., 2013] jak i ich bardziej złożone kompilacje: las losowy - Random Forest – RF [Laña i in., 2016] oraz wzmacniane drzewa regresyjne - Boosted Regression Trees -BRT [Sayegh i in., 2016]. Metodykę lasu losowego RF zastosowałam w pracach [A1], [A2], [A4], metodykę drzew wzmacnianych BRT w pracach [A4], [A6]

Modele deterministyczne, w szczególności te nie odwołujące się do wartości zmiennych z chwil poprzednich (punktowe), obarczone są dużym błędem i główna niedogodność w ich stosowaniu wynika ze słabego dopasowania do danych a co za tym idzie dużych błędów w prognozach. Fakt ten wynika z bardzo dużego zróżnicowania wartości stężeń oraz znacznej liczby czynników na nie wpływających. Alternatywą dla modeli deterministycznych są modele probabilistyczne, które pozwalają przewidywać wartości stężeń bądź ich przekroczenia z zadaniem prawdopodobieństwem [Aznarte, 2017; Balashov i in., 2017]. Własną propozycję modelu probabilistycznego przedstawiłam w pracy [A3].

Główne cele naukowe

Moje rozważania w zakresie opisu i modelowania zależności między stężeniami zanieczyszczeń w powietrzu a czynnikami otoczenia dotyczą modeli opisanych we wstępie jako punktowe, w których przewidywanie wartości aktualnych odbywa się na podstawie znajomości aktualnych wartości zmiennych objaśnianych (czynników otoczenia).

Główne cele pracy naukowej zawartej w przedstawionym w części 2.2. cyklu publikacji są następujące:

1. Rozpoznanie struktury zależności między stężeniem zanieczyszczeń powietrza pochodzenia antropogenicznego na terenie aglomeracji miejskiej a czynnikami otoczenia dla różnych okresów w roku, z zastosowaniem różnych metod opisu matematycznego.
2. Opracowanie metodyki opisu zależności między stężeniami zanieczyszczeń a czynnikami otoczenia, ze znacznie większym niż w dotychczasowej literaturze poziomem dopasowania do danych rzeczywistych.
3. Ocena skuteczności metod opisu zależności między stężeniami zanieczyszczeń a czynnikami otoczenia, opartych na drzewach losowych: regresyjnych drzew wzmacnianych (Boosted Regression Trees - BRT) oraz lasów losowych (Random Forest - RF).
4. Opracowanie prostego, klastrowego modelu probabilistycznego do prognozowania stężeń zanieczyszczeń powietrza na podstawie czynników otoczenia.

Cele główne osiągnęłam poprzez realizację celów szczegółowych:

- 1) Znalezienie poligonu badawczego, na którego przykładzie będę prowadzić eksperymenty obliczeniowe. Musiał to być obszar dużego miasta: skrzyżowanie lub arteria komunikacyjna, objęta monitoringiem jakości powietrza w zakresie stężeń tlenków azotu i pyłów zawieszonych oraz natężenia ruchu pojazdów z częstotliwością nie większą niż godzinową.
- 2) Dobór zbioru dostępnych czynników otoczenia: meteorologicznych, natężenie ruchu oraz czasowych mających największy wpływ na stężenia rozważanych zanieczyszczeń.
- 3) Wybór kilku metod opisu zależności ze zwróceniem uwagi na założenia poszczególnych metod. Przeprowadzenie badania zależności stężeń wybranych zanieczyszczeń i czynników otoczenia z wykorzystaniem wybranych metod obliczeniowych. Ocena jakości dopasowania modeli.
- 4) Rozpoznanie dodatkowych czynników wpływających na opisywane zależności. Ocena wpływu pory roku na wskaźniki ilościowe zidentyfikowanych relacji biorąc pod uwagę podział roku na 2 lub 4 części pod względem meteorologicznym.
- 5) Opracowanie autorskiej metody opisu zależności stężeń zanieczyszczeń od czynników otoczenia posiadającej znaczącą przewagę nad dotychczas stosowanymi. Przeprowadzenie jej weryfikacji oraz identyfikacja ograniczeń.

- 6) Opracowanie alternatywnej do przedstawionej w 5) metody przewidywania stężeń zanieczyszczeń na podstawie chwilowych wartości czynników otoczenia.

Ad.1)

Przedstawione powyżej cele naukowe realizowałam poprzez rozpoznanie struktury zjawiska od podstaw na przykładzie miasta Wrocławia, w którym mieszkam i pracuję. Przez Wrocław przepływa 5 rzek i wiele starorzeczy oraz kanałów, dla których wybudowano ponad 100 mostów oraz 30 kładek. Uwarunkowania te sprawiły, że układ drogowy w mieście ma charakter promienisty i wciąż jest dostosowywany do rosnącego ruchu pojazdów. We Wrocławiu jest aktualnie zarejestrowanych 270 tys. Pojazdów, z czego 200 tys. stanowią samochody osobowe. Struktura drogową w połączeniu ze znaczną liczbą pojazdów poruszających się po mieście skutkują powstawaniem kongestii, co zwiększa emisję spalin z poruszających się pojazdów. Według Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu 56% emisji NO_2 , 44% emisji CO oraz 16% emisji $\text{PM}_{2,5}$ pochodzi z transportu drogowego [Informacja o jakości powietrza...]. Znacznie mniejszy niż w innych krajach udział spalin samochodowych w emisji szkodliwych substancji, w szczególności $\text{PM}_{2,5}$, jest efektem struktury mieszkaniowej miasta. Historyczne losy miasta sprawiły, że na jego terenie znajduje się wiele stuletnich domów i kamienic, do dziś opalanych paliwami stałymi (węgiel, drewno).

We Wrocławiu znajduje się 5 aktywnie działających stacji pomiaru jakości powietrza, administrowanych przez Wojewódzki Inspektorat Ochrony Środowiska we Wrocławiu [A5]. Pomiar NO_2 oraz NO_x są prowadzone w sposób automatyczny, w badanym okresie, na 3 stacjach, a $\text{PM}_{2,5}$ na dwóch. Jedna z nich posiada status stacji komunikacyjnej miejskiej. Położona jest w kanionie komunikacyjnym, w bezpośrednim sąsiedztwie dużego skrzyżowania ulic Hallera – Powstańców Śląskich. Na stacji tej pomiary automatyczne w zakresie stężeń tlenków azotu (NO , NO_2 , NO_x), tlenku węgla oraz pyłu zawieszanego $\text{PM}_{2,5}$ zapisywane są z częstotliwością godzinową.

We Wrocławiu aktywnie działa system ITS (Intelligent Transport System) zarządzany przez Centrum Zarządzania Ruchem i Transportem Publicznym Zarządu Dróg i Utrzymania Miasta we Wrocławiu. W ramach systemu na terenie całego miasta zainstalowanych jest 921 kamer z wideodetekcją, które jako jedną z informacji zbierają liczbę pojazdów przekraczających przekrój pasa/pasów jezdni. Skrzyżowanie Hallera – Powstańców Śląskich jest objęte monitoringiem systemu ITS. Natężenie ruchu zapisywane jest dla kwantów piętnastominutowych. Na potrzeby zachowania jednakowej częstotliwości zbioru danych sumowałam liczby pojazdów z 4 kolejnych kwadransów każdej godziny. Z literatury wynika, że stężenia zanieczyszczeń emitowanych przez transport samochodowy utrzymują się na niemal niezmiennym poziomie w odległości do 50m od źródła zanieczyszczeń [Padró-Martínez i in., 2012]. Opisana wyżej stacja pomiarowa jakości powietrza znajduje się 30 m od centrum skrzyżowania objętego monitoringiem. Uznałam zatem, że jest to właściwy poligon badawczy do badania wpływu czynników meteorologicznych oraz ruchu samochodowego na zanieczyszczenie powietrza w kanionie komunikacyjnym aglomeracji miejskiej.

Ad.2)

Informacje opisujące warunki meteorologiczne również muszą być mierzone co godzinę, aby zachować jednorodną częstotliwość pomiarów z natężeniem ruchu oraz pomiarami stężeń zanieczyszczeń. Uzyskałam dostęp do danych meteorologicznych z automatycznej stacji pomiarowej Instytutu Meteorologii i Gospodarki Wodnej znajdującej się na terenie lotniska Strachowice we Wrocławiu. Obejmują one godzinowe wartości temperatury powietrza, prędkości wiatru, kierunku wiatru, ciśnienia atmosferycznego oraz wilgotności względnej powietrza. Kierunek wiatru pierwotnie był w formie liczbowej ciągłej, ale uznałam, że niepoprawne jest stosowanie miary stopniowej ponieważ w rzeczywistości blisko siebie występujące kąty jak 1° , 360° miałyby odległe od siebie wartości. Z tego powodu skategoryzowałam kierunek wiatru na 8 kategorii co 45° (N, NE, E, itd.).

Warunki klimatyczne Wrocławia wykazują cechy typowe dla dużych aglomeracji miejsko-przemysłowych, w których wpływ gospodarczej działalności przejawia się zmianą cech fizycznych podłoża, zanieczyszczeniem atmosfery, a także emisją ciepła sztucznego, wytwarzanego w procesach komunalnych i przemysłowych. Średnia temperatura roczna we Wrocławiu wynosi $9,1^\circ\text{C}$ i jest to najcieplejsze miasto w Polsce. Rozważane lata 2015 – 2017 nie należały do wyjątkowych pod względem meteorologicznym.

Natężenie ruchu wykazuje dwumodalną zmienność dobową ze szczytami: porannym w godzinach 7-8 oraz popołudniowym w godzinach 15-17. Maksymalne odnotowane natężenie ruchu to 7797 pojazdów przejeżdżających przez skrzyżowanie w ciągu godziny, a minimalne 76 pojazdów. Przeciętnie natężenie ruchu w godzinach od 7-19 oscyluje w granicach od 4000 do 5000 pojazdów.

Stężenie tlenków oraz dwutlenku azotu wykazuje jeszcze wyraźniejszą dwumodalną zmienność dobową ze szczytami opóźnionymi w stosunku do największego natężenia ruchu: poranny 7-9, popołudniowy 18-21. W Polsce dopuszczalna roczna wartość średnia stężenia NO_2 wynosząca $40 \mu\text{g}/\text{m}^3$ została przekroczona na rozważanym skrzyżowaniu w każdym z lat okresu 2015-2017 wynosząc odpowiednio: 53,8; 49,2; $48,1 \mu\text{g}/\text{m}^3$. W ciągu omawianych 3 lat średnia wartość dobowa stężenia NO_2 w powietrzu przekroczyła wartość $40 \mu\text{g}/\text{m}^3$ w 873 dniach (80% dni). Dopuszczalne godzinowe stężenie NO_2 w Polsce wynosi $200 \mu\text{g}/\text{m}^3$, wartość ta nie może być przekroczona więcej niż 18 razy w roku. Poziom alarmowy stężenia NO_2 wynosi $400 \mu\text{g}/\text{m}^3$ dla trzech kolejnych godzin [Rozporządzenie Ministra Środowiska z dnia 24 sierpnia 2012 r. w sprawie poziomów niektórych substancji w powietrzu - Dz.U.2012.1031]. W okresie 2015-2017 poziom dopuszczalny stężenia NO_2 został przekroczony dwukrotnie w 2015 roku, natomiast poziom alarmowy nie został przekroczony. W zakresie tlenków azotu (NO_x) wspomniane rozporządzenie określa jedynie poziom dopuszczalny średnioroczny ze względu na ochronę roślin wynoszący $30 \mu\text{g}/\text{m}^3$. Zarówno natężenie ruchu jak i stężenia NO_2 i NO_x są zauważalnie niższe w weekendy niż w dni pracujące. Stężenia nie wykazują zmienności miesięcznej [A1], [A2], [A6].

Powyższe rozporządzenie nakłada obowiązek utrzymania dopuszczalnego średniego rocznego stężenia $\text{PM}_{2,5}$ na poziomie nie przekraczającym $20 \mu\text{g}/\text{m}^3$. W latach 2015, 2016

wartości te dla stacji Hallera, znajdującej się w ścisłym centrum miasta w tunelu komunikacyjnym, wynosiły odpowiednio 30,3 oraz 27,5 $\mu\text{g}/\text{m}^3$. Stężenia pyłów zawieszonych $\text{PM}_{2,5}$ w powietrzu przyjmują wyraźnie większe wartości w miesiącach zimowych. Stężenia pyłów zawieszonych $\text{PM}_{2,5}$ nie wykazują istotnej zmienności dobowej ani tygodniowej. Wykazują natomiast istotną zmienność miesięczną w związku z bezpośrednim wpływem nasilenia emisji z procesu spalania paliw stałych do ogrzewania domów i mieszkań w miesiącach zimowych [A2].

