Prof. dr hab. inż. Zbigniew Nahorski
Instytut Badań Systemowych PAN

Warszawa, 19.02.2019 r.

Recenzja
poprawionej rozprawy doktorskiej mgr inż. Damiana Zasiny pt. *Analiza zmienności i niepewności szacowania emisji zanieczyszczeń ze spalania paliw w gospodarstwach domowych metodami geostatystycznymi* przygotowanej pod opieką prof. dr. hab. inż. Jarosława Zawadzkiego

Recenzję przygotowano na podstawie zamówienia Dziekana Wydziału Instalacji Budowlanych, Hydrotechniki i Inżynierii Środowiska Politechniki Warszawskiej prof. dr. hab. inż. Krzysztofa Wojdygi z dnia 23.01.2019 r.

1. Informacje i uwagi wstępne

Poprawiona rozprawa zawiera się w podobnej liczbie stron, co wersja wcześniejsza, i ma tak samo 9 rozdziałów. I chociaż na pierwszy rzut oka mogło by się wydawać, że nowa wersja różni się od pierwszej tylko mniejszą liczbą załączników (jeden w nowej wersji w stosunku do czterech w poprzedniej), to w rzeczywistości rozprawa została znacznie zmieniona. Doktorant uwzględnił praktycznie wszystkie ważniejsze uwagi z recenzji, co według mojej oceny znacznie poprawiło zarówno zawartość merytoryczną pracy, w szczególności poprawność sformułowań, jak i jej przejrzystość.

Główne zmiany dotyczące struktury rozprawy polegają na lepszym uporządkowaniu materiału: przemieszczeniu rozdziałów i dużych fragmentów tekstu, aby rozumowanie było bardziej logicznie uporządkowane, oraz opuszczeniu niektórych fragmentów, w szczególności sugerowanego w mojej recenzji podrozdziału 8.4 z poprzedniej wersji rozprawy, dotyczącego uwag o zmienności sezonowej emisji pyłu PM_{10}, co pociągnęło za sobą rezygnację z trzech związanych z tym załączników. Te zmiany struktury oceniam pozytywnie.

Zmiany struktury wymagały znacznych przeredagowań. Doktorant wziął przy tym pod uwagę wiele uwag ogólnych i szczegółowych z recenzji, co znacznie poprawiło zrozumiałość sformułowań oraz czytelność tekstu. Doktorant poprawił także podstawowe zastrzeżenia z recenzji dotyczące matematycznych sformułowań oraz wyprowadzeń wzorów wprowadzając odpowiednie poprawki, co także znacznie poprawiło czytelność rozprawy.

2. Uwagi ogólne o rozprawie

W poprzedniej recenzji dosyć dokładnie przedstawiłem zawartość poszczególnych rozdziałów. Na poziomie ogólnym nie zmieniła się ona zbytnio, więc w tej recenzji odniosę się tylko do kilku zasadniczych kwestii.

Głównym celem rozprawy jest opracowanie metody inwentaryzacji emisji pyłu PM_{10} z gospodarstw domowych w aglomeracji miejskiej z dużą rozdzielczością przestrzenną, z uwzględnieniem rozkładu obszarów zasilanych ciepłem sieciowym. Dokładne szacowanie emisji w drobnych siatkach jest zasadnicze dla jakości modelowania dyspersji zanieczyszczeń, a uwzględnienie obszarów zasilanych z sieci jest zapewne najważniejszym czynnikiem, który wpływa na dokładność szacowania emisji. Do tego celu można użyć dostępnej informacji o zasięgach sieci grzewczych. Według Doktoranta, informacje o sieciach w Polsce nie są dostatecznie kompletne. Doktorant postawił sobie więc za zadanie oszacowanie emisji w małych siatkach bez korzystania z tej informacji. Używa do tego ogólnie przyjętego sposobu

dezagregacji danych o emisjach znanych dla większych obszarów, przede wszystkim administracyjnych. Stosowane dotychczas podejście do dezagregacji w takiej sytuacji, bez uwzględniania zasięgu sieci, to zazwyczaj jedynie odjęcie całościowego ciepła systemowego na wyjściowym poziomie przed dezagregacją, co jest z pewnością mało satysfakcjonujące.

