Пловдивски университет „Паисий Хилендарски“
Факултет по математика и информатика
Катедра Приложна математика и моделиране

Мая Пламенова Стоименова

Моделиране на бързопроменливи времени редове

Автореферат

за присъждане на образователна 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

Пловдив

2018

Дисертационният труд е обсъден и насрочен за защита на разширен катедрен съвет на катедра „Приложна математика и моделиране” при Факултет по математика и информатика на Пловдивския университет „Паисий Хилендарски”, град Пловдив, проведен на 14.03.2018 г.

Дисертационният труд е с общ обем от 139 страници и включва четири глави, заключение и библиография, състояща се от 123 източника. Списъкът на авторските публикации включва 3 заглавия, невключени в библиографията. Дисертацията съдържа 43 фигури и 33 таблици.

Номерацията на формулите, цитиранията, примерите, таблиците и фигурите съвпада с тяхната номерация в дисертационния труд.

Защитата на дисертационния труд ще се състои на 15.06.2018 г. в Заседателната зала на Нова сграда на Пловдивския университет „Паисий Хилендарски”, Пловдив, бул. „България“ №236, етаж 2, на открито заседание на научното жури.

Материалите по защитата са на разположение на интересуващите се в библиотеката на ФМИ, Нова сграда на ПУ, всеки ден от 8:30 до 17:00 часа.

Научно жури

Председател: доц. д-р Дойчин Тодоров Бояджиев, ПУ „Паисий Хилендарски“

Членове:

Проф. д.т.н. Иван Томов Димов, ИИКТ на БАН (рецензент)

Проф. д-р Михаил Димов Тодоров, ТУ София (рецензент)

Доц. д.м.н. Иванка Миткова Желева, РУ “Ангел Кънчев“

Проф. д.м.н. Снежана Георгиева Гочева- Илиева, ПУ „Паисий Хилендарски“

Автор на дисертационния труд: Мая Пламенова Стоименова

Заглавие: Моделиране на бързопроменливи временни редове

Съдържание

Актуалност на изследванията

В съвременния свят все по-често се обръща внимание на така наболелия проблем със замърсяването на въздуха, което е сериозна заплаха за човешкото здраве. В световен мащаб може да се отбележи, че страните с най-голямо замърсяване са Китай, Индия, Индонезия и част от страните в Европа, в това число и България. Непрекъснато наблюдение и контрол за стойностите на вредните емисии във въздуха се извършва от Националната система за мониторинг на качеството на атмосферния въздух. Създадена е интерактивна карта [27], която следи качеството на въздуха в Европа и отразява нивата на замърсителите PM10, PM2.5, NO2, O3, SO2 във всеки конкретен момент. Целта е населението по целия свят да бъде информирано какъв въздух диша. Европейската комисия сътрудничи с държавите членки на ЕС и цели да им съдейства да спазват законоустановените норми за вредните емисии, които да бъдат гаранция за здравето на населението.

Много голям брой научни публикации са свързани с изследване, анализ и прогнозиране на замърсяването на атмосферния въздух с фини прахови частици PM10 и PM2.5. Ще добавим, че към момента има над 50 специализирани списания с импакт фактор, публикуващи статии с математическо моделиране в областта на замърсяването на въздуха. Общият брой статии в базите данни на SCOPUS, с ключови думи “PM10” AND “model” за последните 5 години са над 1050, а в базите на WoS – броят е 2400.

На база на литературния обзор и анализа на състоянието на проблемите, представени в Глава 1 следва, че **в научен и научно-приложен аспект моделирането на временните редове с данни за замърсяването на атмосферния въздух е силно актуална задача.**

В настоящия дисертационен труд се прилагат стохастични и най-съвременни дейта майнинг методи за построяване и анализиране на математически модели на проблемни въздушни замърсители в населени места в България. По-специално е приложен метода CART (класификационни и регресионни дървета), който до сега не е достатъчно добре застъпен в литературата в областта на екологията. Получените модели имат за цел решаване на реален проблем и се базират на реални данни.

Цел и задачи на дисертационния труд

Основен обект на изследване са данни от измервания на въздушни замърсители и подбор на подходящи методи за тяхното статистическо изследване.

Основна цел на дисертационния труд

Разработка на висококачествени статистически модели за бързопроменливи времени редове и приложението им за описание, анализ и краткосрочни прогнози на замърсители на атмосферния въздух.

Основни задачи на дисертационния труд

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 временният ред за PM10 в град Перник в зависимост от метеорологични данни 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.

Обзор на основните резултати на дисертационния труд

Глава 1. Въведение

Направен е подробен анализ на състоянието на изследванията в областта на статистическото моделиране на замърсителите на въздуха и е дадено кратко описание на използваните в дисертационния труд статистически методи. Дефинирани са целта и задачите на този труд.

- ***ARIMA метод***

ARIMA (авторегресия, интегрирана, с плаващи средни) е общ клас методи за моделиране и анализ на времеви редове, въведен от Бокс и Дженкинс през 1970 г. **[**46**].** Общият вид на едномерните модели се записва като ARIMA(p, d, q), където параметрите p, d, q са неотрицателни цели числа и изразяват както следва: p – авторегресионен процес (AR), d – тренд процес, тенденция (I) и q – процес на плаващи средни на остатъците (MA) [11]. ARIMA се прилага за краткосрочни прогнози, предвиждайки бъдещите стойности на реда, като може да се проследи дали тези стойности се увеличават или намаляват в реда. За да може да се приложи този метод, ще считаме, че броят на наблюденията е минимум 40. Изисква се също стойностите на изследвания временен ред да имат нормално или близко до нормалното разпределение. Друго необходимо условие за прилагането на този метод е да няма липсващи данни в извадката, а в случай че има такива, те се запълват например чрез линейна интерполация или друг метод.

