

## **РЕЦЕНЗИЯ**

от проф. д.н. Иван Томов Димов,  
Институт по информационни и комуникационни технологии,  
Българска академия на науките

### **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**

Определен съм за член на научното жури (НЖ) по настоящата процедура със заповед РЗЗ-1350/27.03.2018 на Ректора на ПУ „Паисий Хилендарски“. На първото заседание на НЖ съм определен за рецензент. Настоящата рецензия е написана на база на получения от мен пълен пакет от документи за придобиване на ОНС „доктор“, описани в Правилника на ПУ „П. Хилендарски“, законовите изисквания и Специфичните изисквания на Факултета по математика и информатика (ФМИ).

Мая Пламенова Стоименова завършва висшето си образование във ФМИ на ПУ „Паисий Хилендарски“ през 2011 г. по специалността Приложна математика и магистърска степен през 2013 г. по специалността Управление на човешките ресурси във Факултета по икономически и социални науки на ПУ. От 2015 г. до 2018 г. е редовна докторантка в катедра Приложна математика и моделиране на ФМИ. В периода на обучението си в докторантура е положила успешно необходимия минимум, а също така е завършила и следните допълнителни курсове: “Numerical-analytic and constructive methods for boundary-value problems” в рамките на CEEPUS Summer University – 2015, Мишколц, Унгария; “New Aspects of the Time Frequency Analysis Involving Fréchet Frames”, DAAD Intensive course, Нови Сад, Сърбия (2016 г.), езикови курсове по английски език – ниво А1 и А2. Взела е участие в 3 научни университетски проекта. Учебната ѝ работа включва подготовка и провеждане на учебни занятия по дисциплината Иконометрия със студенти от специалност Бизнес математика на ФМИ. Считаю, че изброените дейности съответстват изцяло на изискванията за образователна и научна степен „доктор”.

Представеният за рецензиране дисертационен труд на Мая Стоименова е разработен на 139 страници. Структурата му обхваща увод, четири глави, заключение и библиография от 123 източника. Темата и съдържанието на дисертационния труд изцяло отговарят на професионалното направление и докторската програма. Всички представени ми документи са оформени в надлежен вид и съгласно изискванията в процедура за защита на дисертационен труд за придобиване на ОНС „доктор”, описани в ЗРАСРБ, Правилника за ЗРАСРБ, ПРАС на ПУ и Специфичните изисквания на Факултета по математика и информатика на ПУ.

## **2. Актуалност на тематиката**

Актуалността на разглежданата в дисертацията тематика се обуславя на първо място от важността на реалния изследван проблем – замърсяванията за атмосферния въздух в българските градове, който безспорно има голямо не само научно, но и обществено значение и пряко влияе на здравето на населението. На второ място – това е прилагането на най-съвременни математически модели на специфичен вид времеви редове, характеризиращи се с бързопроменящи се наблюдения. Става дума за случайни процеси с много ниска гладкост. Фокусът е върху изследване на 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). В тази област има голям брой публикации. Актуалността на тематиката е добре обоснована в цитираната в дисертацията литература. В дисертацията са направени оригинални приложения на тези методи, ориентирани към много съществената и актуална предметна област по моделиране и анализиране на временни редове за аерозолни замърсители на атмосферния въздух.

Тематика на дисертационната работа е изцяло в областта на докторската програма по математическо моделиране и приложение на математиката, а развитата методика на построяване на модели се базира на реални измерени данни, поради което тяхната практическа стойност е съществена.

## **3. Познаване на проблема**

Дисертантката е добре запозната с проблематиката на изследването. Това следва от факта, че е анализирана достатъчно голяма по обем библиография от 123 източника, отразяваща публикации с изследвания в дисертацията тип времеви редове. Представен е много обстоен литературен обзор на използваните методи, приложенията за моделиране на въздушни замърсявания, и в частност с фини прахови частици и суспендирани частици, а също и проблемите, свързани с вредите, нанасяни от такъв тип замърсявания върху здравето на населението в засегнатите градове и региони. Очевидно е, че докторантката познава добре както математическите и компютърни методи за моделиране, най-съвременен софтуер и ИСТ средства, и тематичната област.

## **4. Методика на изследването**

Избраната и приложена методика на анализирани в дисертацията проблеми е насочена за моделиране на времеви редове, неявно съдържащи линейни, нелинейни и авторегресионни зависимости. Смятам, че именно заради това, методиката на изследването е избрана адекватно. ARIMA е предназначен за линеен тип авторегресионни зависимости; CART има добри качества за описание както за линейни, така и за нелинейни локални зависимости за широк клас от данни за различни типове променливи - номинални, порядкови и непрекъснати, и работи добре с големи извадки от наблюдения.

