

С Т А Н О В И Щ Е

от проф. д.м.н. Снежана Георгиева Гочева-Илиева,
Пловдивски университет „Паисий Хилендарски”

на дисертационен труд за присъждане на образователната и научна степен „доктор”

в област на висше образование: 4 Природни науки, математика и информатика,
професионално направление: 4.5 Математика,
докторска програма: Математическо моделиране и приложение на математиката

Автор: Мая Пламенова Стоименова

Тема: МОДЕЛИРАНЕ НА БЪРЗОПРОМЕНЛИВИ ВРЕМЕННИ РЕДОВЕ

Научен ръководител:

Проф. д.м.н. Снежана Георгиева Гочева-Илиева – ПУ „Паисий Хилендарски”

1. Общо представяне на процедурата и докторанта

Със заповед РЗЗ-1350/27.03.2018 на Ректора на ПУ „Паисий Хилендарски“ съм назначена за член на научното жури по настоящата процедура. С първо заседание на НЖ съм определена за написване на становище, в качеството ми на научен ръководител на докторанта. Предоставен е пълен пакет документи за защита.

Мая Стоименова е родена през 1988 година. Възпитаник е на ПУ „Паисий Хилендарски“. В периода 2007-2011 г. е редовна студентка по специалността Приложна математика във Факултета по математика и информатика, която завършва с диплома за бакалавър. През 2011-2013 г. се обучава и завършва магистратура по специалността Управление на човешките ресурси във Факултета по икономически и социални науки. От 01.03.2015 г. до 01.03.2018 г. е редовна докторантка в катедра Приложна математика и моделиране на ФМИ на ПУ „Паисий Хилендарски“. По време на докторантурата полага необходимия минимум, както и се обучава допълнително по 2 специализирани докторантски курса в чужбина: „Numerical-analytic and constructive methods for boundary-value problems“ – летен университет към СЕЕПУС мрежата СИИ-НУ-0028-08-1415 „Active Methods in Teaching and Learning Mathematics and Informatics”, Университета в Мишколц, Унгария и „Intensive course on time series analysis“, DAAD проект, Факултет по математика и информатика, Университета в Нови Сад, Сърбия. Участва активно и в 3 научни проекта към НПД на ПУ. По време на обучението си в докторантура Мая Стоименова подготви и проведе възложените ѝ от катедра ПММ учебни занятия със студенти по дисциплината Иконометрия за специалност Бизнес математика. Счита, че тези дейности напълно отговарят на изискванията за образователна и научна степен „доктор” по смисъла на закона.

Представеният дисертационен труд от Мая Стоименова е с общ обем от 139 печатни страници. Включва увод, 4 глави, заключение, библиография от 123 източника. Темата и съдържанието на дисертационния труд са в съответствие с професионалното направление и докторската програма. Всички представени документи удовлетворяват изискванията по процедурата за защита на дисертационен труд за придобиване на ОНС „доктор”, заложен в ЗРАСРБ и Правилника за неговото прилагане, ПРАС на ПУ и Специфичните изисквания на Факултета по математика и информатика на ПУ.

2. Актуалност на тематиката

В дисертационния труд се разглеждат задачи по моделиране на наблюдения, описвани с временни редове за случая на бързопроменливи редове, които се характеризират с относително голям размах и силно вариране на данните. По-специално се изследват свойствата и приложението на два метода - стохастичният подход на Бокс-Дженкинс за моделиране с авторегресия и плаващо средно (ARIMA) и дейта майнинг методът на класификационните и регресионни дървета (CART). Голям брой публикации са свързани с

тези методи и различни аспекти от тяхното приложение, като актуалността им е демонстрирана добре в цитираната в дисертацията литература. Приложенията на тези методи в дисертационното изследване са ориентирани към много съществената и актуална предметна област по моделиране и анализиране на временни редове за замърсители на атмосферния въздух в българските градове. Подобни изследвания за България по същество са крайно недостатъчни в съществуващата литература, за разлика от останалите страни.

Избраната тематика е изцяло в областта на моделирането и приложението на математиката, а методиката на построяване на модели се базира на реални измерени данни и процедури за прогнозиране, поради което тяхната практическа стойност е висока.

3. Познаване на проблема

Приведена е голяма по обем библиография от 123 източника, отразяваща публикации с изследвания в дисертацията тип временни редове. Направен е обстоен литературен обзор на използваните методи, приложенията за моделиране на въздушни замърсявания, както и на вредните последици от тях за здравето на населението. Направеният литературен обзор показва, че докторантката познава добре конкретната приложна област, тематичните проблеми, математическите и компютърни методи за моделиране.