Szczegółową analizę zmienności stężeń pyłów $\text{PM}_{2,5}$ oraz PM_{10} w latach 2012-2016 we Wrocławiu przedstawiłam w pracy [A5] wykazując, że w okresach zimowych omawianych lat poziom smogu identyfikowany ze stężeniem pyłów zawieszonych był porównywalnie wysoki. Średnie wartości dobowe za okres X-IV wynosiły od 34 do 51,5 $\mu\text{g}/\text{m}^3$ sięgając maksymalnie 249 $\mu\text{g}/\text{m}^3$ w jednej godzinie.

Ad.3)

Ze względu na dużą liczbę danych (ponad 17000 przypadków dla lat 2015-2016, ponad 26000 dla lat 2015-2017) rozważałam głównie metody statystyczne z grupy data mining. Ze względu na regresyjny charakter badanych zależności oraz występowanie zarówno zmiennych ilościowych (natężenie ruchu, temperatura powietrza, prędkość wiatru, ciśnienie atmosferyczne, wilgotność względna powietrza) jak i jakościowych (kierunek wiatru N, NE, E itd., dzień tygodnia, miesiąc, dzień wolny od pracy) zdecydowałam się skupić na metodach opartych na teorii drzew decyzyjnych. W prezentowanym cyklu prac zastosowane zostały trzy metody modelowania zależności wielowymiarowych.

Losowy las – Random forest (RF)

Losowy las (RF) składa się z zadanej liczby prostych drzew decyzyjnych. Każde z drzew składowych tworzących RF korzysta z losowego podzbioru wybranego ze zbioru dostępnych danych. Podzbiory te są niezależne i ten sam przypadek może występować w kilku podzbiorach (losowanie ze zwracaniem). Dla każdego drzewa wybierany jest zadanej liczności podzbiór zmiennych niezależnych (predyktorów) z równym prawdopodobieństwem. Każde drzewo jest uczone na innym podzbiorze danych wejściowych. Przewidywana wartość zmiennej zależnej wyznaczana jest jako średnia arytmetyczna ze wszystkich drzew składowych lasu (RF). Metoda ta, łącząca koncepcję upakowania i losowego wyboru cech, ma na celu poprawę wydajności w porównaniu z innymi algorytmami uczenia maszynowego i modelami regresji liniowej [Archer i Kimes, 2008]. Metoda lasu losowego daje możliwość określenia ważności predyktorów, czyli ich wpływu na wartości zmiennej zależnej. Ważność zmiennych predykcyjnych wyrażana jest jako suma - po wszystkich węzłach drzewa - wzrostów oceny resubstytucji (ΔR). Wartości ważności zmiennej obliczane są jako ułamek maksymalnej sumy (po wszystkich zmiennych, wyrażone w procentach). Oznacza to, że najważniejsza zmienna (o najwyższej sumie resubstytucji) ma przypisaną ważność 100. Ważność pozwala zidentyfikować

zmiennie, które mają istotną moc predykcyjną względem zmiennej zależnej [Breiman i in., 1984].

Metodykę lasów losowych zastosowałam w pracach [A1], [A2] oraz [A4].

Drzewa wzmacniane

Główną ideą wzmacnianych drzew regresyjnych jest tworzenie ciągu decyzyjnych drzew binarnych składających się z korzenia i dwóch potomków (jeden podział), z których każde kolejne drzewo jest zbudowane do predykcji reszt generowanych przez poprzednie [Friedmann 1999a, 1999b]. W metodzie tej budowane są drzewa binarne, podział danych na dwa podzbiory, w każdym węźle podziału. W kolejnych krokach wzmacniania algorytmu określany jest pojedynczy (najlepszy) podział danych i obliczane są odchyłki wartości obserwowanych od średnich (reszty w każdym podziale). Zastosowana w algorytmie metoda stochastycznego wzmacniania gradientowego oznacza, że każde kolejne drzewo budowane jest na podstawie losowej próby zawierającej 50% całego zbioru danych. Zatem kolejne drzewa tworzone są do predykcji reszt w próbach wylosowanych niezależnie. Wprowadzenie losowości do analizy ma na celu zabezpieczenie przeciwko przeuczeniu i zwraca modele, które mają własność generalizacji i dobrą trafność predykcyjną. Opisany algorytm daje w efekcie dobre dopasowanie wartości przewidywanych do wartości obserwowanych, nawet jeśli sama natura relacji pomiędzy predyktorami a zmienną zależną jest bardzo złożona (np. nieliniowa). Zastosowanie drzew decyzyjnych z metodą podziału C&RT wyczerpującego poszukiwania podziałów jednowymiarowych, umożliwia ocenę ilościową ważności zmiennych analogicznie jak dla lasów losowych.

Metodykę drzew wzmacnianych zastosowałam w pracach [A4] oraz [A6].

Krokowa regresja wielowymiarowa

Wielowymiarowa regresja liniowa służy do ilościowego ujęcia związków pomiędzy wieloma zmiennymi niezależnymi (objaśniającymi, predyktorami) a zmienną zależną (objaśnianą). Zakłada ona liniowy związek pomiędzy zmienną objaśnianą a objaśniającymi. Współczynniki liniowego równania regresji wyznacza się metodą najmniejszych kwadratów minimalizując kwadraty błędów dopasowania modelu do danych empirycznych. Metoda regresji wielowymiarowej wymaga aby wszystkie zmienne były ilościowe.

Jako pierwsze przeprowadziłam badanie wpływu czynników otoczenia na stężenia NO_2 oraz NO_x z zastosowaniem regresji liniowej wielowymiarowej oraz BRT [A6]. W modelach regresji liniowej wielowymiarowej wybrałam zbiór 5 zmiennych objaśniających składający się wyłącznie ze zmiennych ilościowych: natężenie ruchu, prędkość wiatru, temperatura powietrza, ciśnienie atmosferyczne i wilgotność względna powietrza. Modele BRT budowałam dla maksymalnego dostępnego zbioru zmiennych objaśniających (9 zmiennych) obejmującego dodatkowo następujące zmienne: dzień wolny/pracujący, dzień

tygodnia, miesiąc oraz kierunek wiatru. Obliczenia przeprowadziłam na podstawie godzinowych wartości wymienionych zmiennych z lat 2015-2016. Celem bardziej szczegółowego rozpoznania wpływu każdej ze zmiennych na stężenia NO_2 wyznaczyłam dla obu metod modele w pięciu wariantach: cały okres 2015-2016, lato (połączone okresy VI-VIII), jesień (połączone okresy IX-XI), zima (XI-II) oraz wiosna (III-V). Z przeprowadzonych obliczeń wynika, że największy wpływ na stężenia obu rodzajów zanieczyszczeń w powietrzu ma natężenie ruchu oraz prędkość wiatru. Regresja krokowa wskazuje natężenie ruchu jako zmienną wyjaśniającą największą część zmienności (do 36% dla NO_2 oraz do 29% dla NO_x w okresie zimowym), natomiast metoda drzew wzmacnianych największą wagę przypisuje prędkości wiatru. Jako drugie co do ważności zmienne występują te same naprzemiennie. Na podstawie współczynników regresji wielowymiarowej stwierdziłam, że wzrost natężenia ruchu o 1000 pojazdów powoduje wzrost stężenia NO_2 o $10 \mu\text{g}/\text{m}^3$.

Do porównania modeli między sobą zastosowałam kryterium Bayesa BIC. Wartości tego kryterium są dla obu rodzajów modeli mniejsze przy modelowaniu stężeń NO_2 , a większe przy modelowaniu stężeń NO_x . Kryterium BIC zależy nie tylko od wartości reszt (różnica między wartością empiryczną, a modelową) ale również od liczby zmiennych w modelu. W modelach z zastosowaniem drzew wzmacnianych użyto 9 zmiennych, podczas gdy w modelach regresyjnych liniowych 5 zmiennych. W okresach letnich oraz wiosennych lepsze dopasowanie mierzone wskaźnikiem BIC otrzymałam dla modeli drzew wzmacnianych (BRT). Dla całego okresu 2015-2016 lepsze dopasowanie wykazał model regresji liniowej wielowymiarowej. Dla okresów jesiennego oraz zimowego preferencje wskazywane przez BIC są rozbieżne dla różnych zanieczyszczeń. Trudno zatem jednoznacznie rozstrzygnąć o preferowaniu jednego z opisywanych modeli. Wyższe wskaźniki współczynnika determinacji dla modeli drzew wzmacnianych wynikają z większej liczby parametrów tych modeli. Koszt zwiększenia liczby parametrów jest jednak porównywalny ze wzrostem dobroci dopasowania dzięki nim otrzymanej.

Ad. 4)

W kolejnym etapie badań przeprowadziłam rozważania nad zależnością stężeń NO_2 , NO_x oraz $\text{PM}_{2,5}$ od czynników otoczenia z zastosowaniem lasów losowych – RF [A2]. Zauważyłam, że wpływ poszczególnych czynników otoczenia zmienia się w zależności rozważanego okresu czasu. Określiłam 9 okresów następująco:

- pełny zakres danych – lata 2015-2016 - [1] (17598 przypadków)
- połączone półrocza ciepłe (IV-IX) 2015, 2016 - [2] (8774 przypadki)
- połączone półrocza chłodne (X-III) 2015, 2016 - [3] (8724 przypadki)
- dni pracujące - [4] (12121)
- dni wolne od pracy - [5] (5477)
- wiosna (III-V) - [6] (4412)
- lato (VI-VIII) - [7] (4407)

- jesień (IX-XI) - [8] (4357)

- zima (XII-I) - [9] (4323)

Do tworzenia drzew decyzyjnych wzięłam pod uwagę 9 zmiennych objaśniających pogrupowanych w obszary tematyczne:

- natężenie ruchu,
- zmienne czasowe (dzień tygodnia, dni świąteczne, miesiąc),
- czynniki meteorologiczne (temperaturę powietrza [$^{\circ}\text{C}$], prędkość wiatru [m/s], kierunek wiatru, wilgotność względna [%], ciśnienie powietrza [hPa]).

Zmienna „dni świąteczne” identyfikuje dni będące świętami państwowymi bądź niedzielami.

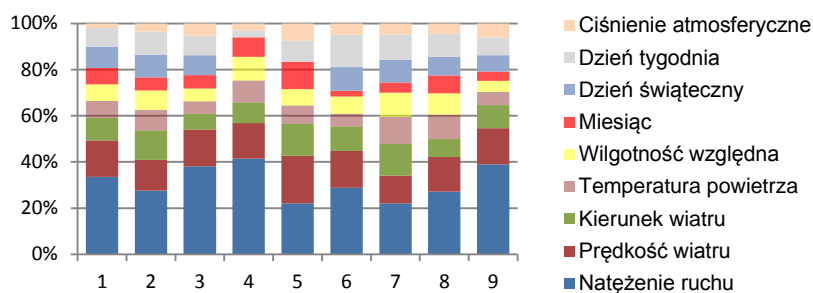
Modelowałam stężenia jednego z trzech zanieczyszczeń powietrza: NO_2 , NO_x , $\text{PM}_{2,5}$.

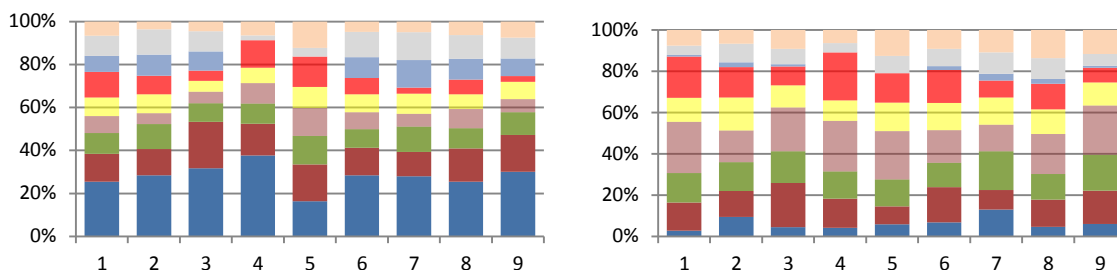
Na podstawie analizy przebiegu wartości błędów zdecydowałam się zbudować lasy składające się z 25 drzew. Próba ucząca stanowi 50% wszystkich przypadków, próba testowa 30%. Do budowy każdego drzewa losowanych jest 5 zmiennych objaśniających ze zbioru 9 zmiennym opisanych wyżej. Wyjątkiem są lasy losowe budowane dla okresów [4] i [5]. Dla nich usunięto ze zbioru zmiennych „dni świąteczne” ze względu na zbyt małą zmienność wartości. W związku z redukcją liczby zmiennych do 8, zredukowano tam również liczbę predyktorów do 4.