## **5. Характеристика и оценка на дисертационния труд и приносите**

В представения дисертационния труд е формулирана целта на изследването, а именно: *разработка на висококачествени статистически модели за бързопроменящи се времеви редове и приложението им за описание, анализ и краткосрочни прогнози на замърсители на атмосферния въздух*. Изследванията са концентрирани основно върху

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). Важно е да се знае, че източниците на тези замърсявания за всяка област или населено място са различни. Те могат да зависят от множество фактори като климат, географско разположение на населеното място, надморска височина, ландшафт, производствени мощности, наличие на източници от битов характер, производствени процеси с горивни и отпадъчни процеси, засилен автомобилен трафик, отделяне на въглероден оксид и други прекурсори на фините прахови частици, и др. По тези причини, събраните данни във вид на времеви редове имат комплексен характер. Всички използвани данни са от наблюдения на среднодневни концентрации на фини прахови частици – PM10 и PM2.5. Комплексността на проблема идва и от това, че метеорологичните фактори, адвекцията и дифузията на частиците, както и влажността и валежите могат да доведат до резки изменения дори и на средните стойности. С една дума, това е задача с много слаба регулярност на входните данни, което я прави и много трудна.

**Основни задачи на дисертационния труд са:**

- 1) Създаване и приложение на стохастични модели за изследване на данни за PM10, проблемен замърсител на въздуха в град Перник за период от 5 години.
- 2) Построяване на високоефективни математически модели базирани на CART метод за изследване на временния ред за PM10 в град Перник в зависимост от метеорологични данни и основните прекурсори, и приложение на моделите за краткосрочни прогнози.
- 3) Построяване и приложение на многомерни стохастични модели за анализиране на временни редове за PM10 и PM2.5 за градовете Пловдив и Асеновград.
- 4) Моделиране на временни редове за PM10 в град Плевен, базирано на CART методологията, като се използват трансформации на данните.
- 5) Анализиране на CART моделите за Плевен без трансформация на данни и приложение на резултатите за бъдещи прогнози.

**Глава 1** представя подробен анализ на състоянието на изследванията в областта на статистическото моделиране на замърсителите на въздуха. Дадено е кратко описание на използваните в дисертационния труд статистически методи. Дефинирани са целта и задачите на дисертационния труд. Дефинирани са ARIMA (авторегресия, интегрирана, с плаващи средни) метод и методът CART (Classification And Regression Trees).

В **Глава 2** са представени резултати от 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. За обработка на временния ред на PM10 са построени и изследвани едномерни ARIMA модели. Чрез избрани критерии е подбран най-добър модел ARIMA(1,0,5). Този модел отчита влиянието на замърсяването с един ден назад спрямо текущия и използва изглаждане на грешките с пет дни назад. Моделът постига съвпадение с данните с коефициент на детерминация  $R^2=56.4\%$ . Общата среднаквадратична грешка на модела е  $RMSE=32.762$ . Направен е математически анализ на грешките на модела. Моделът е приложен за предсказване на замърсяването с PM10 на град Перник за 7 дни напред и демонстрира много добри прогнозиращи качества. 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. За същите данни е приложен CART метод, като за предиктори са използвани 14 на брой временни редове, от които 8 са метеорологични променливи, две променливи, които отчитат влиянието на замърсяването от последните два дни, CO, NO<sub>2</sub>, SO<sub>2</sub> и стойността на CO от предишния ден. По същите критерии е избран най-добър CART модел с R<sup>2</sup> =93.7. Моделът е изследван от гледна точка на неговата точност и има RMSE= 12.851. Избраният най-добър CART модел отчита, че най-съществено влияние за замърсяването с PM<sub>10</sub> за текущия ден имат следните променливи: на първо място концентрацията на CO, следвана от стойността на PM<sub>10</sub>, измерена от вчерашния ден, а трети и четвърти по ред са концентрациите съответно на NO<sub>2</sub> и SO<sub>2</sub>. Моделът е приложен за предсказване на PM<sub>10</sub> за следващите 7 дни и показва добри прогнозни резултати.

В **Глава 3** са представени резултати от стохастичен едномерен и многомерен анализ на времеви редове с ARIMA метод на въздушните замърсители PM<sub>2.5</sub> и PM<sub>10</sub> 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. Направено е приложение на моделите за краткосрочен период на прогнозиране от 3 дни, за които съответните данни не са били използвани в построяването на моделите. В същата глава 3 са построени многомерни (векторни) модели с ARIMA от тип 2D, 3D и 4D, които описват анализиранията концентрации на замърсителите чрез системи уравнения. Те са от тип ARIMA(1,0,3) - ARIMA(1,0,5) за PM<sub>10</sub> и ARIMA(1,0,0), когато са използвани данни за PM<sub>2.5</sub>. От многомерните ARIMA модели може да се направи изводът, че замърсяването с фини прахови частици на целия район от Тракийската низина, включващ Пловдив и Асеновград е почти еднакъв. В тази глава е използван пакетът Wolfram Mathematica.