Основни стъпки на ARIMA методите:

а) Идентификация и определяне на началните приблизителни стойности на параметрите p, d, q

За целта се провежда изследване на данните с изчисляване и начертаване на графиката на автокорелационните функции (ACF) и частични автокорелационни функции (PACF). Търсят се възможно най-малките стойности на параметрите p, d, q по методите на Бокс-Дженкинс. Когато стойността е 0, елементът не е необходим в този модел. Средният елемент d (тренд) се изследва първи. Целта е да се определи дали процесът е стационарен, а ако не е, да се преобразува към такъв. Стационарният процес има постоянна средна стойност и малка дисперсия през целия времеви период на изследване. Стойността на p е 0, ако няма връзка между 2 съседни наблюдения.

б) Построяване на модел и оценка на неговите коефициенти

На този етап се прилага алгоритъм, предложен от Бокс-Дженкинс.

в) Диагностика на модела

За целта се изследват остатъците. Това са разликите между наблюдаваните и предсказаните по модела стойности. Теоретично се приема, че остатъците трябва да са случайни и да имат нормално разпределение (бял шум).

г) Избор на модел

Могат да се използват много критерии, включително т.нар. информационни критерии. В статистиката критерият на Шварц, известен като BIC (Bayersian Information Criterion) често се използва за избор на модел. Моделът с най-малката стойност на BIC се счита за най-добър.

- ***Метод на класификационни и регресионни дървета CART***

Методът CART (Classification And Regression Trees) е предложен в монографията **[**13**]** през 1984 г. Днес той се използва много активно за класификации и изследване на зависимости в почти всички научни сфери и се счита като един от най-ефективните методи на извличане на знания от данни – дейта майнинг **[**67**].** В настоящия дисертационен труд се използва оригиналната версия на метода с двоични дървета от монографията [13].

Като регресионна техника CART методът се определя като рекурсивно-разделяща регресия. Целта е разделяне на случаите (наблюденията) с данни в относително хомогенни (с ниско стандартно отклонение или с минимална обща грешка по метода на най-малките квадрати) крайни възли и получаване на средна наблюдавана стойност при всеки краен възел като прогнозна стойност.

Целта е да се построят регресионни CART модели за установяване зависимостта на нивата на замърсяване с PM10 от осемте метеорологични променливи. Ограниченията за минимум случаи в родителски възел (m_1) и минимум в наследствен възел (m_2) след провеждане на голям брой предварителни анализи задаваме в два варианта – 20 и 10 за m_1, и 10 и 5 за m_2, съответно. Въвеждаме означение за получените модели за PM10 с $M(m_1, m_2)$.

При построяването на V – разделна кросвалидация, извадката се разделя на случаен принцип на V равни под-извадки. Всяка една от под-извадките се използва за тестване на модела, а останалото множество от данни се използва като обучителна извадка, като процесът се повтаря V пъти. Предимството на този метод е, че всички наблюдения използвани за обучение и тестване на извадката, се използват само веднъж.

Оценката за качеството на моделите ще извършваме с помощта на средната квадратична грешка (RMSE), средната абсолютна процентна грешка (MAPE) и средната абсолютна грешка (MAE) и по формулите:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{1.5}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{1.6}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{1.7}$$

където, $\hat{y}_t$ е предсказаната от модела стойност във всеки момент t на времевия ред.

Глава 2. Стохастично и CART моделиране на времеви редове на проблемни замърсявания с фини прахови частици на въздуха на град Перник

В тази глава задачата е да се създадат, анализират и сравнят:

А) Стохастични авторегресионни модели на PM10, зависещи само от времето;

Б) CART модели, с използване на метеорологични и други променливи като предиктори.

2.1. Изследвана област

Настоящето изследване се фокусира върху моделирането на данни за проблемния замърсител PM10 във въздуха на град Перник. Търсят се адекватни оптимални модели, при предварително зададени критерии. Моделите се прилагат за прогнозиране на замърсяванията в рамките на няколко дни.

2.2. Използвани данни и начална статистическа обработка

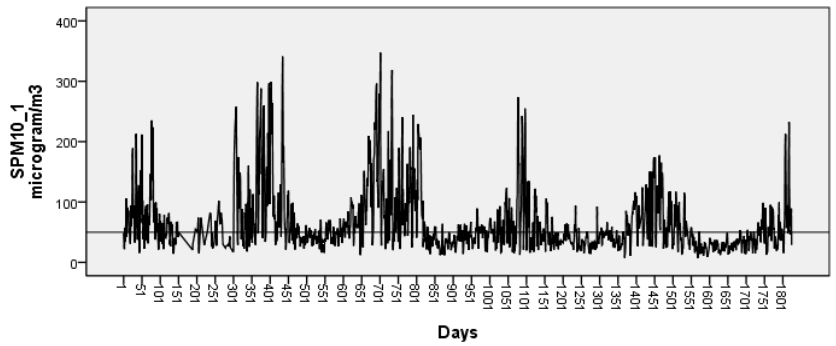
Използваните данни са за период от 5 години – от 1 януари 2010 г. до 31 декември 2014 г., на база среднодневни наблюдения.

Резултатите от първоначалната обработка на данните са представени в Таблица 2.1. Броят на наблюденията е N=1826.

Таблица 2.1 Описателна статистика за началните данни за концентрациите на PM10 (променлива SPM10_1) на град Перник

Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Minimum	Maximum
63.18	47.700	50.273	2527.36	2.272	0.057	5.895	0.114	7	348

Максималната стойност от 348 $\mu g / m^3$ на PM10 превишава почти 7 пъти среднодневната норма от 50 $\mu g / m^3$. Такъв тип превишения не са единични. На Фигура 2.1 е показан характерът на концентрациите в течение на времето.