Do oceny dopasowania każdego z modeli wykorzystałam następujące współczynniki jakości dopasowania: R^2 , MFB, MAD, MAPE. Nie brałam pod uwagę popularnych kryteriów informacyjnych jak BIC lub AIC ze względu na z góry zadaną i stałą liczbę zmiennych w modelu. Współczynnik R^2 jest jedną z podstawowych miar jakości dopasowania modelu. Przyjmuje wartości w zakresie $\langle 0,1 \rangle$. Im wartość R^2 jest bliższa 1 tym wartości estymowane zmiennej objaśnianej mniej się różnią od wartości empirycznych. Innymi miarami dopasowania, niezależnymi od wartości średniej, są MAD – *Mean Absolute Deviation Error*, MAPE – *Mean Absolute Percentage Error* oraz MFB – *Mean Fractional Bias*. MFB jest miarą rekomendowaną w literaturze dla celów analizy stężeń zanieczyszczeń (Boylan i Russel, 2006) ponieważ mierzy tendencje modelu do niedoszacowania lub przeszacowywania. MAD określa przeciętny błąd bezwzględny, czyli średnią różnicę między wartością empiryczną, a modelową. Podobną miarą jest MAPE, jednak w odróżnieniu od MAD określa on przeciętny błąd względny. Współczynnik R^2 waha się dla 27 wyznaczonych modeli od 0,26 do 0,57. Błąd względny oscyluje pomiędzy 73,3%, a 29,3%. Najlepsze okazało się dopasowanie modeli przy przewidywaniu stężeń cząstek stałych $\text{PM}_{2,5}$.

Do oceny wpływu poszczególnych czynników otoczenia na stężenia zanieczyszczeń wyznaczyłam względne ważności zmiennych (*Share of Validity* – SoV), obliczając je jako procentowy udział w całkowitej sumie ważności w danym modelu. Dzięki temu zabiegowi widoczne jest zjawisko względnej ważności predyktorów w modelu, a zatem ich struktura i wkład w proces wyjaśniania zmienności zmiennej zależnej/objaśnianej. Skutkiem ubocznym zabiegu jest zakamufłowanie informacji, która zmienna jest najważniejsza. Zatem dla pełnego obrazu ważności predyktorów w modelu należy równolegle analizować

wartości ważności jak i zaproponowany wyżej ich procentowy rozkład udziałów (SoV). W modelowaniu stężeń tlenków azotu najważniejszą zmienną jest natężenie ruchu, oprócz modelu [5] gdzie najważniejsza okazała się prędkość wiatru (ważność 100), natężenie ruchu miało wtedy ważność 95. Kolejne co do ważności są charakterystyki wiatru: prędkość i kierunek (Rys.1.). Najmniej istotną zmienną w 5 modelach na 9 było ciśnienie powietrza. Widoczna jest różnica w strukturze ważności zmiennych między modelami [2], [7] oraz [3] i [9]. Pierwsze dwa dotyczą okresów ciepłych. Cechuje je w miarę wyrównany podział udziałów ważności zmiennych: względna ważność - SoV najważniejszej zmiennej natężenie ruchu wynosi 27 i 22% odpowiednio, a najmniej ważnych: miesiąc i ciśnienie powietrza 9 i 13%. Odmiennie przedstawia się SoV dla okresów chłodnych [3] i [9], gdzie również zdecydowanie najważniejsze jest natężenie ruchu zajmując 38% i 39% udziałów odpowiednio następnie, z o ponad połowę mniejszym SoV, prędkość wiatru (po 16%) przy niemal równym wpływie pozostałych zmiennych (od 4% do 9%). Najbardziej równomierny wpływ predyktorów występuje w modelu o najmniejszej liczebności [5] - dni wolnych od pracy (od 6 do 22%), który znacząco wyróżnia się w ważności predyktorów w modelowaniu zarówno NO₂ jak i NO_x. Obserwuje się tam większy wpływ zmiennych innych niż natężenie ruchu na stężenia tlenków azotu, co należy rozumieć jako większe znaczenie pozostałych zmiennych względem ruchu samochodowego w generowaniu zanieczyszczeń niż dla innych ciągów pomiarowych. W modelowaniu NO_x zaznaczają się modele [4] i [5] cechujące się największą względną ważnością dla ciśnienia atmosferycznego, a najmniejszym dla dnia tygodnia ze wszystkich rozważanych modeli. Małe znaczenie dnia tygodnia jest oczywiste z racji małego zróżnicowania wartości tej zmiennej objaśnianej szczególnie w modelu [5] - dni wolne od pracy. Modele [4] i [5] cechuje również odpowiednio największe (38%) oraz najmniejsze (17%) SoV dla natężenia ruchu mimo, że w pierwszym z modeli ważność zmiennej traffic wynosi 100, a w drugim 95. Wynika to z bardziej wyrównanej ważności innych zmiennych w modelu [5], co zwiększa sumę ważności wszystkich zmiennych, obniżając wartość SoV.

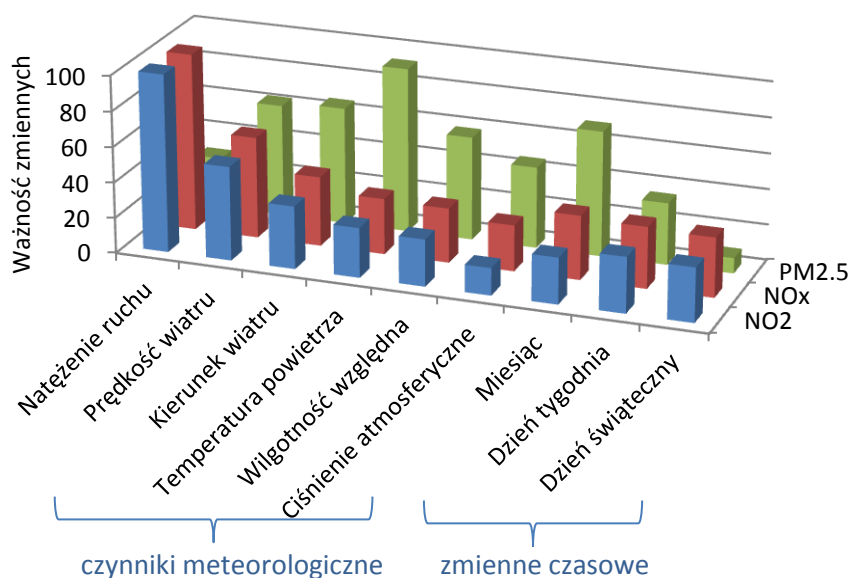




Rysunek 1. Ważność względna (SoV) dla poszczególnych podzbiorów czasowych [1] – [9] oraz zanieczyszczeń: NO₂ – u góry, NO_x – u dołu po lewej, PM_{2,5} – u dołu po prawej

Zdecydowanie inaczej przedstawia się wpływ zmiennych objaśniających na stężenia pyłów zawieszonych PM_{2,5}. Tu największy wpływ ma temperatura powietrza, następnie prędkość i kierunek wiatru oraz miesiąc. Natężenie ruchu wykazuje największe SoV w modelowaniu PM_{2,5} dla modeli tworzonych dla okresów ciepłych [2] i [7] (10% i 13% odpowiednio). Zjawisko to związane jest ze zdecydowanie mniejszą emisją pyłów z sektora komunalno-bytowego w okresach ciepłych niż w pozostałe części roku. Dla pozostałych okresów stężenie pyłów zmienia się w sposób zdecydowanie mniej zależny od natężenia ruchu, zatem ważność tej zmiennej jest niższa.

Uśrednioną ważność przedstawiłam na rysunku 2. Ważność uśredniono dla wszystkich modeli celem zobrazowania różnic w ważności zmiennych dla różnych rodzajów zanieczyszczeń. Widoczne jest podobieństwo w znaczeniu zmiennych objaśniających w modelowaniu stężeń tlenków azotu i opisana wyżej różnica pomiędzy pyłami zawieszonymi PM_{2,5}. Analogiczny wykres z podziałem na pory roku znajduje się w pracy [A2].



Rysunek 2. Ważność zmiennych w modelach RF dla rozważanych zanieczyszczeń

Ad. 5)

W modelowaniu stężeń zanieczyszczeń na podstawie otaczających warunków (bez uwzględnienia zależności wstecznej z opóźnieniem) zasadniczym problemem są niskie miary dopasowania modeli. Na podstawie ogólnoswiatowych danych z 5220 stacji pomiarowych umieszczonych na wszystkich kontynentach z zastosowaniem informacji o użytkowaniu terenu, Larkin i in. [2017] otrzymali dopasowanie modelu dla NO_2 mierzone R^2 równe 0,52. Oznacza to, że jedynie 52% zmienności stężeń zanieczyszczeń jest wyjaśniane przez zmienność wartości predyktorów. Sayed i in. [2016] wyznaczyli 112 modeli określających stężenia pięciu zanieczyszczeń powietrza dla 4 różnych zestawów predyktorów. W zakresie rozważanych przeze mnie zanieczyszczeń oraz zmiennych objaśniających obejmujących warunki meteorologiczne, czasowe i natężenie ruchu, otrzymali oni współczynniki R^2 od 0,49 do 0,54 dla dwutlenku azotu, od 0,37 do 0,48 dla PM_{10} oraz od 0,33 do 0,44 dla tlenku azotu. Ja otrzymałam w pracy [A2] wariancję wyjaśnianą przez model R^2 od 0,31 do 0,58. Niskie wartości R^2 wynikają ze znacznego zróżnicowania wartości zmiennych objaśnianych, oddziaływania nieznanymi czynnikami oraz trudności w modelowaniu z zastosowaniem jednego modelu, zarówno typowych, jak i ekstremalnych wartości zanieczyszczeń. W odpowiedzi na tak zidentyfikowany problem, zaproponowałam modyfikację ogólnego modelu, określającego zależność między stężeniami zanieczyszczeń w powietrzu a występującymi w tym czasie czynnikami otoczenia w taki sposób, aby poprawić jego dokładność bez utraty ogólności. Znane i opisywane w literaturze jest rozwiązanie cechujące się dokładnością modelu do $R^2=0.92$ [Catalano i in., 2016], przy czym jest to model prognostyczny oparty głównie na autokorelacji, zawierający wśród predyktorów wartości stężeń zanieczyszczeń z poprzedniej chwili w czasie x_{t-1} . W omawianym zagadnieniu poszukiwana jest ogólna zależność dla dowolnej chwili w czasie, bez odnoszenia się do wartości poprzednich zmiennej objaśnianej. Porównanie modeli RF z BRT, przedstawione w pracy [A4] wskazało, że modele RF lepiej dopasowują się do wartości wysokich, natomiast gorzej do niskich. W poszukiwaniu rozwiązania tego problemu postanowiłam dokonać podziału całości zbioru danych empirycznych na wartości „niskie” („Lower”) oraz „wysokie” („Upper”) i modelowanie każdego z podzbiorów niezależnie. Nowy model, ze względu na jego strukturę, nazwałam modelem dzielonym – Partition Model.

Mimo możliwości wszechstronnego zastosowania metody lasu losowego (RF) nie opisuje ona w zadawalający sposób skomplikowanej relacji między stężeniami zanieczyszczeń i czynnikami otoczenia w całym zakresie występowania stężeń zanieczyszczeń. Konieczność jednorazowego określenia parametrów tworzenia drzew, takich jak liczba poziomów drzewa, minimalna liczność węzła oraz potomka, dla całej procedury tworzenia losowego lasu sprawia, że model nie opisuje dobrze rzadko występujących wartości ekstremalnych lub niewystarczająco dopasowuje się do wartości niskich. Zaproponowałam zatem by podzielić cały zbiór danych na dwa podzbiory według wartości zmiennej objaśnianej (tu stężenia NO_2 w powietrzu) modelując każdy z podzbiorów niezależnie. Model końcowy przedstawia już zależności dla całego zakresu wartości zmiennej zależnej.

Dokonałam podziału całości zbioru danych empirycznych na „niskie” (L) oraz „wysokie” (U) wartości i modelowałam każdy z podzbiorów niezależnie. Model (RF) wyznaczony dla wartości niskich (L) oznaczono \mathcal{M}_L , a dla wysokich (U) - \mathcal{M}_U :

$$L = \{(x_1, x_2, \dots, x_n, y); y < \tilde{y}\} \quad (1)$$

$$U = \{(x_1, x_2, \dots, x_n, y); y \geq \tilde{y}\}$$

Zatem model dzielony \mathcal{M} wyraża się następująco:

$$\mathcal{M} = F(X_1, \dots, X_n) = \begin{cases} \mathcal{M}_L & \text{dla } L \\ \mathcal{M}_U & \text{dla } U \end{cases} \quad (2)$$

gdzie X_1, \dots, X_n – zmienne objaśniające (predyktory), y – wartość empiryczna zmiennej zależnej Y , \tilde{y} – wartość graniczna podziału zmiennej Y .

Wartość graniczną \tilde{y} wyznaczono w procesie minimalizacji funkcji błędu całego modelu \mathcal{M} składającego się z dwóch podmodeli \mathcal{M}_L i \mathcal{M}_U .