4. Методика на изследването

Раглежданите задачи за математическо моделиране обхващат временни редове, съдържащи линейни и нелинейни зависимости, включително авторегресионни. По тази причина изключително подходяща методика на изследването е подборът на ARIMA, предназначен за линеен тип авторегресионни зависимости и CART, който има гъвкави качества за описание както за линейни така и за нелинейни локални шаблони за широк клас от данни с всички типове променливи - номинални, порядкови, непрекъснати.

5. Характеристика и оценка на дисертационния труд и приносите

Целта на представения дисертационния труд е **разработка на висококачествени статистически модели за бързопроменливи временни редове и приложението им за описание, анализ и краткосрочни прогнози на замърсители на атмосферния въздух.** По-конкретно приложната предметна област са замърсяванията на въздуха с PM2.5 и PM10 (прахови частици с аеродинамичен размер съответно под 2.5 и до 10 микрона). Следва ясно да се отбележи, че източниците на тези замърсявания за всяко населено място са с разнороден характер и се обуславят от специфични географски, атмосферни, метеорологични, климатични, транспортни, производствени, битови и други фактори. По тези причини, събраните данни във вид на временни редове са с комплексно поведение, което ги прави трудни за моделиране. Изследваните данни са от наблюдения на среднодневни концентрации на фини прахови частици, а където са налични – и от временни редове на метеорологични и друг тип променливи.

Глава 1 съдържа въведение с подробен анализ на състоянието на изследванията в областта на моделирането на замърсителите на въздуха и прилаганите в литературата методи, кратко описание на приложената методика с ARIMA и CART, критериите за оценка на моделите и др. Определени са целта и задачите на дисертационния труд.

В Глава 2 е проведено ARIMA и CART моделиране на временни редове на проблемни замърсявания с фини прахови частици (PM10) на въздуха на град Перник за период от 5 години. Построени и анализирани са едномерни АРИМА(1,0,5) модели на PM10, описващи около 56% от данните. Полученият оптимален модел е приложен за краткосрочно прогнозиране в хоризонт от 7 дни напред. За целта са използвани реални допълнителни данни, които не са участвали при построяването на модела. Сравнението показва много добро съответствие на модела с реалните измерени стойности на PM10. Във втората част на главата към същите данни са използвани и данни за трите основни въздушни замърсители – прекурсори на PM10 – въглероден оксид (CO), серен диоксид (SO2) и азотен диоксид (NO2), както и 8 метеорологични променливи, авторегресионни (лагирани) редове на PM10 от предходните 1 или 2 дни за всеки член на реда, а също и времеви променливи.

Приложен е CART метод и са построени голям брой CART модели. От тях са подбрани 4, обясняващи до над 90% (R^2) от данните и притежаващи отлични статистически показатели за съвпадение с изследваните, и за прогнозиране със 7-дневен хоризонт. Установено е, че най-съществени фактори за замърсяването с PM10 на въздуха на Перник са концентрацията на CO, следвана от стойността на PM10 от вчерашния ден, а с по-слабо влияние са концентрациите съответно на NO2 и SO2. Направено е сравнение с ARIMA модела. Моделите са получени и изследвани с помощта на SPSS и Salford Predictive Modeler.

Глава 3 представя резултати от стохастичен едномерен и многомерен анализ на времеви редове с ARIMA метод на замърсителите на въздуха PM2.5 и PM10 за градовете Пловдив и Асеновград. Данните са среднодневни, измерени в 4 измервателни станции за период от 5 години. Тъй като променливите нямат нормално разпределение, изисквано в ARIMA методологията, е направена начална трансформация за подобряване на разпределението с помощта на формулата на Йео-Джонсън. Избраните най-добри едномерни модели обясняват от 59% до 66% от данните. Резултатите показват, че изходните данни и съответните модели са много близки по характер и от тип ARIMA(1,0,5), ARIMA(1,0,4). Моделите са приложени за прогнозиране на концентрациите на замърсителите с 3 дни напред, с данни извън използваните в модела. По-нататък в глава 3 са построени многомерни (векторни) модели ARIMA от тип 2D, 3D и 4D, които описват замърсяването чрез система взаимни уравнения. Моделите са приложени за прогнозиране с три дни напред. Те са от тип ARIMA(1,0,3) - ARIMA(1,0,5) за PM10 и ARIMA(1,0,0) с участие на PM2.5. С получените модели се установява, че няма съществена разлика в замърсяването с фини прахови частици на целия район от Тракийската низина, включващ Пловдив и Асеновград. Оттук индиректно следва, че то не зависи от разположението на измервателните станции. Изследването е проведено със софтуерния пакет Wolfram Mathematica.