Фигура 2.1 Графика на измерените среднодневни данни на концентрациите на PM10 за град Перник. Хоризонталната линия показва допустимата средноденощна горна граница от 50 $\mu g / m^3$

2.3. Построяване на ARIMA модели и анализ

За построяването на параметрични модели прилагаме метода ARIMA, т.е. търсим модели от вида ARIMA(p,d,q) [11]. За определяне на най-адекватен модел със съответни стойности на параметрите (p,d,q) бяха построени редица модели, при които се тръгва от по-малки стойности на параметрите към по-големи. При близки резултати от моделите следваме правилото на по-простия модел [11]. Проведените изследвания показват, че очакваните приблизителни стойности за p са: 1<p<5, че няма тренд т.е. d=0 и очакваните стойности за q са: 1<q<7.

Получените най- добри модели и техните основни статистики са показани на Таблица 2.3. За модела (1,0,5) R^2=0.564, т.е. полученият ARIMA описва около 56% от данните, за модела (5,0,7) имаме R^2=0.566 и (3,0,7) има R^2= 0,566. За ARIMA (1,0,5), авторегресионният компонент е (p=1), т.е. най-силното влияние върху нивото на замърсяване е това на стойностт от предишния ден. Компонентът на плаващите средни (q=5) е индикатор, че локалните стохастични изменения са зависими с 5 предишни стохастични члена в рамките на времевия ред. От последния ред на Таблица 2.3 Sig=0.568; 0.596 и 0.284 , т.е. са незначими, което се изисква. Към RMSE в скоби са изчислени стойностите на грешките в микрограма на кубичен метър.

От получените статистически индекси в Таблица 2.3 можем да заключим, че и трите модела дават приблизително еднакво приближение, но избираме най-простия от тях модел АРИМА(1,0,5), с който да работим.

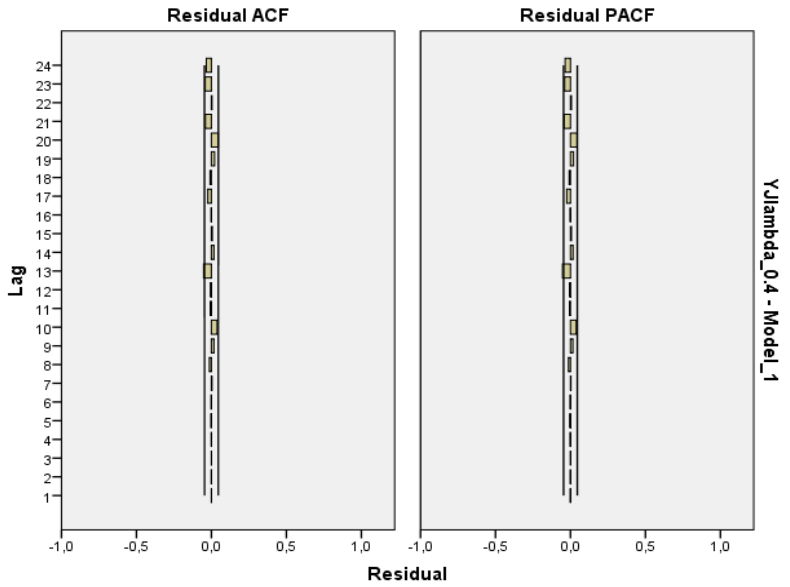
Таблица 2.3 Статистики на избраните ARIMA моделите на замърсителя PM10 за град Перник

		ARIMA (1,0,5)	ARIMA (5,0,7)	ARIMA (3,0,7)
Model Fit statistics	**Stationary R-squared**	0.564	0.566	0.566
	R-squared	0.564	0.566	0.566
	RMSE	0.086 (32.762)	0.085 (32.926)	0.085 (32.756)

	MAPE	3.253	3.247	3.240
	MAE	0.063	0.062	0.062
	Normalized BIC	-4.889	-4.866	-4.875
Ljung-Box	**Statistics**	10.543	4.597	9.743
	DF	12	6	8
	Sig.	0.568	0.596	0.284

2.4. Диагностика на ARIMA моделите

Проведен е обстоен анализ за диагностики на грешките на получените стохастични модели. Като например на Фигура 2.4 са изобразени графиките на ACF и PACF на грешките (остатъците – разликите между стойностите, предсказани от модела и измерените данни) на избрания модел (1,0,5). При 5% доверителен интервал, грешки са пренебрежимо малки, което ни дава възможност да ги приемем за „бял шум“, а моделът – за достатъчно точен.

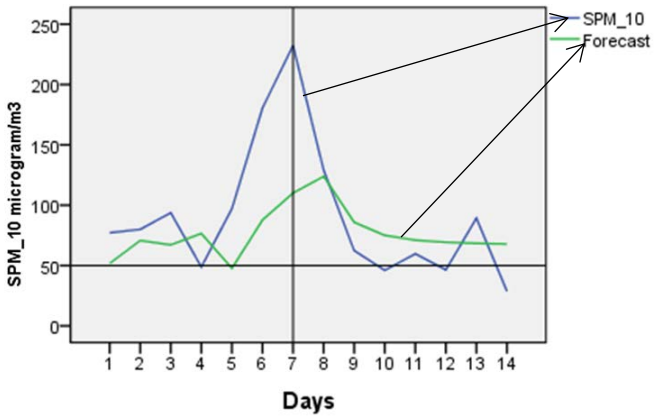


Фигура 2.4 ACF и PACF на остатъците за модел АРИМА(1,0,5)

2.5. Приложение на моделите за предсказване на измерените и прогнозиране на бъдещи концентрации на PM10

Силата на ARIMA моделите се крие в добрите резултати, постигнати при предсказване на бъдещи събития. В нашия случай на Фигура 2.10 е представено приложението на избрания модел ARIMA (1,0,5) за краткосрочно прогнозиране в рамките на 7 дни – от 25 декември до 31 декември 2014 г. За целта са използвани реални допълнителни данни, които не са включени при построяването на модела и могат да бъдат сравнени с прогнозите, получени с помощта на избрания модел. Както се вижда има много добро съответствие с наблюдаваните стойности.