Modele \mathcal{M}_L i \mathcal{M}_U zostały wyznaczone z zastosowaniem metody random forest (RF). Przy czym dla każdego z modeli określiłam inne parametry tworzenia drzew decyzyjnych. Dla modelu \mathcal{M}_L opisującego zależność wartości niskich stężeń NO_2 od zmiennych objaśniających określiłam liczniejsze ograniczenia tworzenia kolejnych podziałów: minimalna liczność węzła – 40, minimalna liczność potomka – 5 przypadków. Dla modelu \mathcal{M}_U mającego modelować wartości wysokie stężeń, w tym również ekstremalne, zadeklarowałam możliwość tworzenia większych drzew (o trzy poziomy więcej) oraz mniej licznych węzłów, by dokładniej opisać mało liczne wartości ekstremalne. Ustaliłam minimalną licznosc węzła na 5, a minimalną licznosc potomka na 2 przypadki. W obu modelach próba ucząca stanowi 50% wszystkich przypadków, próba testowa 30%. Do budowy każdego drzewa losowane są 4 zmienne objaśniające ze zbioru 8 zmiennych: natężenie ruchu, dzień tygodnia, miesiąc, temperatura powietrza, prędkość wiatru, kierunek wiatru, wilgotność względna oraz ciśnienie powietrza. W porównaniu z pracą [A2] nie uwzględniłam zmiennej „dzień świąteczny” ze względu na silne skorelowanie z dniem tygodnia.

Otrzymany model dzielony \mathcal{M} porównałam z modelem klasycznym, w którym wyznaczyłam jeden las losowy dla całego zbioru danych.

Rozważyłam kilka funkcji błędu: MAPE, MADE oraz BIC dla zobrazowania jakości dopasowania modelu. Każdorazowo wyznaczyłam również R^2 .

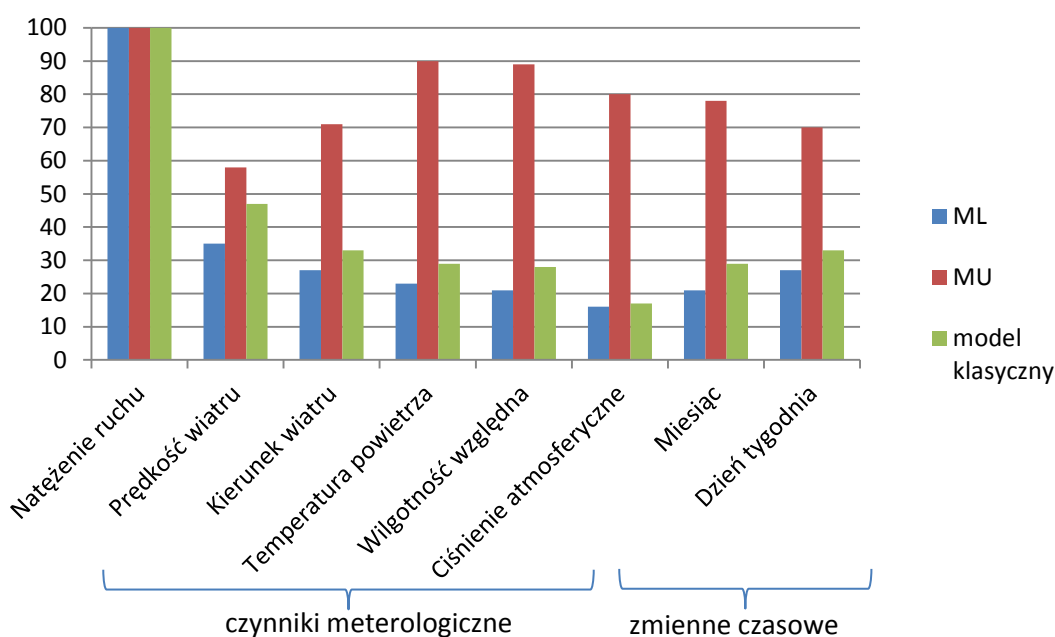
Za rozwiązanie optymalne uznałam taki podział zbioru danych na dwa podzbiory w punkcie \tilde{y} , dla którego średni błąd bezwzględny (MADE), wyznaczony dla modelu dzielonego \mathcal{M} , przyjmuje najmniejszą wartość. Do wyznaczenia optymalnego podziału zbioru danych na niskie i wysokie przeprowadziłam minimalizację funkcji błędu MADE. Zastosowałam numeryczną metodę wyznaczania ekstremum. W wyniku iteracji wyznaczyłam wartość stężenia NO_2 stanowiącą punkt podziałowy na poziomie $57,5 \mu\text{g}/\text{m}^3$ (dokładność oszacowania $\varepsilon = 0,25$). Podzbiory były równoliczne dla podziału przy wartości $\tilde{y} = 49,3 \mu\text{g}/\text{m}^3$. Wartość $57,5 \mu\text{g}/\text{m}^3$ dzieli zbiór danych na podzbiory 16878 oraz 9172 elementowe. Dla modelu dzielonego \mathcal{M} wartość MADE została zredukowana od 10.36 dla tradycyjnego RF bez podziału do $7,35 \mu\text{g}/\text{m}^3$, co oznacza redukcję błędu o

29% kosztem dodania jednego parametru do modelu – punktu podziału zbioru. Pozostałe miary jakości dopasowania również wskazują na znacznie lepsze dopasowanie modelu dzielonego \mathcal{M} w porównaniu z klasyczną analizą pełnego zbioru danych. Średni błąd względny (MAPE) został zredukowany od 22,38% dla klasycznego modelu do 16,65% dla modelu \mathcal{M} (redukcja o 26%). Podobnie dla kryterium informacyjnego Bayesa (BIC): redukcja od 139 656 do 119 408 (o 14%). Współczynnik R^2 zwiększył się od 0.605 dla modelu klasycznego do 0.818 dla modelu dzielonego (o 35%). Wszystkie przedstawione miary wskazują, że procedura podziału zbioru danych dotyczących stężeń NO_2 w powietrzu na dwie części: niskie oraz wysokie, jest uzasadniona i model \mathcal{M} powstały w jej wyniku jest istotnie lepszy od klasycznego modelu RF budowanego dla całej zbiorowości danych.

Wykazałam, że stworzenie modelu dzielonego w zagadnieniu modelowania stężenia NO_2 w powietrzu z zastosowaniem metody RF znacząco zwiększa dokładność, a zatem możliwości stosowania modelu do celów rozpoznania struktury zjawiska. Na podstawie wygenerowanych lasów losowych, dobrze dopasowanych do rzeczywistości ($R^2=0.82$), można wnioskować na temat wpływu poszczególnych zmiennych na wartości stężeń NO_2 w powietrzu w kanionie komunikacyjnym.

Otrzymane w wyniku modelowania ważności zmiennych dla modeli \mathcal{M}_L - niskich stężeń NO_2 , oraz \mathcal{M}_H – wysokich stężeń (Rys.3.) znajdują uzasadnienie w zjawiskach z zakresu fizyki oraz chemii atmosfery. Zanieczyszczenia generowane przez ruch samochodowy modyfikują stężenie NO_2 jedynie przez zwiększanie ładunku zanieczyszczeń. Meteorologiczne zmienne objaśniające mogą modyfikować stężenie NO_2 zarówno w kierunku jego wzrostu jak i redukcji. Wiatr może być czynnikiem redukującym stężenie NO_2 jeśli działa przewietrzająco i powoduje przemieszczanie zanieczyszczeń na zewnątrz miasta lub może powodować kumulowanie zanieczyszczeń jeśli będzie zerowy. Reakcje chemiczne zachodzące w przyziemnych warstwach atmosfery również modyfikują stężenia zanieczyszczeń. Rodzaj i tempo zachodzących reakcji zależą od warunków meteorologicznych. Jednym z najbardziej niebezpiecznych wtórnych zanieczyszczeń powietrza jest ozon troposferyczny, który powstaje z tlenków azotu przy udziale między innymi fotonów oraz lotnych związków organicznych. Podczas procesów ich degradacji oprócz wydzielania się ozonu następuje również przekształcanie się NO w NO_2 . Zatem emisja lotnych związków organicznych również wpływa na wzrost stężenia NO_2 w powietrzu. Proces ten jest tym bardziej intensywny, im więcej substratów znajduje się w powietrzu oraz im bardziej sprzyjające są warunki atmosferyczne: wysokie promieniowanie słoneczne, wysoka temperatura oraz słaby wiatr. W stanie niskich poziomów stężeń NO_2 , którym odpowiada model \mathcal{M}_L , kluczową rolę odgrywa dostarczany do atmosfery ładunek zanieczyszczeń, które w zakresie NO_2 pochodzą głównie z gazów spalinowych pojazdów silnikowych. Natomiast dla wysokich stężeń NO_2 , którym odpowiada model \mathcal{M}_H , reakcje chemiczne przekształcające zanieczyszczenia pierwotne we wtórne zachodzą bardziej intensywnie. Na znaczeniu przybiera wtedy temperatura powietrza oraz wilgotność powietrza (ujemnie skorelowana z promieniowaniem słonecznym). Duże stężenia NO_2 obserwuje się w okresach ze słabym wiatrem, zatem jego wpływ jest również największy ze wszystkich czynników meteorologicznych. Potwierdza

to tezę o słuszności niezależnego rozważania sytuacji kiedy zanieczyszczenie jest „małe” oraz kiedy jest „duże” ponieważ inny jest wpływ poszczególnych czynników na wysokość stężenia zanieczyszczeń NO_2 w każdej z opisanych grup przypadków.



Rysunek 3. Ważność zmiennych niezależnych w modelowaniu stężeń NO_2

Ad 6)

Zasadniczy problem w deterministycznym modelowaniu stężeń zanieczyszczeń, bez zakładania znajomości wartości wstecznych tych stężeń, leży w dużej zmienności wartości stężeń nawet przy podobnych wartościach mierzalnych czynników otoczenia. Alternatywą dla takich modeli są modele probabilistyczne, w których prognozuje się rozkład prawdopodobieństwa dla wartości stężenia w zależności od wartości zmiennych niezależnych. W tematyce zanieczyszczenia powietrza modele te określają prawdopodobieństwo (lub rozkład prawdopodobieństwa) wystąpienia stężenia zanieczyszczenia przy zadanych warunkach otoczenia. Modele probabilistyczne szczególnie sprawdzają się w przypadku przewidywania wartości zmiennych charakteryzujących się ciężkimi ogonami, kiedy dokładność modeli deterministycznych drastycznie spada.

W pracy [A3] zaproponowałam klastrowy model probabilistyczny umożliwiający przewidywanie stężeń zanieczyszczeń powietrza, który w łatwy sposób może być zaimplementowany w dowolnej lokalizacji w kanonie komunikacyjnym. Podstawowym założeniem skuteczności modelu jest bezpośrednia bliskość miejsca dokonywania pomiarów natężenia ruchu oraz jakości powietrza. Wybrana przeze mnie lokalizacja – skrzyżowanie Hallera- Powstańców Śląskich we Wrocławiu, jak wykazano w Ad 1), spełnia ten warunek. Drugim czynnikiem determinującym skuteczność modelu jest właściwy dobór zmiennych objaśniających (predyktorów). Na podstawie wartości współczynników korelacji Pearsona oraz rezultatów moich badań w zakresie analizy

wpływu (ważności) różnych zmiennych na stężenia NO₂ przeprowadzonych dla tego właśnie skrzyżowania zarówno w latach 2015-2016 [A2], jak i latach 2015-2017 [A1] uznałam, że największy wpływ na stężenia dwutlenku azotu w powietrzu mają dwie zmienne: natężenie ruchu i prędkość wiatru. Te zmienne przyjął jako predyktory w dalszej analizie.

Przedstawiona w pracy [A3] idea polega na podzieleniu zbioru danych na klastry według przedziałów wartości zmiennych niezależnych. Przy założeniu uwzględnienia dwóch zmiennych niezależnych klastry będą znajdowały się w układzie macierzowym. Wartości każdej ze zmiennych X_1 , X_2 poddaje się podziałowi na n_1 , n_2 przedziałów odpowiednio. W efekcie podziału powstaje $n_1 \cdot n_2$ podzbiorów (klastrów) danych reprezentowanych przez macierze $D_{i,j} = [Y_{ij}, X_{1ij}, X_{2ij}]$, gdzie $i = 1, 2, \dots, n_1$, $j = 1, 2, \dots, n_2$. Dla każdego klastra niezależnie wyznacza się rozkład prawdopodobieństwa zmiennej $Y_{i,j}$. Znajomość rozkładu prawdopodobieństwa stężeń zanieczyszczeń w ograniczonej przestrzeni warunków otoczenia pozwala określić prawdopodobieństwo przewyższenia zadanej wartości stężenia y_0 przy założeniu zaistnienia takich warunków.

$$P_{a_{1i}a_{2j}}(y_0) = P(Y > y_0 | a_{1i} < X_1 \leq a_{1i+1} \wedge a_{2j} < X_2 \leq a_{2j+1})$$

Gdzie a_{1i} , a_{2j} to wartości graniczne podziału odpowiednio zmiennej X_1 , X_2 .