Doktorant opisuje i dyskutuje w rozprawie kilka metod szacowania emisji. W rozdziale 4 uzyskuje niejako oszacowania referencyjne dezagregując dane o spalaniu podstawowych paliw: węgla kamiennego, drewna i gazu ziemnego zgodnie z zapotrzebowaniem na ogrzewanie powierzchni mieszkalnych. Jest to dosyć typowe podejście. Następnie używa wskaźników emisji pyłu z poszczególnych paliw, w tym opracowany przez siebie (autorski) wskaźnik emisji pyłu PM_{10} ze spalania węgla. W ten sposób Doktorant uzyskuje mapę emisji w rozpatrywanej aglomeracji śląskiej z rozbiciem na poszczególne gminy. W tych obliczeniach nie uwzględnia, które gminy i w jakim stopniu korzystają z ciepła systemowego.

W pozostałych metodach korzysta przy dezagregacji ze zmiennych stowarzyszonych (ang. *proxy*), zwanych w pracy surogatami: gęstości zaludnienia lub gęstości zabudowy, używając głównie pierwszej z nich, co jest rzeczywiście częściej przyjmowane w takich obliczeniach. Postępowanie z dezagregacją wprost względem gęstości zaludnienia nazwano w rozprawie metodą bezpośrednią.

Następnie, obie gęstości są, w uproszczeniu, przetworzone z zastosowaniem krigingu (po zlogarytmowaniu danych). Dla gęstości zaludnienia używa się potem dodatkowo sekwencyjnej symulacji gaussowskiej (ang. *sequential gaussian simulation* – SGS) – w rozdziale 5, a dla gęstości zabudowy metody walidacji krzyżowej (ang. *leave-one-out crossvalidation* – LOO) połączonej z metodą scyzoryka (ang. *jacknife*) – w rozdziale 6, przy czym wyniki uzyskane dla gęstości zabudowy są właściwie użyte tylko do wykazania, że oba rozkłady – gęstości zaludnienia i gęstości zabudowy – są podobne. W obu przypadkach dodatkowe metody (SGS lub LOO) służą do poprawy oszacowania wariancji rozkładów.

W metodzie zwanej „metodą pośrednią z redukcją zaludnienia w obrębie *hot-spotu*" (metoda pośrednia poziom 1), na podstawie tak przetworzonych danych Doktorant szacuje rozkłady prawdopodobieństwa gęstości zaludnienia w oczkach siatki, a następnie zmniejsza o połowę gęstość zaludnienia w obszarach tzw. *hot-spotu*, czyli gęstości zaludnienia wystających ponad przyjęty próg (na podstawie przyjętych percentyli rozkładu dla prawdopodobieństw 0,75; 0,85; 0.95) i dopiero takich danych używa jako *proxy* przy dezagregacji. Jest to właśnie nowa metoda zaproponowana przez Doktoranta.

W metodzie pośredniej z tak zwaną „maską" (metoda pośrednia poziom 2) Doktorant korzysta z informacji o zasięgu sieci ciepłowniczej i zmniejsza przy dezagregacji o połowę gęstość zaludnienia w tych oczkach siatki, w których jest doprowadzone ciepło sieciowe.

Porównano wyniki uzyskane tymi trzema metodami (metoda bezpośrednia i metody pośrednie: poziom 1 i poziom 2), zarówno pod względem wartości średnich, jak i różnych wskaźników uwzględniających wariancję lub odchylenie standardowe. Jednak wyniki porównania wydają się być niejednoznaczne. Z porównania tzw. scenariuszy w Tabl. 7.1 i Tabl. 7.2 wynika, że wyniki szacowania emisji metodą bezpośrednią i metodą pośrednią poziom 1 mają zbliżone rozkłady, gdy tymczasem rozkład wyników uzyskanych metodą pośrednią poziom 2 bardzo od nich odbiega. Natomiast porównanie rozkładów tzw. wskaźnika walidacji w Tabl. 8.4 wykazuje, że rozkłady pokrywają się w zasadzie dokładnie dla obu metod pośrednich, gdy tymczasem rozkład dla metody bezpośredniej (zwanej tu bazowym), różni się od nich wyraźnie. Wynik drugiego porównania wydaje mi się intuicyjnie bardziej przemawiający, bo właśnie w obu metodach pośrednich następuje redukcja emisji w obszarach podejrzewanych o dopływ ciepła sieciowego. Może najwyżej zaskakująca jest aż tak duża zbieżność rozkładów dla obu metod pośrednich w drugim przypadku.