От Фигура 2.10 се вижда, че в дадения интервал много добре са предсказани и прогнозирани превишенията спрямо официалния лимит от 50 $\mu g / m^3$.



Фигура 2.10 Прогнозни стойности на ФПЧ10 с ARIMA(1,0,5) за 7 дни напред, сравнени с реалните измерени стойности. Хоризонталната линия показва допустимата горна граница от 50 $\mu g / m^3$, а вертикалната линия е между последните използвани 7 дни (от лявата страна) и предсказаните 7 дни (от дясната страна на линията)

2.6. Постановка на задачата за изследване на PM10 в зависимост от метеорологичните данни с CART метод

В това изследване CART методът е проведен със същите данни, както в случая на ARIMA моделите. За приложението на анализа са използвани измерените стойности на 8 метеорологични променливи – минимална и максимална дневна температура (^{o}C), скорост (m/s) и посока (radians) на вятъра, валежи (%), влажност на въздуха (%), атмосферно налягане (*mb*), облачно покритие (%) и влиянието на три отровни газа – въглероден оксид ($\mu g / m^3$), серен диоксид ($\mu g / m^3$) и азотен диоксид ($\mu g / m^3$), прекурсори на PM10. Като предиктори са използвани също и лагирани променливи със стойностите от предишни дни, отбелязани в скоби от вида < >, например PM10<1>, PM10<2> и т.н.

В Таблица 2.11 са показани получените резултати на максималните модели от прилагането на CART метода, със съответните стойности за коефициент на детерминация, броя на възлите във всеки един модел и тяхната грешка.

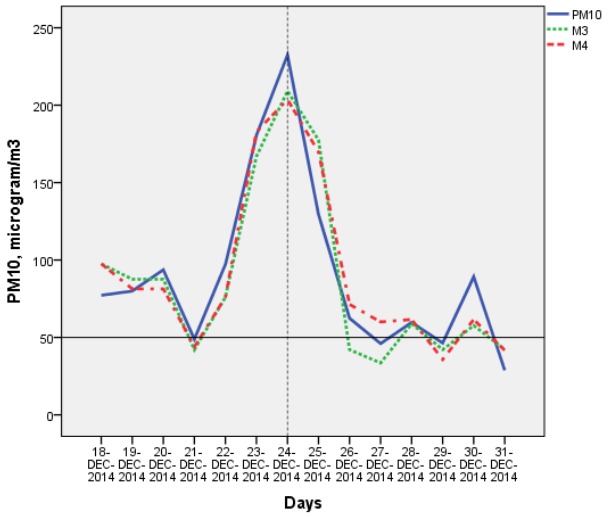
Всички четири модела са добри и дават много добро приближение към началните данни. Разглеждайки и четирите модела, се забелязва, че M1 е с най- малка средна квадратична грешка (RMSE). Това показва, че при отчитането на замърсяването, всички метеорологични условия и отровни газове оказват влияние и така получения модел описва данните в най-висока степен. Вижда се, че всички модели отчитат сравнително близки резултати.

Таблица 2.11 Обобщение на получените максимални CART модели

	Модел	(m1,m2)	Предиктори	R^2 Learn	Брой крайни възли	Relative Error	RMSE
PM10	**M1**	(10,5) 20fold	Метео данни + CO, NO2,SO2,PM10<1>, PM10<2>, CO<1>	0.937	267	0.237	12.850
	M2	(10,5) 10fold	min_temp, max_temp, CO,NO2,SO2, PM10<1>, PM10<2>	0.914	144	0.263	15.033
	M3	(10,5) 20fold	CO,NO2,SO2, PM10<1>, CO<1>	0.915	265	0.261	14.908

M4	(10,5) 10fold	CO,NO2,SO2, PM10<1>, PM10<2>	0.904	152	0.257	15.809

За да се предскаже какво ще бъде замърсяването в следващите 7 дни, са използвани данни от реални измервания за периода от 25.12.2014 до 31.12.2014 г., които не се включват в анализа, но се използват за сравнение с получените стойности, предсказани от модела. На Фигура 2.12 са илюстрирани отрязък на последните 7 дни (от лявата страна на вертикалната линия) от изследвания период и прогнозираните стойности за следващите 7 дни (вдясно на вертикалната линия) за моделите M3 и M4. Хоризонталната линия показва пределно допустимия праг за замърсяване с фини прахови частици от 50 $\mu g / m^3$. Резултатите показват много добро съвпадение на прогнозите с изходните данни, което означава, че методът е подходящ за прилагане на краткосрочни прогнози за следващи замърсявания.



Фигура 2.12 Графика на предсказване за 7 дни напред с моделите M3 и M4

2.7. Сравнение на CART моделите с едномерни ARIMA модели

В този параграф се прави сравнение на CART модели, построени само с отчитане на влиянието на замърсяването от предходните два дни с едномерни ARIMA модели.

В Таблица 2.16 са показани два модела от приложението на CART метода, като модел C_M1 е изграден с отчитане влиянието на замърсяването от предишните два дни, а C_M2 зависи само от концентрациите на PM10 от вчерашния ден. Резултатите показват, че първия модел е по-добър, R^2 е по- високо, а грешките на модела са по- малки, което означава, че е важно, какво е било замърсяването в последните два дни.

Таблица 2.2 Обобщение на получените оптимални CART модели

	Модел	**(m1,m2)**	**R^2 Learn**	**Брой крайни възли**	**Relative Error**	**RMSE**
PM10	C_M1	(10,5) 10fold	0.660	16	0.447	29.797
	C_M2	(10,5) 10fold	0.614	14	0.465	31.782

За да се сравнят двата метода ARIMA и CART е необходимо да се сравнят стойностите на средно квадратичната грешка RMSE и коефициента на детерминация т.е. до каква степен модела се доближава до реалните данни. От Таблица 2.3 се вижда, че за най – добрия ARIMA модел (1,0,5), R^2 е 0.564, което означава приближение 56.4% до изходните данни, а RMSE е 32.762. В Таблица 2.16 от резултатите за CART моделите се забелязва, че

най-добрият CART модел C_M1 дава 66% описание на действителните данни, а грешката на модела е 29.797.