Znając teoretyczny rozkład prawdopodobieństwa zgodny z rozkładem empirycznym zmiennej $Y_{i,j}$, a zatem i jego dystrybuantę, metoda pozwala też wyznaczać wartość stężenia jaka zostanie przekroczona, przy określonych warunkach, z zadaniem prawdopodobieństwem.

Zagrożeniem dla skuteczności metody jest zbyt mała liczebność podzbiorów danych. Występują wtedy trudności z dopasowaniem rozkładu teoretycznego i uzyskaniem statystycznie istotnej zgodności. Podział wartości zmiennych objaśniających musi być zatem dokonany w taki sposób aby zapewnić minimalną ilość wartości zmiennej $Y_{i,j}$ umożliwiającą przeprowadzenie procedury dopasowania teoretycznego rozkładu prawdopodobieństwa. Zaleca się, aby liczebność klastrów była tak dobrana, by podczas tworzenia klas w procesie wyznaczania wartości statystyki testowej, liczebność w każdej klasie była równa co najmniej 5. W praktyce przyjął minimalną liczebność klastra równą 30 przypadków.

Stosując opisaną metodykę na przykładzie danych z lat 2015-2017 podzieliłam wszystkie wartości stężeń NO₂ na klastry według wartości zmiennych niezależnych, tutaj natężenie ruchu i prędkość wiatru. Wartości graniczne predyktorów przyjęto a priori na podstawie analizy różnicowania predyktorów oraz podstawowych praw fizyki zanieczyszczeń w atmosferze. Natężenie ruchu podzieliłam na przedziały o szerokości 1000 pojazdów, ze względu na zapewnienie wystarczającej liczebności klastrów. Prędkość wiatru podzieliłam co 2 m/s ponieważ:

- pierwszy przedział [0,2] m/s opisuje warunki bezwietrzne i bardzo słabego wiatru, kiedy to jego rola w ewakuowaniu zanieczyszczeń z kanionu komunikacyjnego jest znikoma, natomiast zaleganie zanieczyszczeń sprzyja zachodzeniu reakcji chemicznych w powietrzu,

- kolejne przedziały (2,4], (4,6] oraz (6,8] opisują warunki wiatrowe coraz silniej wpływające na tempo horyzontalnego przemieszczania zanieczyszczeń,
- ostatni przedział (8,+∞) opisuje wiatry silne (jak na lokalne warunki), które w znaczący sposób, poprzez przewietrzanie miasta, modyfikują wartości stężeń NO₂.

Dla wartości stężeń NO₂ odpowiadających przypadkom znajdującym się w każdym klastrze dopasowałam parametry niesymetrycznych ciągłych rozkładów teoretycznych: Webulli, Jonsona, uogólniony wartości ekstremalnych (GEV) i log-normalny. Następnie dla każdego rozkładu teoretycznego oraz każdego klastra przeprowadziłam testy statystycznej zgodności rozkładu teoretycznego z empirycznym: χ^2 oraz Kołmogorowa-Smirnova (K-S). Najlepsze dopasowanie (biorąc pod uwagę wszystkie klastry) otrzymałam dla rozkładu log-normalnego, którego funkcja gęstości opisana jest równaniem (3):

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (3)$$

Wykorzystując znajomość rozkładów teoretycznych stężeń NO₂ przy stabilizowanych warunkach otoczenia, wyznaczyłam prawdopodobieństwa przekroczenia zadanej wartości stężenia w każdym z klastrów reprezentujących określone warunki. Jedną z tablic prawdopodobieństwa przekroczenia przedstawiam poniżej (tab.1.)

Tabela 1. Prawdopodobieństwa przekroczenia stężenia NO₂ równego 40 $\mu\text{g}/\text{m}^3$

$P(Y > 40 \frac{\mu\text{g}}{\text{m}^3})$	[0,2]	(2,4]	(4,6]	(6,8]	(8, +∞)
(0,1000]	39.1%	23.7%	14.2%	7.4%	3.2%
(1000,2000]	64.5%	48.9%	34.0%	22.3%	
(2000,3000]	74.6%	59.0%	46.7%	36.2%	17.9%
(3000,4000]	79.4%	64.2%	54.7%	41.9%	27.8%
(4000,5000]	82.5%	77.1%	71.4%	63.8%	52.8%
(5000, +∞)	85.7%	81.0%	74.4%	66.5%	

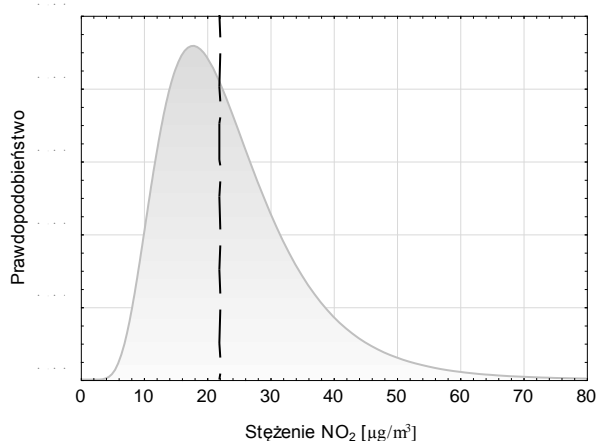
Wartości tabeli należy interpretować następująco: przy natężeniu ruchu nie przekraczającym 1000 pojazdów oraz prędkości wiatru nie większej niż 2 m/s stężenie NO₂ w powietrzu przekroczy 40 $\mu\text{g}/\text{m}^3$ z prawdopodobieństwem 39%. Prawdopodobieństwo wystąpienia zadanego stężenia NO₂ maleje wraz ze wzrostem jego wartości. Najbardziej prawdopodobne są przekroczenia najniższych wartości. Prawdopodobieństwo przekroczenia stężenia 40 $\mu\text{g}/\text{m}^3$ przy najmniej sprzyjających warunkach (klastr (6,5): natężenie ruchu >5000 poj oraz prędkość wiatru ≤ 2 m/s) przekracza 85%, co oznacza, że statystycznie stężenie to jest przekroczone przez 7507 z 8760 godzin w roku. Ponieważ stężenie NO₂ w powietrzu wykazuje zdecydowaną zmienność dobową, to wartość uznawana jako bezpieczna jest przekraczana codziennie. Wraz ze spadkiem natężenia ruchu oraz wzrostem prędkości wiatru prawdopodobieństwo przekroczenia średniej rocznej wartości dopuszczalnej maleje do 3.2% dla klastra (1,5) co oznacza jej przekroczenie przez 282h w ciągu roku. Prawdopodobieństwa przekroczenia

stężenia $100 \mu\text{g}/\text{m}^3$ są kilkakrotnie mniejsze. Wahają się od 0% (przekroczenie przez 10h w roku) dla warunków z silnym wiatrem oraz małym natężeniem ruchu (klaster (1,5) do 22,6% (przekroczenie przez 1997 godzin w roku) dla natężenia ruchu z przedziału (2000,3000] oraz wiatru nie przekraczającego 2 m/s (klaster (3,1)).

Wyznaczone rozkłady prawdopodobieństwa wskazują, że w 7 z 28 opisanych warunkach otoczenia wartość dopuszczalna stężenia NO_2 ($200 \mu\text{g}/\text{m}^3$) występuje rzadziej niż raz w roku. Przy niekorzystnych warunkach otoczenia, klaster (3,1), poziom dopuszczalny może być przekraczany z prawdopodobieństwem sięgającym 3,4% (300 h w roku). Prawdopodobieństwo przekroczenia poziomu alarmowego $400 \mu\text{g}/\text{m}^3$ jest zawsze mniejsze niż 0,2%. Największe prawdopodobieństwa przewyższenia stężenia NO_2 zarówno 100, 200 jak i $400 \mu\text{g}/\text{m}^3$ zidentyfikowano w klastrze (3,1) przy słabym wietrze ($\leq 2\text{m}/\text{s}$) oraz dla natężenia ruchu z przedziału (2000, 3000]poj. Jest to klaster o liczności 985 przypadków obejmujących głównie godziny wieczorne: od 20:01 do 22:00 (70% przypadków) kiedy następuje akumulacja zanieczyszczeń emitowanych przez całodzienny ruch, a samo natężenie ruchu nie jest już ekstremalnie wysokie. Mała prędkość wiatru nasila zjawisko zalegania zanieczyszczeń oraz sprzyja zachodzeniu reakcji chemicznych.

Zaobserwowałam szybszy spadek wartości prawdopodobieństw przekroczenia wraz ze wzrostem prędkości wiatru niż ze spadkiem natężenia ruchu. Bardzo silny wpływ prędkości wiatru na stężenia NO_2 w powietrzu związany jest głównie z geograficznym położeniem skrzyżowania. Przy przeważających we Wrocławiu wiatrach WNW ułożenie arterii rozważanego skrzyżowania zgodne z kierunkiem wiatru sprzyja ewakuacji zanieczyszczeń. Kluczową rolę w zapewnieniu optymalnych warunków do ewakuacji zanieczyszczeń jest zrównoważona zabudowa i rozbudowa miasta. W warunkach Wrocławia, racjonalne kształtowanie zagospodarowania przestrzennego sprzyjającego utrzymaniu korzystnych warunków środowiskowych nie jest dotychczas realizowane [Szewrański i in. 2017]. Niekorzystne oddziaływanie nowej zabudowy zauważa się przy tym nie tylko na obszarze miasta, ale i na nowo zabudowywanych obszarach podmiejskich [Tokarczyk-Dorociak i in. 2018]. Większość nowej zabudowy we Wrocławiu nie wpisuje się w zwarte zespoły urbanistyczne, mogące zagwarantować kliny umożliwiające przewietrzanie miasta. Nadziej na zmianę obecnych wzorców jest jednak przyjęta na szczeblu krajowym strategia adaptacji do zmian klimatu (ang. *Polish National Strategy for Adaptation to Climate Change with the perspective by 2030*), która nakłada na miasta powyżej 100 tysięcy mieszkańców obowiązek opracowania miejskich planów adaptacji do zmian klimatu. Pomimo iż wiodącą rolę w tym zakresie przypisano do zagospodarowania wód opadowych, to jednak dokument ten stwarza możliwość zintegrowanego i systemowego zarządzania różnymi komponentami środowiska, w tym jakością powietrza.

Przedstawioną metodę probabilistycznego przewidywania stężenia NO_2 w powietrzu zweryfikowałam na niezależnym materiale pochodzącym z pierwszego półrocza 2018 roku. Przykład prognozy w postaci wykresu funkcji gęstości wyznaczonego rozkładu log-normalnego oraz zaznaczonej wartości odnotowanej w rzeczywistości wartość w dniu 18 marca 2018 o godz.13 przedstawiłam na rys.4. Odnotowano wtedy wiatr o prędkości 10 m/s oraz natężenie ruchu 2891 pojazdów. Był to zatem klaster (3,5) w macierzy z tabeli 1. Rzeczywiste odnotowane stężenie NO_2 wyniosło $22 \mu\text{g}/\text{m}^3$.



Rysunek 4. Przykład prognozy probabilistycznej dla godziny 14:00 dnia 18 marca 2018 r.

Do ilościowej oceny skuteczności prognozy i jej niepewności zastosowałam współczynnik CRPS (*Continuous Rank Probability Score*), który ma na celu ocenę niezawodności, poprawności i niepewności prognozy probabilistycznej [Hersbach, 2000; Baran i Lerch, 2015]. Wartość CRPS dla pojedynczej prognozy (godziny) wahała się od 1,38 do 53,2 $\mu\text{g}/\text{m}^3$ i wyniosła średnio 9,15 $\mu\text{g}/\text{m}^3$. Biorąc pod uwagę prostotę zastosowanej metody prognozowania oraz jej częstotliwość (godzinowa) rezultat uznałam za zadowalający. Weryfikacja modelu na niezależnym materiale potwierdziła zatem skuteczność modelu klastrowego do probabilistycznego prognozowania stężeń zanieczyszczeń w kanionie komunikacyjnym.

Podsumowanie i wnioski

Polityka Polski oraz Unii Europejskiej w zakresie ustanawiania i egzekwowania prawa dotyczącego utrzymania określonych poziomów stężeń zanieczyszczeń w powietrzu skłania do prowadzenia nasilonych badań w zakresie identyfikacji czynników oraz ich wpływu na stężenia zanieczyszczeń. Umiejętność określenia czynników najsilniej wpływających na zmianę stężeń zanieczyszczeń w powietrzu oraz szacowania wpływu zmiany wartości wybranego czynnika otoczenia na stężenie zanieczyszczenia daje możliwości scenariuszowania i przewidywania skuteczności różnych metod zarządzania. Prace zaprezentowane w części 2.2. autoreferatu wpisują się w tę potrzebę.

Do najważniejszych osiągnięć cyklu prezentowanych prac zaliczam:

1. Rozpoznanie czynników wpływających na stężenia tlenków azotu oraz pyłów zawieszonych w powietrzu w kanionie komunikacyjnym z zastosowaniem technik uczenia maszynowego.
2. Określenie wpływu każdego z rozważanych czynników otoczenia na stężenia zanieczyszczeń dla całego roku oraz z wyszczególnieniem okresów ciepłych oraz chłodnych.