Последната глава 4 разглежда приложението на CART метод за моделиране и предсказване на замърсяванията с PM10 на град Плевен в период от 6 години. Целта в тази глава е създаване на висококачествени математически CART модели чрез изследване влиянието на различни елементи на процеса на моделиране: прилагане на машинно обучение с кросвалидация, влияние на типа на разпределението, моделиране със и без авторегресионни предиктори, подбор и влияние на метеорологични променливи, оценка на качеството на моделите с контингентни таблици, определяне уловията за максимално замърсяване (ouliers), и др. От големия брой получени CART модели в дисертацията са избрани и описани свойствата на 5 оптимални модела без кросвалидация и 6 – с кросвалидация за машинно обучение. Получено е много добро съвпадение с експеримента с до 78% на модели без кросвалидация и с до 84% на модели с кросвалидация. Направено е приложение на моделите за прогнозиране на замърсяването с 2 дни напред извън периода на моделиране. Установена е силата на влияние на отделните променливи, като най-голямо значение за високите концентрации на PM10 са концентрациите на PM10 от предишния ден, а след това - минимална температура от вчерашния ден. Моделите са получени и анализирани с помощта на SPSS и Salford Predictive Modeler.

Всички получени в дисертацията модели, анализи и резултати следва да се разглеждат като алтернатива на официалните числени и друг тип методи, на които се базират докладите на регионалните инспекции по околната среда и водите в България. Резултатите показват, че прилаганите ARIMA, CART и дейта майнинг техники с машинно обучение са потенциална възможност за провеждане на независими изследвания, прогнозиране, превенция и контрол на чистотата на въздуха в българските градове. Те са напълно приложими и за реални прогнози, с използване на метеорологични данни –прогнозни или лагирани.

Приемам напълно претенциите за научни приноси, формулирани в заключението на дисертационния труд.

6. Преценка на публикациите и личния принос на докторанта

Основната част от резултатите от изследванията са публикувани в три публикации, от които първата е самостоятелна, а останалите са в съавторство. Две от публикациите са издадени в трудове на конференции на престижния American Institute of Physics (AIP), индексирани в Scopus и Web of Science. Приносът на докторантката в публикациите със съавторство е съществен както в построяване на моделите, провеждане на статистическите анализи и анализиране на моделите, така и в тяхното оформяне и представяне.

7. Автореферат

Авторефератът е с обем от 32 страници и коректно отразява резултатите на дисертационния труд.

8. Препоръки за бъдещо използване на дисертационните приноси и резултати

Нямам съществени забележки. Препоръчвам провеждането на по-нататъшни изследвания и разширяване на обхвата на постигнатите приноси в избраната актуална област на моделиране и приложение на математиката.

ЗАКЛЮЧЕНИЕ

На база на постигнатите от докторантката Мая Стоименова оригинални научни резултати от областта на моделиране на временни редове с актуален приложен характер, изложени в представения дисертационен труд, автореферат и публикации, мога убедено да твърдя, че целите и задачите на изследването са постигнати. В качеството ми на научен ръководител изказвам мнението, че и с цялостната си дейност – обучението в докторантурата, участие в семинари и конференции, работа по научни проекти, работа с модерен софтуер и методи, докторантката показва, че притежава задълбочени знания и професионални умения за провеждане на научни изследвания, способности за самостоятелна научна работа и работа в екип.

Въз основа на по-горе изложеното, моята оценка за дисертационния труд на Мая Пламенова Стоименова е ПОЛОЖИТЕЛНА.

Напълно подкрепям получените научни резултати и приноси и предлагам на Научното жури да присъди образователната и научна степен „доктор” на **Мая Пламенова Стоименова** по: Област на висше образование: 4. Природни науки, математика и информатика, Професионално направление: 4.5 Математика, докторска програма: Математическо моделиране и приложение на математиката.

08.05.2018 г.

Изготвил становището:

/проф. д.м.н. Снежана Гочева-Илиева/