От така получените резултати може да заключим, че ARIMA моделите дават предимство на класификационната и регресионна техника. Ясно се вижда, че CART моделите се доближават повече до актуалните данни и получените средноквадратични грешки на моделите са по-малки, което е изискване за по-добър модел.

Изводи към Глава 2

Глава 2 представя резултатите от статистическото изследване на замърсителя PM10 на въздуха на град Перник. За обработка на временния ред на PM10 са построени и изследвани едномерни ARIMA модели. Чрез избрани критерии, описани в Глава 1, 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 описва данните с коефициент на детерминация R2=56.4% , т.е. описва 56% от наблюденията. 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 показва много добри прогнозиращи качества.

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, NO2, SO2 и стойността на CO от предишния ден. По същите критерии е подбран най-добър CART модел с R2 =93.7. Моделът е изследван от гледна точка на неговата точност и RMSE= 12.851. Избраният най-добър CART модел отчита, че най-съществено влияние за замърсяването с PM10 за текущия ден имат следните променливи: на първо място концентрацията на CO, следвана от стойността на PM10, измерена от вчерашния ден, а трети и четвърти по ред са концентрациите съответно на NO2 и SO2. Моделът е приложен за предсказване на PM10 за следващите 7 дни и показва отлични прогнозни резултати.

Глава 3. Стохастичен едномерен и многомерен анализ на времеви редове на замърсителите на въздуха PM2.5 и PM10: сравнителен анализ за Пловдив и Асеновград

Настоящата глава разглежда средните дневни данни за замърсяването на въздуха с PM2.5 и PM10, в градовете Пловдив и Асеновград между 2011 и 2015 г. Целта е да се намерят и анализират основните взаимозависимости в данните, математическите модели и да се разработят краткосрочни прогнози. Общият брой изследвани данни е N=1826.

3.1. Изследвана област

Изследваме среднодневните данни за PM2.5 и PM10, събрани от три мониторингови станции, от които две станции в Пловдив и една станция в Асеновград за 5 години в периода от 2011 до 2015 г. Станциите са разположени съответно в Пловдивски райони: Баня Старина (PM2.5, PM10), Каменица (PM10), а в Асеновград: гара Кметство – Долни Воден, (PM10).

3.2. Използвани данни и начална статистическа обработка

Статистическият анализ на данните показва превишения на максимално допустимите норми [6], [23], [4], както за PM2.5, чиято норма е 25 $\mu g/m^3$, така и за замърсителя PM10 и в трите автоматично-измервателни станции. Достигнат е максимум за PM2.5 в измервателна станция Баня Старинна от 334.6 $\mu g/m^3$, което е почти 14 пъти над допустимата норма. За

PM10 и в трите мониторингови станции са отчетени също много високи максимални стойности, които са около 7-8 пъти над максимално допустимата норма.

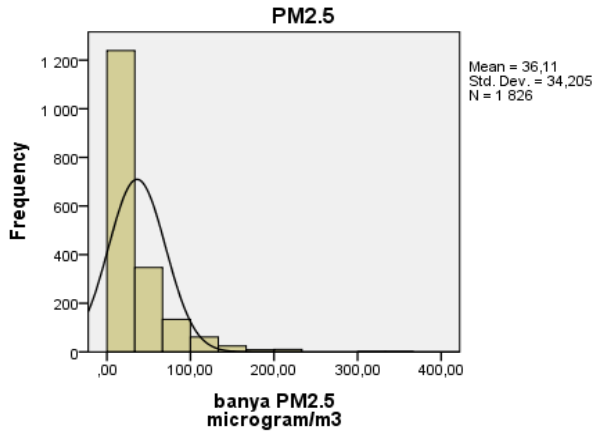
3.3. Трансформация на данни

За да се прилагат параметрични модели е необходимо разпределението на извадката да бъде нормално. В нашия случай изходните данни нямат нормално разпределение.За да се подобри нормалността и да се стабилизира разсейването на данните, използваме формулата на Йео – Джонсън трансформация (2.1) [113]:

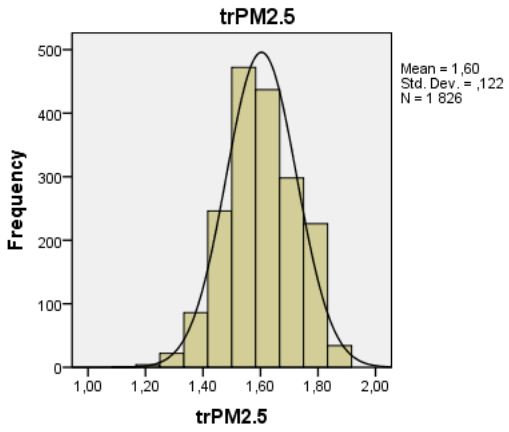
$$trx = \psi_{YJ}(\lambda, x) = \begin{cases} \left\{(x+1)^{\lambda} - 1\right\} / \lambda & x \geq 0,\ \lambda \neq 0 \\ \log(x+1) & x \geq 0,\ \lambda = 0 \\ -\left\{(-x+1)^{2-\lambda} - 1\right\} / (2-\lambda) & x < 0,\ \lambda \neq 2 \\ -\log(-x+1) & x < 0,\ \lambda = 2 \end{cases}, \quad \lambda \in [-2, 2] \qquad (2.1)$$

където x е първоначалната променлива, trx е трансформираната променлива, а λ е неизвестен параметър. За нашите данни, коефициентите за Йео – Джонсън трансформацията за всички наблюдавани променливи бяха определени с помощта на проста процедура чрез опити от редицата числа {-2, -1.9, …, 2} и K-S (Колмогоров-Смирнов) теста за нормалност. При подбор на оптималната стойност на λ сравняваме оптималната стойност на показателя на Колмогоров-Смирнов за нормално разпределение на данните.