3. Opracowanie autorskiego, z zastosowaniem podziału, modelu opisu zależności między stężeniami zanieczyszczeń a czynnikami otoczenia ze znacznie większym niż w dotychczasowej literaturze poziomem dopasowania do danych rzeczywistych.
4. Wskazanie różnic we wpływie (mierzonym ważnością zmiennej) czynników otoczenia na niskie oraz wysokie stężenia NO₂.
5. Wykazanie merytorycznej zgodności oszacowania ważności zmiennych w modelu dzielonym dla niskich i wysokich stężeń z prawami fizyki i chemii niskich warstw troposfery.
6. Opracowanie prostego klastrowego modelu probabilistycznego do prognozowania stężeń zanieczyszczeń powietrza na podstawie czynników otoczenia.
7. Wskazanie wartości natężenia ruchu i prędkości wiatru o największym prawdopodobieństwie przekroczenia zadanych wartości stężeń NO₂.

Poprzez opisane powyżej osiągnięcia wniosłam istotny wkład w rozwój dyscypliny naukowej Kształtowanie i Ochrona Środowiska.

Bibliografia

1. Aldrin M., Haff I.H., 2004. Generalized additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment* 2005, 39,11 p.2145-2155. <https://doi.org/10.1016/j.atmosenv.2004.12.020>
2. Archer K.J., Kimes R.V., 2008. Empirical characterization of random forest variable importance measures. *Computer Statistics Data Analysis* 52(4), 2249-2260. <https://doi.org/10.1016/j.csda.2007.08.015>
3. Aznarte J.L. Probabilistic Forecasting for extreme NO₂ pollution episodes. *Environmental Pollution* 2017, 229 321-328 <https://doi.org/10.1016/j.envpol.2017.05.079>
4. Balashov N.V., Thompson A.M., Young G.S., 2017. Probabilistic forecasting of surface ozone with a novel statistical approach. *Journal of Applied Meteorology and Climatology* 2017, 56 297-316 <https://doi.org/10.1175/JAMC-D-16-0110.1>
5. Baran S., Lerch S., 2015. Log-normal distribution based Ensemble Model Output Statistics model for probabilistic wind-speed forecasting. *Quarterly Journal of the Royal Meteorological Society*, 141, 2289-2299 doi:10.1002/qj.2521
6. Battista G., de Lieto Vollaro R., 2017. Correlation between air pollution and weather data in urban areas: Assessment of the city of Rome (Italy) as spatially and temporally independent regarding pollutants. *Atmospheric Environment* 2017, 165, 240-247 <https://doi.org/10.1016/j.atmosenv.2017.06.050>
7. Boylan J. W., Russell A. G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmospheric Environment* 40, p.4946-4959. <https://doi.org/10.1016/j.atmosenv.2005.09.087>
8. Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984. *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
9. Catalano M., Galatioto F., Bell M., Namdeo A., Bergantino A.S., 2016. Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environment Science&Policy* 60, 69-83 <https://doi.org/10.1016/j.envsci.2016.03.008>
10. EEA Report No 9/2017, European Union emission inventory report 1990–2015 under the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP), doi:10.2800/478321
11. EEA, 2017, Exceedances of air quality objectives due to traffic. [online] (Accessed 06.04.2018) Available from: <https://www.eea.europa.eu/data-and-maps/indicators/exceedances-of-air-quality-objectives-7/assessment>

12. Elangasinghe M.A., Singhal N., Dirks K.N., Salmond J.A., 2014. Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research* 2014, 5, 4, 696-708 <https://doi.org/10.5094/APR.2014.079>
13. Flores-Pajot M.-C., Ofner M., Do M.T., Lavigne E., Villeneuve P.J., 2016. Childhood autism spectrum disorders and exposure to nitrogen dioxide, and particulate matter air pollution: A review and meta-analysis, *Environmental Research* 151(2016) p.763-776. <https://doi.org/10.1016/j.envres.2016.07.030>
14. Friedman J. H., 1999a. *Greedy function approximation: A gradient boost-ing machine*. IMS 1999 Reitz Lecture.
15. Friedman, J. H., 1999b. *Stochastic gradient boosting*. Stanford Univer-sity1999.
16. Hersbach H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting* 2000, 15, 559-570 [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
17. Hien T.T., Linh H.N., Luong L.M.T., Thai P.K., 2016. Air pollution and risk of respiratory and cardiovascular hospitalizations in the most populous city in Vietnam, *Science of the Total Environment*, 557–558, 2016, p. 322-330. <https://doi.org/10.1016/j.scitotenv.2016.03.070>
18. Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, Kaufman JD., 2013. Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environ Health*. 28;12(1):43. <https://doi.org/10.1186/1476-069X-12-43>
19. Informacja o jakości powietrza na terenie miasta Wrocławia, Wojewódzki Inspektorat Ochrony Środowiska, Wrocław 2016.
20. Laña I., Del Ser J., Pedró A., Vélez M., Casanova-Mateo C., 2016. The role of local urban traffic and meteorological conditions in air pollution: A data-based study in Madrid, Spain. *Atmospheric Environment* 145, p.424-438. <https://doi.org/10.1016/j.atmosenv.2016.09.052>
21. Larkin A., Geddes J.A., Martin R.V., Xiao Q., Liu Y., Marshall J.D., Brauer M., Hystad. P., 2017. Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environmental Science and Technology*, 51, p.6957-6964. doi :10.1021/acs.est.7b01148
22. Nejadkoorki F., Baroutian S., 2012. Forecasting extreme PM10 concentrations using artificial neural networks. *International Journal of Environmental Research* 2012, 6, 277-284 doi: 10.22059/IJER.2011.493
23. Padró-Martínez L.T., Patton A.P., Trull J.B., Zamore W., Brugge D., Durant J.L., 2012. Mobile monitoring of particle number concentration and other traffic-related air pollutants in a near-highway neighborhood over the course of a year. *Atmospheric Environment* 61, 253- 64 <https://doi.org/10.1016/j.atmosenv.2012.06.088>
24. Papanastasiou D.K., Melas D., Kioutsioukis I., 2007. Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. *Water Air and Soil Pollution* 182, 325-334 doi:10.1007/s11270-007-9341-0
25. Lee P.-C., Liu L.-L., SunMD Y., Chen Y.-A., Liu Ch.-Ch., Li Ch.-Y., Yu H.-L., Ritz B., 2016. Traffic-related air pollution increased the risk of Parkinson's disease in Taiwan: A nationwide study. *Environment International* 96 (2016) p.75–81. <https://doi.org/10.1016/j.envint.2016.08.017>
26. Ping Shi J., Harrison R.M., 1997. Regression modelling of hourly NO_x and NO₂ concentration in urban air in London. *Atmospheric Environment* 31, 24, 4081-4094 [https://doi.org/10.1016/S1352-2310\(97\)00282-3](https://doi.org/10.1016/S1352-2310(97)00282-3)
27. Rozporządzenie Ministra Środowiska z dnia 24 sierpnia 2012 r. w sprawie poziomów niektórych substancji w powietrzu (Dz.U.2012.1031)
28. Sayegh A., Tate J.A., Ropkins K., 2016. Understanding how roadside concentrations of NO_x are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees. *Atmospheric Environment* 127, 163-175. <https://doi.org/10.1016/j.atmosenv.2015.12.024>

29. Singh K.P., Gupta S., Rai P., 2013. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*, 80, 426-437 doi: 10.1016/j.atmosenv.2013.08.023
30. Szewrański S., Kazak J., Żmuda R., Wawer R., 2017. Indicator-based assessment for soil resource management in the Wrocław larger urban zone of Poland. *Polish Journal of Environmental Studies*, 26, doi:10.15244/pjoes/70178.
31. Tang G., Zhao P., Wang Y., Gao W., Cheng M., Xin Y., Li X., Wang Y., 2017. Mortality and air pollution in Beijing: The long-term relationship. *Atmospheric Environment*, 150, p.1238-243. <https://doi.org/10.1016/j.atmosenv.2016.11.045>
32. Tokarczyk-Dorociak K., Kazak J., Szewrański S., 2018. The impact of a large city on land use in suburban area - The case of Wrocław (Poland). *Journal of Ecological Engineering* 19, doi:10.12911/22998993/81783.
33. World Health Organization Office for Europe, 2012. Burden of disease from Household Air Pollution for 2012
http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf
34. Zhang Z., Zhang X., Gong D., Quan W., Zhao X., Ma Z., Kim S-J., 2015. Evolution of Surface O₃ and PM_{2.5} concentrations and their relationships with meteorological conditions over the last decade in Beijing, *Atmospheric Environment* 108, 67-75 <https://doi.org/10.1016/j.atmosenv.2015.02.071>

3. Pozostałe osiągnięcia naukowo – badawcze

Pracę naukowo–badawczą rozpoczęłam wraz z rozpoczęciem studiów doktoranckich na Uniwersytecie Przyrodniczym we Wrocławiu w 2003 roku. Do czasu uzyskania stopnia naukowego doktora, moje zainteresowania skupiały się wokół zagadnień związanych z bilansem cieplnym i klimatycznym bilansem wodnym różnych roślin uprawnych [poz.40-43, tab.2.] oraz lasu iglastego [poz.39, 44, tab.2.]. Na podstawie mojej rozprawy doktorskiej została wydana monografia [poz.1., tab.4.]. Wraz z współautorami analizowałam zmiany klimatu w Bydgoszczy w różnych wieloletniach [poz.37, tab.2.; poz.2., tab.5.]. W związku z wprowadzaniem zyskującej na popularności na początku XX wieku w Polsce metody automatycznych pomiarów czynników meteorologicznych, rozważaliśmy wpływ metody pomiaru na wartości ewapotranspiracji [poz.35, 38, tab.2.] oraz elementów bilansu cieplnego [poz.34, tab.2.]. Poruszaliśmy również problem samego modelowania ewapotranspiracji [poz.36, tab.2.] i, w innym zespole współautorów, opracowaliśmy aplikację Evapo do jej wyznaczania [poz.20, tab.2.].

Po obronie doktoratu moje zainteresowania naukowe skierowały się w stronę analizy środowiskowych aspektów transportu samochodowego. Zagadnienie to jest bardzo szerokie i rozważałam je na różnych płaszczyznach, regularnie publikując wyniki badań. W zakresie badań związanych z relacjami pomiędzy transportem samochodowym a środowiskiem, uzyskane przeze mnie wraz z współautorami wyniki można podzielić na następujące części tematyczne:

1. Bezpieczeństwo ruchu drogowego mierzone liczbą wypadków drogowych – 6 publikacji,
2. Edukacja kierowców oraz wpływ ich umiejętności na występowanie wypadków drogowych – 10 publikacji,
3. Modelowanie ruchu samochodowego w mieście – 13 publikacji,

Pozostałe prace prezentują mój dorobek we współautorstwie z różnymi grupami naukowców. Mój wkład w poniższe prace polegał głównie na wykonaniu analizy statystycznej oraz interpretacji jej wyników. Publikacje z tej grupy można podzielić tematycznie następująco:

- a) badanie zależności między parametrami mikrostrugi gazów chłodzących w procesie spawania a jakością spoiny [poz.1., tab.1.].
- b) zagadnienia związane ze stężeniami zanieczyszczeń w wodach rzeki Trzemny [poz.2., tab.1.] oraz zbiornika Dobromierz [poz.3., tab.1.],
- c) badania rozkładów kropel deszczu naturalnego [poz.28., 30., 31., tab.2.; poz.1., tab.5.],
- d) ocena wpływu niedokładności określenia współczynnika filtracji elementów uszczelniających wały przeciwpowodziowe, na warunki przepływu przez nie wody [poz.22., tab.2.],
- e) analiza wykorzystania roweru miejskiego we Wrocławiu [poz.6., tab.2.]

- f) logistyka elementów ponadgabarytowych [poz.1., tab.2.],
- g) probabilistyczne prognozowanie zużycia pary ślizgowej pierścieni tłokowy-tuleja cylindrowa [poz.18., tab.2.] – zagadnienie związane z moją pracą inżynierską na Wydziale Mechanicznym Politechniki Wrocławskiej.

Lista wszystkich publikacji została przedstawiona w załączniku 3.

Ad. 1. Bezpieczeństwo ruchu drogowego mierzone liczbą wypadków drogowych.