В **глава 4** са е приложен CART метод за моделиране и предсказване на замърсяванията с PM<sub>10</sub> на град Плевен в период от 6 години (2011 – 2016). Избрани и описани са свойствата на 5 оптимални CART модела без кросвалидация и 6 оптимални модела с кросвалидация за моделиране на нивата на PM<sub>10</sub> в зависимост от 8 метеорологични променливи и 1 времева променлива. Освен това модели са построени и за трансформация на PM<sub>10</sub>, с която е получено близко до нормалното разпределение. Моделите са построени и изследвани със софтуерните пакети SPSS и Salford Predictive Modeler. Проведени са анализ и диагностика на моделите. Най-добрите модели достигат R<sup>2</sup>=84% и RMSE=11.694. Установено е влиянието на използваните променливи, което показва, че най-съществена е авторегресионната връзка с предходния ден за PM<sub>10</sub> и същия тип връзка за минималната температура. Моделите обясняват много добре измерените стойности на PM<sub>10</sub>, с коефициент на детерминация R<sup>2</sup>=78% без кросвалидация и с R<sup>2</sup>= 84% с кросвалидация. Направено е приложение на получените модели за получаване на прогнози на PM<sub>10</sub> за 2 дни напред след използваните измервания. В дадения случай най-голямо значение за високите концентрации на PM<sub>10</sub> са авторегресионната зависимост с PM<sub>10</sub> от предишния ден и измерената минимална температура от вчерашния ден.

Получените в дисертацията емпирични резултати във вид на математически модели, проведените с тях анализи и изводи могат да се използват като една добра алтернатива на официалните числени методи, използвани от агенцията по околна среда и водите в България. Приложената методологията показва, 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 населени места и региони на България. Те могат да бъдат използвани и за реални прогнози с използване на метеорологични данни.

## 6. Приноси и значимост на разработката за науката и практиката

Приемам авторските претенции за научни приноси, формулирани в заключението на дисертационния труд.

## 7. Преценка на публикациите и личния принос на докторанта

Резултатите от изследванията са публикувани в общо три публикации. От тях една е самостоятелна, а останалите две са в съавторство, с участие на научни ръководител. Съвместните публикации са докладвани на международна конференция и са издадени в трудовете на American Institute of Physics (AIP), индексирани са в Scopus и Web of Science. Предполагам, че приносът на докторантката в публикациите със съавторство е съществен и равностоен, тъй като не е декларирано друго твърдение.

## 8. Автореферат

Авторефератът коректно отразява резултатите на дисертационния труд.

## 9. Забележки и препоръки за бъдещо използване на дисертационните приноси и резултати

Имам някои забележки и препоръки към дисертантката и нейния дисертационен труд.

- Независимо, че в дефиницията на страница 28 от дисертацията се казва „временен (или времеви) ред“ смятам, че е правилно да се използва понятието „времеви“. За съжаление, на много места в дисертацията се използват и двете понятия, а те в общия случай означават различни неща. Смятам, че използването на понятието „бързопроменлив“ също е неправилно; би следвало да се използва „бързопроменящ се“.
- В работа [85] се цитира метод за идентификация на източници на замърсяване, който използва методът на главните елементи. Този метод има редица недостатъци. Бих препоръчал използването на метода, изложен в **I. Dimov, U. Jaekel, H. Vereecken, D. Wendt, A numerical approach for determination of sources in reactive transport equations**, Second International Conference on Monte Carlo and Quasi-Monte Carlo methods in scientific computing, University of Salzburg, 8-12 July, 1996, (in: **Proceedings of MC & QMC 96, Springer Notes in Statistics** (H. Niederreiter, P. Hellekalek, G. Larcher and P. Zinterhof, Eds)), 1997, pp. 189 - 204. Методът, предложен в горната работа използва дуално (спрегнато) решение и е значително по-общ и точен.
- Основните разглеждания в дисертацията са фокусирани върху PM10 и PM2.5. В същото време последните резултати за въздействието на аерозолни частици върху човешкото здраве показват, че наночастиците (които се отделят от катализаторите на ауспусите на автомобилите) са канцерогенни и могат да бъдат значително по-вредни). Смятам, обаче, че използваните в дисертацията методи са доста общи и биха могли да третират и този проблем, ако има данни за концентрацията на тези наночастици.

Искам да подчертая, че направените забележки и препоръки на омаловажават постиженията на дисертантката, а биха могли да се разглеждат като идеи за по-нататъшни изследвания.

### **ЗАКЛЮЧЕНИЕ**

Като има предвид изложеното по-горе, мога да заявя, че докторантката Мая Стоименова е получила оригинални научни резултати в областта на моделиране на времеви редове с актуален приложен характер. Убедено твърдя, че целите и задачите на изследването са постигнати.

**Въз основа на проведения анализ, моята оценка за дисертационния труд на Мая Пламенова Стоименова е ПОЛОЖИТЕЛНА.**

Високо оценявам 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.

15.05.2018 г.

Изготвил рецензията:

(проф. д.т.н. Иван Томов Димов)