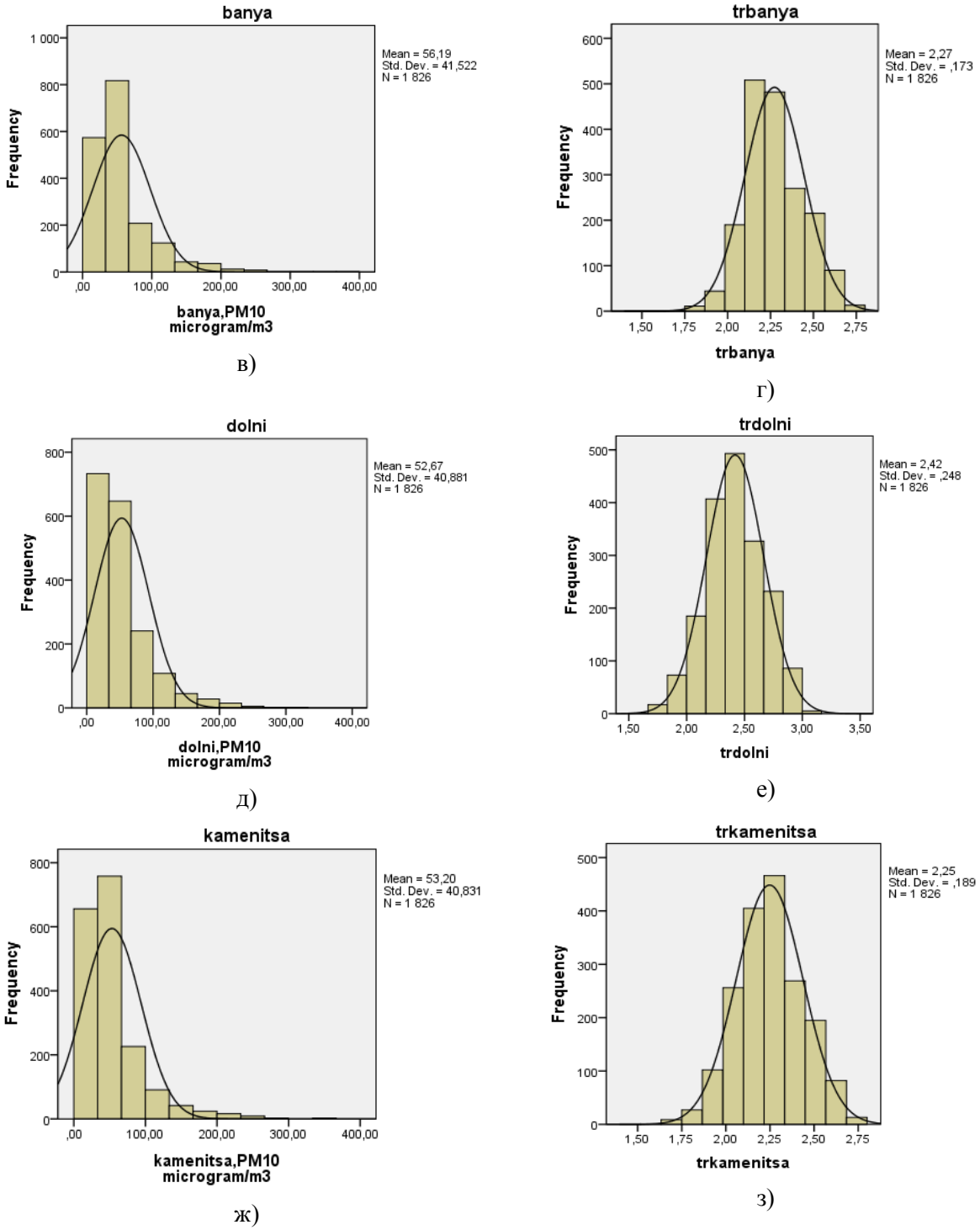
На Фигура 3.3 са изобразени разпределенията на времевите редове, за изследваните данни. От лявата част на фигурата са посочени хистограмите на началните данни, а в дясната част са разпределенията след трансформация на данните. Данните и за четирите станции след трансформацията имат нормално разпределение.



а)



б)



Фигура 3.3 Графики на честота на първоначалните данни и след трансформацията.

3.4. Резултати от моделирането с едномерни модели

За трансформираните променливи бяха извършени нормалност на разпределението и неговата слаба стационарност, автокорелационни тестове и тест за наличие на тренд (Дики- Фулър и др.) и беше проверена и обратимост на временния ред [38]. Установено е, че всички трансформирани редове са нормално разпределени, слабо стационарни (или нямат тенденции d, di) и нямат сезонен компонент. Трябва да се отбележи, че е направен и тест за съществуването на годишна сезонност s = 365, което доведе до отрицателен резултат. Използвайки функциите за автокорелация (ACF) и частичните функции за автокорелация

(PACF), бяха разгледани и зададени следните граници на параметрите: $1 \le p \le 3, 1 \le q \le 6$ за всички времеви редове.