Wypadek drogowy to zdarzenie drogowe zaistniałe, lub mające początek, na drodze publicznej w związku z ruchem przynajmniej jednego pojazdu, w wyniku którego są ofiary w ludziach: śmiertelne lub ranne. Zależność między czynnikami warunkującymi ruch drogowy a liczbą wypadków drogowych pierwotnie opisywano deterministycznymi równaniami regresji liniowej. Jednak zmienne w czasie natężenie ruchu oraz częste występowanie zerowej liczby wypadków w rozważanym okresie czasu na badanym odcinku drogi sprawiają, że standardowa regresja liniowa prowadziła do błędnych estymatorów parametrów. Wykorzystano zatem regresję Poissona, ze względu na często występujące wartości zerowe oraz nieujemne pozostałe wartości zmiennej objaśnianej – liczby wypadków drogowych. W modelu Gamma-Poissona (NB) wprowadzono składnik losowy ε , dla którego $\exp(\varepsilon)$ ma rozkład gamma. Jeśli natomiast ε podlega rozkładowi normalnemu, model nazywany jest log-normalnym-Poissona (PLN). W moich badaniach zmienną zależną była liczba wypadków drogowych w zadanym okresie czasu na określonych odcinkach dróg zwanych sekcjami. Zbiór zmiennych objaśniających to: średnie dzienne natężenie ruchu (AADT – *annual average daily traffic*), logarytm naturalny z minimalnej liczby pasów ruchu plus jeden, logarytm naturalny z liczby skrzyżowań w sekcji, ograniczenia prędkości oraz rodzaj drogi. W pracy [poz.21., tab.2.] dla danych pochodzących z norweskiej jednostki administracyjnej Hordaland wyestymowałam współczynniki modeli Poissona oraz NB i wykazałam lepsze dopasowanie modelu NB. Wyzaczyłam podział wariancji na systematyczną, losową oraz nie wyjaśnioną przez model. Model wyjaśnił 70% zmienności systematycznej. W pracy [poz.23., tab.2.] rozszerzyłam zbiór zmiennych objaśniających o logarytm naturalny z AADT oraz porównałam modele, dla danych z poprzedniej pracy (Poissona, NB), z modelem PLN. Najlepsze dopasowanie do danych otrzymałam dla modelu PLN. Dla najlepszego z prezentowanych w pracy modeli przedstawiłam ocenę jakości dopasowania za pomocą metody CURE – kumulowanych residuów dla zmiennej objaśniającej AADT. Kolejnym rozszerzeniem analiz było modelowanie wpływu czynników drogowych na liczbę wypadków drogowych w całej Norwegii. W pracy [poz.11., tab.2.] podjęłam się badania wpływu przyjętego związku funkcyjnego (functional form) między zmiennymi objaśniającymi na rezultaty modelowania metodą PLN. Zaproponowałam 6 funkcji i wyznaczyłam modele z 11 zmiennymi objaśniającymi w różnych związkach. Porównałam wyniki wskazując na podobieństwa i różnice między poszczególnymi modelami w zakresie jakości dopasowania oraz wpływu każdej ze zmiennych objaśnianych. Najbardziej złożona forma funkcyjna zawierająca wszystkie zmienne objaśniane oraz ich

przekształcenia ($\ln(\text{AADT})$, ograniczenie prędkości w kwadracie itp.) okazała się być najlepiej dopasowaną.

Oslo jest największym miastem Norwegii, zamieszkanym przez 8% ludności kraju. W pracy [poz.3., tab.3.] postawiłam pytania: W jaki sposób na liczbę wypadków drogowych wpływa lokalizacja odcinka drogi? Czy można modelować na podstawie połączonych danych ze wszystkich lokalizacji liczbę wypadków drogowych? Czy modelowanie dla dużych miast wymaga osobnej analizy? Wprowadziłam do modelowania metodą PLN dodatkową, binarną zmienną „dummy of Oslo” oznaczającą przynależność odcinka drogi do miasta Oslo, lub nie (wartości zmiennej 1 lub 0 odpowiednio). Dodanie zmiennej określającej przynależność do Oslo poprawiło jakość dopasowania modelu wykonanego dla całego kraju wskazując tą zmienną jako statystycznie istotną.

Prace [poz.1., tab.6.; poz.4., tab.1.] będące głównym rezultatem mojego stażu naukowego w Instytucie Ekonomiki Transportu w Oslo poruszają temat wpływu stosowania opon kolcowanych na bezpieczeństwo drogowe w aspekcie liczby wypadków drogowych. Stosując model NB zbadaliśmy wpływ używania opon kolcowanych (w udziałach procentowych ze wszystkich pojazdów) dla każdego z większych miast Norwegii: Oslo, Drammen, Stavanger, Bergen oraz Trondheim. Badania przeprowadzono dla lat 1990-2009. W zbiorze zmiennych objaśniających uwzględniliśmy: opad śniegu, grubość pokrywy śnieżnej, opad deszczu, termin wystąpienia pokrywy śnieżnej, temperaturę powietrza, AADT, wystąpienie dnia wolnego, świątecznego (z wyszczególnieniem świąt). Wykazaliśmy, że redukcja stosowania opon kolcowanych w norweskich miastach wpłynęła na zwiększenie liczby wypadków drogowych.

Ad.2. Edukacja kierowców oraz wpływ ich umiejętności na występowanie wypadków drogowych

Z analiz Komendy Głównej Policji wynika, że 78% wypadków drogowych miało miejsce z winy kierowcy. W opinii Komendy Głównej Policji główny wpływ na niski poziom bezpieczeństwa na drodze mają niedostateczne kwalifikacje i dyscyplina kierujących pojazdami. 77% wszystkich sprawców wypadków będących kierowcami prowadziło samochód osobowy. Dominujący wpływ czynnika ludzkiego w postaci umiejętności kierujących pojazdami skłonił mnie do rozważenia zagadnienia wpływu świadomości oraz edukacji kierowców na zmianę ich umiejętności i redukcji liczby wypadków przez nich powodowanych. Przeprowadzona wśród studentów studiów stacjonarnych ankieta wykazała, że 28% badanych, w wieku 19-21 lat, oceniło swoje umiejętności kierowania samochodem najwyżej w proponowanej skali [poz.24., 27., tab.2.]. W odpowiedzi na niewystarczające umiejętności kierowców, skutkujące około pięcioma tysiącami ofiar śmiertelnych rocznie w Polsce, na rynku pojawiły się oferty szkoleń doskonalących technikę jazdy, przeznaczone dla kierowców posiadających uprawnienia do kierowania pojazdem. Dla kierowców pojazdów specjalnych szkolenia takie są obowiązkowe, dla kierowców niepracujących w zawodzie kierowcy dobrowolne. Na podstawie informacji pozyskanych z Ośrodka Doskonalenia Techniki Jazdy we Wrocławiu dokonałam analizy struktury wiekowej kierowców świadomych niedoborów

swoich umiejętności w zakresie kierowania samochodem porównując ją ze strukturą wiekową sprawców wypadków [poz.33., tab.2.]. Najwyższy wskaźnik liczby wypadków na 10 000 populacji dotyczy grupy wiekowej 18-29 lat, podczas gdy uczestnicy szkoleń najczęściej reprezentowali grupę wiekową 30-39 oraz 40-49 lat. Dokładną analizę błędów najczęściej popełnianych przez kierowców w aspekcie przyczyn wypadków drogowych wraz z elementami szkolenia mającymi na celu ich unikanie rozważyłam w pracy [poz.32., tab.2.]. Dokładną analizę świadomości kierowców w zakresie najbardziej popularnego systemu bezpieczeństwa czynnego ABS (Anti Block System) jego stosowalności, skuteczności i ograniczeń przedstawiłam w pracy [poz.25., tab.2.]. Badanie przeprowadzono na grupie 623 kierowców czynnie użytkujących samochód osobowy wyposażony w system ABS. Jedynie 51% kierowców wiedziało jak poprawnie i efektywnie wykorzystywać system ABS. Kolejnym krokiem było ilościowe określenie skuteczności szkoleń prowadzonych przez Ośrodek Doskonalenia Techniki Jazdy [poz.19., tab.2.]. Badania przeprowadzono na grupie 212 kierowców pojazdów uprzywilejowanych ocenianych w zakresie 3 elementów kierowania pojazdem w trudnych warunkach: czasu przejazdu (określającego umiejętność doboru prędkości do warunków drogowych), omijaniu przeszkody, odzyskaniu panowania nad pojazdem w poślizgu. 94% z nich wykazało wzrost umiejętności, natomiast 76% wykazało podniesienie umiejętności do poziomu zadowalającego dla kierowców pojazdów uprzywilejowanych. Metodyka ilościowej oceny jakości szkoleń została opisana i zastosowana w pracach [poz.13., 17., tab.2.]], w których przedstawiłam, wraz z współautorem, sposób wskaźnikowej oceny jakości szkolenia oraz oprogramowanie EwaS będące rozszerzeniem aplikacji wykonanej na zlecenie przedsiębiorcy – Ośrodka Doskonalenia Techniki Jazdy w ramach Dolnośląskiego Bonu na Innowacje w 2013 roku.

W związku ze stażem naukowym w Instytucie Ekonomiki Transportu w Oslo, który odbyłam w 2010 roku, dokonałam również porównania stanu bezpieczeństwa ruchu drogowego w Polsce i w Norwegii [poz.29., tab.2.] wraz z procesem edukacji kierowców i kandydatów na kierowców [poz.26., tab.2.]. W Norwegii przypada 540 samochodów osobowych na 1000 mieszkańców podczas gdy w Polsce 380. Mimo wysokiego poziomu zmotoryzowania społeczeństwa liczba ofiar śmiertelnych w wypadkach drogowych w Norwegii oscyluje około 300 podczas gdy w Polsce około 5000. Przyczyną istnienia tak radykalnej różnicy w bezpieczeństwie ruchu drogowego na korzyść Norwegii są stosowane tam: system szkoleniowy, przepisy ruchu drogowego, kultura osobista społeczeństwa oraz wysokie kary za wykroczenia drogowe.

Ad.3. Modelowanie ruchu samochodowego w mieście

Zagadnienie modelowania ruchu pojazdów w sieci drogowej jest popularną metodą stosowaną do identyfikacji zjawisk zachodzących w układach komunikacyjnych oraz ich skutków społecznych i środowiskowych. Prace z tego zakresu tematycznego powstawały we współpracy z doktorem Chalfenem, będącym pracownikiem Katedry Matematyki Uniwersytetu Przyrodniczego we Wrocławiu. Dr Chalfen zajmował się głównie tworzeniem programu komputerowego według omówionej metodologii. Podjęliśmy próbę adaptacji zasad modelowania ruchu wody gruntowej do zjawisk przemieszczania się

pojazdów [poz.9., tab.2.]. Zastosowaliśmy metodykę opartą na równaniu Boussinesq'a i jego numerycznym rozwiązaniu. W stworzonym modelu obszar miasta rozważany był jako powierzchnia ciągła. Zastosowanie modelu ciągłego w przestrzeni pozwoliło na uzyskanie rozkładu przestrzennego wektorów prędkości wraz z wyznaczeniem trasy przejazdu pojazdu, od zadanego punktu startu (S) do punktu końcowego (F). Wyznaczyliśmy trajektorie 5000 pojazdów przypisując każdemu odrębną parę punktów S-F. Wygenerowany metodą dwuwymiarowego ciągłego modelu ruchu z analogią do przepływu wód gruntowych ruch miejski porównano do wyznaczonego w standardowy sposób modelu opartego na teorii grafów. Metoda oparta na algorytmie Dijkstry wyznaczania najkrótszej ścieżki w grafie, opisana wraz ze wszystkimi stosowanymi założeniami w pracy [poz.16., tab.2.], pozwalała wyznaczać krótsze pod względem czasu przejazdu ścieżki według zadanego z góry kryterium. Rozważaliśmy kryterium najkrótszej odległości oraz najkrótszego czasu przejazdu. W dalszych pracach stosowaliśmy opracowany samodzielnie autorski program do symulowania ruchu w mieście z zastosowaniem algorytmu Dijkstry, oparty na siatce trójkątów symulującej siatkę ulic hipotetycznego miasta. Kolejne prace były podsumowaniem badania charakterystyk ruchu samochodowego w zależności od przyjętych założeń wejściowych do modelu. Zbadaliśmy wpływ parametrów obwodnicy śródmiejskiej na gęstość ruchu samochodowego w całym mieście [poz.15., tab.2.]. Zakładając jednorodne pole prędkości poruszania się pojazdów po ulicach w całym mieście, wyznaczyliśmy optymalne promienie obwodnic według dwóch kryteriów: minimalizacji średniego czasu przejazdu z punktu S do F oraz minimalizacji maksymalnej lokalnej gęstości ruchu (lokalnej kongestii/korków). Otrzymaliśmy rozwiązania różne dla każdego z założeń, określające optymalny promień obwodnicy na 2,5 km oraz 4 km odpowiednio. Badania wskazały potrzebę budowy również dwóch kolejnych obwodnic, odpowiedników miejskiej oraz pozamiejskiej dalej oddalonych od centrum miasta. Następnie rozważaliśmy modelowanie ruchu w mieście o niejednorodnej prędkości przejazdu, przyjmując cztery warianty rozkładu prędkości dopuszczalnej w mieście oraz piątej z ciągłą zmianą ograniczenia prędkości na obszarze hipotetycznego miasta [poz.14., tab.2.]. Wyznaczyliśmy dla każdego wariantu mapy gęstości ruchu i porównaliśmy je tworząc mapy różnic między wariantami. Bez względu na przyjęty wariant układu stref ograniczenia prędkości przejazdu, gęstość ruchu zależy przede wszystkim od założonej strategii wyboru optymalnej trasy: najkrótszej w odległości lub w czasie.