Chociaż sprawa dokładności oszacowań emisji uzyskanych na podstawie dostępnej informacji o zasilaniu systemowym lub metodą Doktoranta może być dyskusyjna, to metoda

zaproponowana przez Doktoranta może być także stosowana bez informacji o zasięgach sieci cieplnych, jeżeli tylko wiadomo, że one istnieją, na przykład w krajach, w których taka szczegółowa informacja nie jest dostępna. Do tego, opracowana w rozprawie metoda ma charakter ogólny i może być stosowana także w innych zagadnieniach. Doktorant opracował więc oryginalną metodę bazującą na zaawansowanym aparacie teoretycznym oraz wymagającą użycia zaawansowanego oprogramowania komputerowego. Jest to wynik o poziomie naukowym właściwym dla przewodu doktorskiego.

Doktorant szacuje niepewności emisji, jednak w zdecydowanej większości przypadków rozumie przez to oceny związane ze zmiennością przestrzenną ocen emisji. Oszacowania takie są oczywiście potrzebne do wykazania tezy 3 (T3) pracy. Jednak sformułowanie tezy jest precyzyjne, a sformułowania w tytułach podrozdziałów 7.4 i 8.2 mogą wprowadzać w błąd.

Prawdziwe oszacowanie niepewności emisji jest właściwie dokonane w podrozdziale 8.1. Uzyskany wynik 51% nie jest zaskakujący dla ocen emisji w tak drobnych siatkach, jak rozpatrywane z rozprawie. Jest to oszacowanie eksperckie, jednakowe dla wszystkich metod. Nie pozwala ono na rozstrzygnięcie, która z metod jest dokładniejsza. Porównanie wskaźników walidacji w Tabl. 8.4 sugeruje, że mogą one być jednakowo dokładne. Mam jednak pytanie do Doktoranta, czy i jak można by było przynajmniej spróbować sprawdzić jeszcze w inny sposób, która metoda daje dokładniejsze wyniki.

W sumie ważne jest jednak to, że Doktorant zwrócił uwagę na konieczność uwzględniania obszarów objętych sieciami cieplnymi w takich metodach szacowania emisji w drobnych siatkach, które nazwałbym „automatycznymi", co jest lub przynajmniej było nowością przy rozpoczęciu badań, biorąc pod uwagę istniejącą literaturę. Trudno w tej chwili zdecydować, w którym kierunku pójdą badania w tym obszarze naukowym, niemniej jednak obserwuje się wyraźny trend w rozwoju właśnie „automatycznych" metod szacowania emisji, których można by było użyć w dowolnych przypadkach, bez lub z niewielkimi adaptacjami. Badania Doktoranta mieszczą się właśnie w tym nurcie.

Być może perspektywicznym rozwiązaniem tego zagadnienia będzie jednak mocno pracochłonne podejście bazujące na inwentaryzacji cieplnej indywidualnych budynków, na przykład przez ankietowanie, jak użyte przy tworzeniu atlasów obejmujących charakterystyki poszczególnych domów, na przykład w Danish Heat Atlas, Heat Atlas Saxony, czy London Heat Map. Takie atlasy istnieją (chociaż nie do końca mogą być kompletne) i są chyba prowadzone prace, lub przynajmniej się o nich myśli, nad użyciem ich do estymacji emisji.

Jeżeli Doktorant będzie chciał kontynuować badania, to proponowałbym także zastanowić się, czy nie mogłyby być przydatne w nich mapy zapotrzebowania na ciepło opracowane całkiem niedawno dla Europy w Pan-European Thermal Atlas (Peta) – https://heatroadmap.eu/peta4/, chociażby do porównań z oszacowaniami z rozdziału 4.

3. Uwagi szczegółowe

Poniżej przedstawiam zauważone przeze mnie drobne błędy i przeoczenia. Łatwo zauważyć, że są to drobne niedociągnięcia, które łatwo można poprawić przy starannej redakcji.

Wzór (2.9) – Nie poprawiono znaków, pomimo tego, że pisałem o tym w poprzedniej recenzji.