3.5. Диагностика на едномерни модели

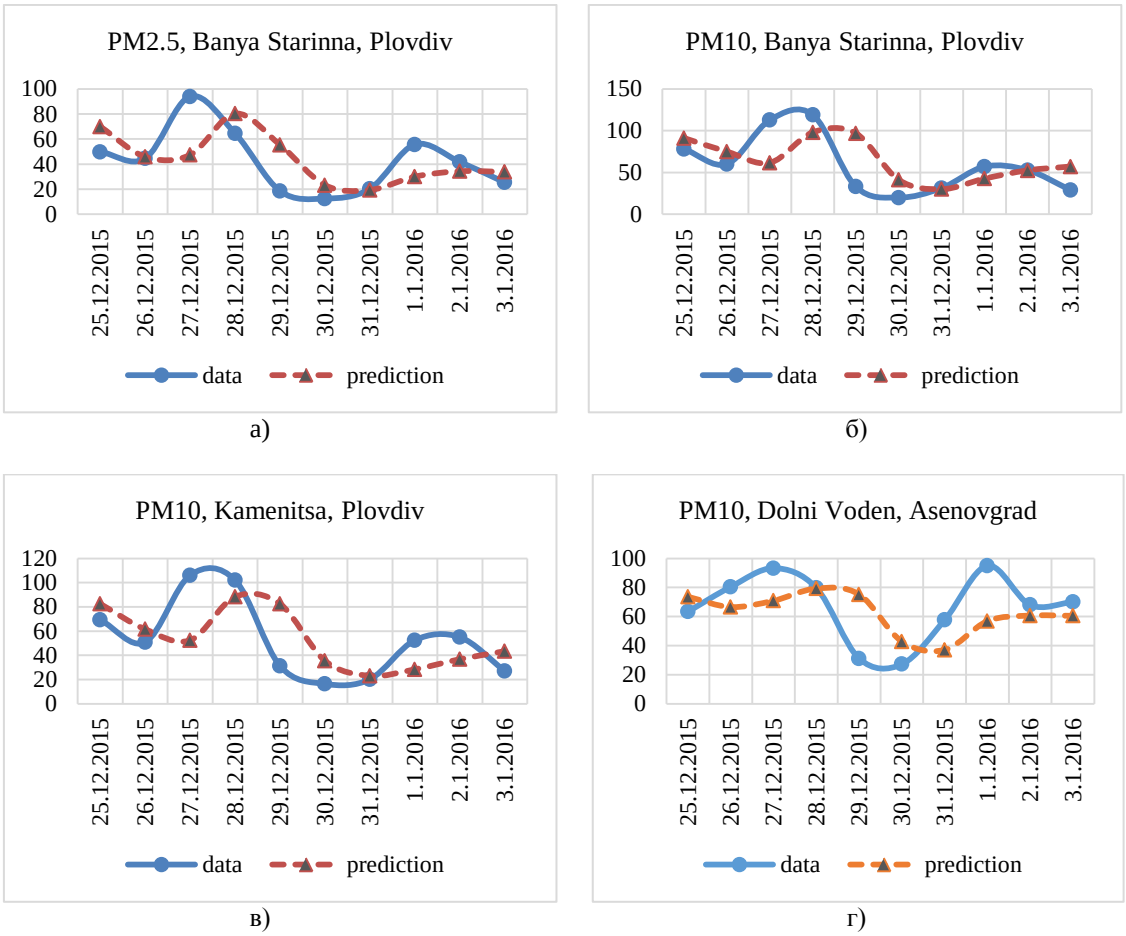
Основен показател за качество на избраните модели са коефициентът на детерминация R^2 и грешките на моделите. Получените стойности на R^2 са в около 60-66%, т.е. степента на съвпадение на моделите с измерените стойности на замърсителите. Всички Люнг- Бокс статистики са незначими (Sig>0.05), което потвърждава статистическия авторегресионен характер на изследваните редове. Получените RMSE са приемливо малки, в рамките на 5% от максималната стойност от Таблица 3.5.

Таблица 3.5 Получени ARIMA модели за трасформираните редове със статистика за данните

Variable	ARIMA Model	R^2	Ljung-Box Statistics	RMSE	BIC
trpm25	(1,0,5)	0.662	0.467	0.071	-5.253
trbanya	(1,0,4)	0.589	0.105	0.111	-4.371
trkamenitsa	(1,0,5)	0.589	0.201	0.122	-4.183
trdolni	(1,0,4)	0.616	0.299	0.185	-3.347

3.6. Приложение на едномерните модели за краткосрочни прогнози

Използвайки моделите с данните до 31 декември 2015 г., прогнозите бяха направени за концентрациите на замърсителите през следващите 3 дни (1-3 януари 2016 г.). Резултатите от тези прогнози за всички едномерни модели са показани на Фигура 3.7. Вижда се, че моделите отразяват много добре характера на промените в динамичните редове с течение на времето.



Фигура 3.7 Сравнение между измерените и предсказаните стойности на концентрациите на замърсителя от 24 до 31 декември 2015 и прогнозите за 3 дни за 1-3 януари 2016 г.

3.7. Резултати от моделирането с многомерни модели

Някои от получените многомерни модела на ARIMA са дадени в Таблица 3.7. Прогнозите за бъдещо замърсяване са много сходни с тези от едностранните модели. Поради тази причина ние не ги привеждаме тук.

Таблица 3.7 Избрани многомерни ARIMA модели за трансформирани времеви редове

Променливи	**ARIMA модели**
2D: trkamenitsa, trbanya (PM10)	(1,0,4)
2D: trbanya, trdolni (PM10)	(1,0,5)
3D: trkamenitsa, trbanya, trdolni (PM10)	(1,0,3)
4D: trkamenitsa, trpm25, trbanya, trdolni	(1,0,0)

Изводи към Глава 3

В тази Глава се изследват данни за замърсяване на атмосферния въздух с PM10 и PM2.5 за градовете Пловдив и Асеновград, измерени в три станции. Не са използвани данни от други временни редове. Извършена е предварителна трансформация на данните.