Jedną z zasadniczych funkcji, którą trzeba zadać w modelu ruchu, jest określenie relacji między prędkością ruchu a gęstością pojazdów. Dla niewielkiej liczby pojazdów prędkość ich poruszania się jest ograniczona przepisami ruchu drogowego. Jednak wraz ze wzrostem gęstości ruchu, pojazdy nie jadą już z maksymalną dopuszczalną prędkością, tylko zredukowaną ze względów bezpieczeństwa przez kierowcę. Zjawisko to określane jest jako zachowanie bezpiecznej odległości między pojazdami. Tej relacji poświęcone są kolejne dwie prace. Wraz ze współautorem wykazaliśmy, że parametrem funkcji prędkości od gęstości ruchu jest bezpieczna odległość między pojazdami mierzona w jednostkach czasu [poz.7., tab.2.]. Wyznaczyliśmy funkcję hiperboliczną poprawnie opisującą zjawisko w zakresie prędkości większych od 10 km/h. Następnie uogólniliśmy postać funkcji do

znajomości rozkładu prawdopodobieństwa bezpiecznych odległości między pojazdami w całej populacji kierowców [poz.5., tab.2.]. Badania modelowe na schematycznym grafie zawarto również w pracy [poz.4., tab.2.].

Przedstawione powyżej rozważania były prowadzone przy założeniu, że kierowca ustala trasę przejazdu w chwili startu i konsekwentnie ją realizuje. Rozważyliśmy również możliwość ciągłego w czasie informowania kierowcy o aktualnej sytuacji na drodze. Wykazaliśmy, że aktualna informacja o gęstości ruchu w danej chwili w całym mieście i modyfikacja trasy w trakcie jej realizacji, pozwala na redukcję czasu przejazdu nawet do 34% [poz.8., tab.2.].

Rozpoznaliśmy strukturę rzeczywistego ruchu we Wrocławiu na podstawie danych empirycznych natężenia ruchu z 2012 roku pochodzących z pętli indukcyjnych zainstalowanych w ciągu Autostradowej Obwodnicy Wrocławia (AOW). Wyznaczyliśmy dwumodalną funkcję opisującą zmienność dobową natężenia ruchu dla każdego przekroju pomiarowego wskazującą na szczyt poranny oraz popołudniowy natężenia ruchu [poz.12., tab.2.]. Wykonaliśmy również bardziej dogłębną analizę widmową natężenia ruchu na AOW [poz.10., tab.2.], dokonując tygodniowej dekompozycji sezonowej szeregów czasowych dla każdego z 17 przekrojów pomiarowych. Przeciętna wartość błędu dopasowania wyniosła 5%, co uznaliśmy za zadawalające. Rozpoznane funkcje opisujące natężenie ruchu na AOW zostały zastosowane do modelowania ruchu samochodowego na grafie reprezentującym rzeczywistą sieć ulic Wrocławia.

W 2017 roku zbudowaliśmy graf reprezentujący rzeczywistą sieć ulic Wrocławia. Na podstawie danych pochodzących z OpenStreetMap utworzyliśmy i zweryfikowaliśmy graf ulic rzeczywistego miasta. Na jego podstawie przeprowadziliśmy szereg symulacji komputerowych generowania ruchu samochodowego w mieście przy określonych założeniach. Badanie podziału czasu przejazdu na czas jazdy z maksymalną dozwoloną prędkością, prędkością ograniczoną przez gęstość ruchu oraz oczekiwania na skrzyżowaniach przedstawiliśmy w pracy [poz.3., tab.2.] wskazując, że wraz ze wzrostem liczby pojazdów poruszających się po mieście do 57% rośnie czas jazdy z prędkością ograniczoną ze względu na dużą gęstość ruchu. Przeprowadziliśmy również badania symulacyjne dla rzeczywistych warunków i rozważanych przez władze miasta rozwiązań połączenia ulic w rejonie Dworca Świebodzkiego we Wrocławiu [poz.2., tab.2.]. Dokonaliśmy oceny obu proponowanych rozwiązań pod względem rozkładu natężenia ruchu na projektowanych łącznikach oraz okolicznych ulicach. Na podstawie rzeczywistych danych o natężeniu ruchu na 6 największych skrzyżowaniach we Wrocławiu, przeprowadziliśmy identyfikację parametrów modelu numerycznego w taki sposób, aby ruch symulowany jak najbardziej odpowiadał sytuacji rzeczywistej w skali makro. Następnie zweryfikowaliśmy poprawność modelu i parametrów na danych z 2016 roku otrzymując błąd względny liczby pojazdów w ruchu w godzinach 6-22 około 10%, a w godzinach nocnych do 20% [poz.1., tab.3.].

Modelowanie ruchu samochodowego we Wrocławiu zastosowaliśmy również w pracy dotyczącej możliwości zastosowania zagadnień modelowania ruchu miejskiego do geodynamicznego procesu wspomagania decyzji w zarządzaniu ruchem [poz.2., tab.3.].

4. Podsumowanie dorobku habilitantki

Mój dotychczasowy dorobek naukowo-badawczy obejmuje 61 publikacji w tym 54 po doktoracie. 7 prac zostało opublikowanych w czasopismach z listy Journal Citation Reports (JCR) o łącznym wskaźniku impact factor (IF), na rok opublikowania, równym 16,278. Łączna liczba punktów zgodnie z listą MNiSW na rok publikacji wynosi 544pkt. Wśród dorobku publikacyjnego 11 pozycji jest wyłącznie mojego autorstwa w tym: 3 znajdujące się na liście JCR (włączone do dorobku w pkt.2.2.), 2 indeksowane w bazie Web of Science, 1 monografia będąca podsumowaniem rozprawy doktorskiej. Szczegółowy wykaz publikacji przedstawiłam w tabelach 1-6 załącznika 3. Artykuły naukowe mojego autorstwa lub współautorstwa według bazy Web of Science zostały zacytowane 24 razy. Publikacje wymienione w punkcie 2.2. według bazy Scopus zostały dotychczas zacytowane 4 razy, a według Google Scholar 13 razy co wskazuje na wysokie zainteresowanie moimi pracami w zakresie przedstawionym w osiągnięciu naukowym. Mój indeks Hirscha według bazy Web of Science wynosi 3. Byłam członkiem zespołu badawczego 3 projektów badawczych realizowanych na Wydziale inżynierii Kształtowania Środowiska i Geodezji Uniwersytetu Przyrodniczego we Wrocławiu. Byłam kierownikiem 6 grantów finansowanych przez Uniwersytet Przyrodniczy we Wrocławiu. Byłam beneficjentem w programie stypendialnym Funduszu Stypendialnego i Szkoleniowego finansowanego z Mechanizmu Finansowego Europejskiego Obszaru Gospodarczego oraz Norweskiego Mechanizmu Finansowego. Otrzymałam 3 nagrody Rektora za działalność naukową oraz jedną za działalność organizacyjną. Uczestniczyłam w 40 konferencjach w tym 4 międzynarodowych. Wygłosiłam 36 referatów. Odbyłam półroczny staż naukowy w Instytucie Ekonomii Transportu w Oslo.

Wykonałam 5 recenzji artykułów naukowych dla zagranicznych czasopism. Od 2014 roku nieprzerwanie jestem Ekspertem Narodowego Centrum Badań i Rozwoju do spraw oceny wniosków o dofinansowanie przedsięwzięć finansowanych przez NCBR. Brałam udział w ocenie 10 wniosków z różnych programów. Jestem promotorem pomocniczym w jednym przewodzie doktorskim. Byłam promotorem 6 prac inżynierskich studentów kierunku gospodarka przestrzenna.

W latach 2011-2018 byłam członkiem komitetu organizacyjnego corocznego Seminarium Zastosowań Matematyki organizowanego przez Katedrę Matematyki Uniwersytetu Przyrodniczego we Wrocławiu oraz redaktorem wydawanych materiałów konferencyjnych.

Aktywnie realizowałam ideę współpracy nauki i gospodarki poprzez realizację usługi na rzecz przedsiębiorcy w ramach Dolnośląskiego Bonu na Innowacje oraz półroczny staż w przedsiębiorstwie VagTune.

Szczegółowe informacje na temat mojego dorobku naukowo-badawczego przedstawiłam w załączniku 3, wykaz osiągnięć w działalności dydaktycznej, organizacyjnej i popularyzującej naukę w załączniku 4.

Mój identyfikator:

ORCID: 0000-0002-0157-516X

SCOPUS: 50861636000

ResearcherID: S-7765-2016

Tabela 1. Zestawienie dorobku naukowego

	ogółem	przed doktoratem	po doktoracie
Prace opublikowane:	61	7	54
- w czasopiśmie z listy JCR	7	0	7
- w innych czasopiśmie	46	6	40
- w materiałach konferencji międzynarodowych	4	0	4
- w monografiach	2	1	1
Monografia	1	0	1
Dokumentacja prac badawczych - raport	1	0	1

Tabela 2. Zestawienie dorobku naukowego z uwzględnieniem oceny punktowej czasopism wg MNiSW oraz IF za rok publikacji (stan na grudzień 2018 r.)

Czasopismo	Język publikacji	Liczba publikacji	IF*	Punkty wg MNiSW *
Czasopisma naukowe posiadające współczynnik Impact Factor (IF)				
Science of the Total Environment	A	1	4,610**	40
Journal of Environmental Management	A	1	4,005**	35
Accident Analysis and Prevention	A	1	2,571	40
Sustainability	A	1	2,075	20
Desalination and Water Treatment	A	1	1,272	20
Polish Journal of Environmental Studies	A	1	1,120	15
Archives of Metallurgy and Materials	A	1	0,625**	30
Łącznie		7	16,278	200
Czasopisma naukowe nieposiadające współczynnika IF, indeksowane w bazie Web of Sciences				
- w materiałach konferencji międzynarodowych				
Lecture Notes in Geoinformation and Cartography.	A	1		15
Procedia Social and Behavioral Sciences	A	1		9
ITM Web Conferences	A	2		30
Łącznie		4		54
Monografie				
Monografie w języku polskim****	P	1		12
Dokumentacja prac badawczych - raport	A	1		0
Łącznie		2		12
Rozdział w monografii				
Rozdziały w monografiach w języku polskim	P	2		6

Pozostałe czasopisma recenzowane***				
Czasopismo	Język publikacji	Liczba publikacji	IF*	Punkty wg MNiSW *
Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska	P	1		6
Woda, Środowisko, Obszary Wiejskie	P	4		24
Annals of Warsaw Agricultural University SGGW Land Reclamation	P	2		12
Zeszyty Problemowe Postępów Nauk Rolniczych	P	4		24
Infrastruktura i Ekologia Terenów Wiejskich	P	2		11
Acta Agrophysica	P	2		12
Logistyka	P,A	11		62
Transport Problems	A	1		6
Aparatura Badawcza i Dydaktyczna	P	1		6
Acta Scientiarum Polonorum	P	1		6
Technika Transportu Szynowego	P	4		16
Journal of Transdisciplinary Systems Science	P	1		0
Autobusy Technika, Eksploatacja, Systemy Transportowe	P	9		57
Logistics and Transport	A	1		10
Roads and Bridges – Drogi i Mosty	A	1		11
Inżynieria Ekologiczna - Ecological Engineering	P	1		9
Łącznie		46		272
RAZEM		61	16,278	544

* IF oraz punktacja MNiSW zgodna z rokiem opublikowania

** IF na rok 2017

*** Zgodnie z Uchwałą nr 30/833/2015 Rady Wydziału Inżynierii Środowiska i Geodezji Uniwersytetu Przyrodniczego we Wrocławiu z dnia 22.04.2015 r. dla publikacji wydanych przed rokiem 2010 ma zastosowanie lista MNiSW z dnia 25.06.2010 r.

**** Zgodnie z Uchwałą nr 30/833/2015 Rady Wydziału Inżynierii Środowiska i Geodezji Uniwersytetu Przyrodniczego we Wrocławiu z dnia 22.04.2015 r. w odniesieniu do monografii (rozdziałów w monografii) stosuje się zasady obowiązujące w przepisach dotyczących oceny parametrycznej jednostek naukowych. Dla monografii wydanych przed rokiem 2009 przyjmuje się zasady jak w latach 2009–2012.

Tabela.3. Wykaz cytowań habilitantki (stan na 17 grudnia 2018 r.)

Baza	Liczba cytowań	Indeks H
Web of Science Core Collection	24	3
Scopus	28	4
Google Scholar	106	6

Joanna A. Kamińska