Wzór (2.19) – Po prawej stronie równości powinno być $\hat{Z}(x_0) - E\{\hat{Z}(x_0)\}$.

Wzór (2.21) – Kilka razy podobny błąd, jak w (2.19).

Wzór (2.35) – Po lewej stronie równości powinno być $L(\lambda, \mu)$.

Objaśnienie do wzoru (3.4) – Indeks p nie jest wzięty w nawias, więc nie ma co o tym pisać.

Drugi i trzeci wiersz po równaniu (3.8) – Chyba do stałości wartości oczekiwanej, niezależnie od położenia, potrzebne jest założenie stacjonarności procesu, a nie jego ergodyczności?

Str. 65, 9-10 w. od góry – Zamiast 19% powinno być chyba 29%, a opis, co to jest 43%, jest niejasny.

Str. 77, 14 w. od góry – Precyzyjniej „powyżej której jest duże prawdopodobieństwo wystąpienia sieci ciepłowniczej".

Str. 77, 6 w. od dołu – W tabl. 1.2 nie ma odsetków odbiorców ciepłownictwa sieciowego.

Str. 88, pierwszy akapit podrozdziału 5.4 – To prawda, że scenariusze w rozkładach ciągłych mają jednakowe prawdopodobieństwo wystąpienia; wynosi ono zero. Czy o to chodziło w tym urywku? Natomiast jeżeli rozważamy skończoną liczbę scenariuszy, to mam wątpliwości, czy mają one jednakowe prawdopodobieństwa zaistnienia.

Rys. 6.3 – Jak się mają wartości rzędnych do współrzędnych geograficznych?

Rys. 6.5 – Co przedstawia wykres (a), a co (b)?

Str. 98, 6 w. od góry – Warto tu wspomnieć, że porównanie wartości SSE jest tu poprawne, ponieważ wszystkie modele (z wyjątkiem samorodka) mają tyle samo dopasowywanych współczynników (ten wątek jest dyskutowany dalej).

Rys. 6.8 – Duże optyczne różnice między (a) i (b) są chyba wynikiem przyjęcia różnych skal na wykresach. Z tabl. 6.2 wynika, że wartości MSE i MAE dla różnic są bardzo małe.

Punkt 7.2.2 – Czy rzeczywiście wystarczy wystąpienie sieci, nawet o znikomej długości, aby przyjąć, że jest tam potrzebna maska? Może lepiej uzależnić wartość maski od procentowego pokrycia oczka siatki przez sieć?

Wzór (7.4) – Co to jest „cover"?

Podrozdział 7.4 – W tym tytule powinno być przynajmniej dodane „rozkładu przestrzennego".

Str. 144, od w. 2 od góry – Niezbyt rozumiem, jak „koncepcja liniowej gęstości cieplnej okazała się przydatna na potrzeby walidacji metody".

Str. 144, ostatni punkt podrozdziału 9.2 – Po usunięciu podpunktu o rozkładzie czasowym powinno się usunąć także ten punkt.

4. Uwagi końcowe

Uważam, że najważniejsze osiągnięcia Doktoranta to:

- Zwrócenie uwagi na to, że przy zautomatyzowanym szacowaniu pola emisji pyłu PM_{10} w drobnych siatkach powinno się wziąć pod uwagę rozkład obszarów zasilanych ciepłem sieciowym, aby zwiększyć dokładność szacowania pola emisji, a tym samym dokładność modelowania dyspersji zanieczyszczeń w powietrzu.
- Opracowanie kilku metod szacowania emisji z zastosowaniem zaawansowanych narzędzi statystycznych. W tym opracowanie nowej, zaproponowanej przez Doktoranta metody z szacowaniem obszarów zasilanych ciepłem sieciowym.
- Wykonanie oszacowań tymi metodami dla aglomeracji śląskiej i porównanie uzyskanych wyników.

Warto też zauważyć, że Doktorant wykazał się w pracy dużą erudycją, niezbędną zresztą w tak interdyscyplinarnych badaniach, łączących w sobie wątki z wielu dziedzin nauki. Doktorant potwierdził także w rozprawie wszystkie postawione przy otwieraniu przewodu doktorskiego tezy.

Uważam, że wyniki uzyskane przez Doktoranta spełniają wymagania na rozprawy doktorskie nałożone przez odpowiednie przepisy. Wnioskuję o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.