За период от 5 години са получени и анализирани едномерни и многомерни ARIMA модели. Всички конструирани модели показват много добри статистически качества, като адекватност, висока способност за предсказване и краткосрочно прогнозиране. Едномерните модели са от тип (1,0,4) за Баня Старинна и Долни Воден и получените R^2 и RMSE за данните от двете измервателни станции са съответно R^2=58.9%, RMSE =0.111 и R^2=61.6%, RMSE =0.185. Получените модели за PM2.5 и данните от Каменица са от вида (1,0,5) и коефициенти са R^2=66.5% и RMSE =0.071 за PM2.5, а за Каменица са R^2=58.9% и RMSE =0.122.

Установи се тясна зависимост в поведението на времевите редове, измерени от трите станции. Получени са много сходни по тип и коефициенти едномерни модели. Това позволява да се заключи, че измереното замърсяване на въздуха с фини прахови частици е почти еднакво в района на двата града. Може да се очаква, че това включва и зоната между градовете.

Построени са и многомерни (векторни) модели ARIMA тип 2D, 3D и 4D, които описват замърсяването чрез система взаимни уравнения. Моделите са приложени за прогнозиране с три дни напред и показват висока степен на съвпадение с реалните измервания.

Сравнителният анализ на времевите редове показва, че няма тенденции, т.е. тенденцията за намаляване на замърсяването на въздуха и качеството на въздуха не се е подобрило в района на Пловдив и Асеновград през последните 5 години. Това предполага, че източниците са общи и постоянни; сред които индустриалните фабрики, освобождаването на големи и малки количества прахови частици в атмосферата, сравнително интензивен автомобилен трафик, както и използването на твърди горива за битово отопление през зимните месеци. Географското положение на градовете в ниско разположената Тракийска долина и континенталният климат също допринасят за задържането и слабото разсейване на вредните емисии на PM2.5 и PM10 във въздуха.

Моделирането показва, че подходът, използван в настоящето изследване, може да се прилага, както за анализиране на минали периоди и за установяване на тенденции, така и за краткосрочно прогнозиране на нивата на замърсители, служеща като независима алтернатива на официалните методи за мониторинг и контрол на качеството на въздуха, предоставени от Националната агенция по околна среда.

Глава 4. Приложение на CART метод за моделиране и предсказване на замърсяванията с PM10 на град Плевен

Целта в тази глава е да се създадат висококачествени математически модели за ефективно предсказване и прогнозиране на нивото на замърсяване с PM10, чрез различни подходи:

1. Построяване и анализ на CART модели без кросвалидация и без авторегресионни предиктори
2. Построяване и анализ на CART модели с кросвалидация и лагирани променливи
3. Изследване степента на влияние на метеорологичните редове
4. Анализ на точността
5. Сравнение на моделите
6. Приложение на моделите за:
   - Предсказване на концентрациите на PM10
   - Проверка на точността на предсказаните стойности с помощта на контингентни таблици
   - Прогнозиране на бъдещи замърсявания

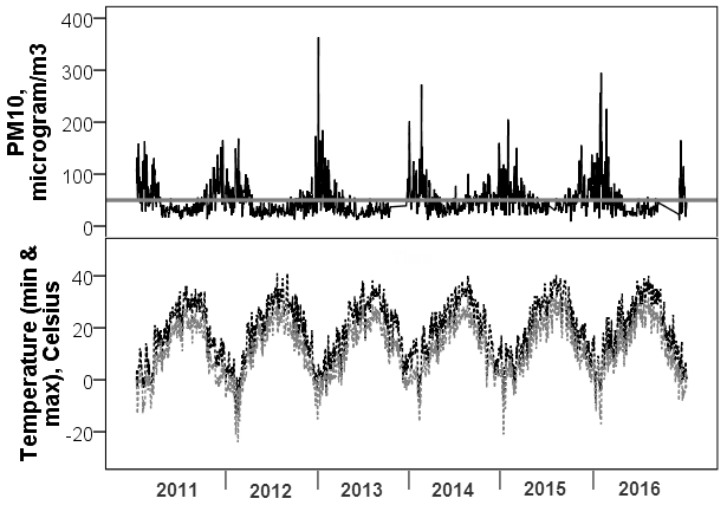
4.1. Изследвана област

Предмет на изследване в тази глава е замърсяването с PM10 на гр. Плевен.

4.2. Използвани данни и начална статистическа обработка

Проведеният в това изследване анализ се базира на измерените среднодневни концентрации на въздушния замърсител PM10 в град Плевен, България в период от 6 години, от 1 януари 2011 до 31 декември 2016 г. За провеждане на моделирането и прогнозите са използвани също измерените стойности на 8 метеорологични променливи- минимална и максимална дневна температура (^{o}C), скорост (m/s) и посока (radians) на вятъра, валежи (%), влажност на въздуха (%), атмосферно налягане (*mb*), облачно покритие (%). Броят на наблюдаваните стойности е N=2190.

На Фигура 4.1 са представени графики на данните. В горната част на фигурата са илюстрирани среднодневните измерени концентрации на PM10, а в долната част – съответните максимална и минимална дневни температури. За PM10 ясно се виждат многократни превишавания на допустимия среднодневен лимит от 50 $\mu g/m^3$ (означен с хоризонтална плътна линия), предимно през зимните периоди. Като цяло се наблюдава съответствие на тези пикове с най-ниските стойности на минималната и максимална дневни температури през зимните месеци.



Фигура 4.1 Графика на първоначалните данни – дневни концентрации на PM10, максимални и минимални дневни температури

За шест-годишния период средната стойност е 49 $\mu g/m^3$. По години този показател от 2011 г. насам взима средни стойности съответно 52.3, 45.4, 41.7, 51.1, 53.9, 48.4 $\mu g/m^3$. Това е системно надвишаване на допустимата средногодишна норма от 40 $\mu g/m^3$.

4.3. Построяване, анализ и приложение на CART модели на PM10 и tr_PM10

4.3.1 CART модели без кросвалидация и без лагирани променливи

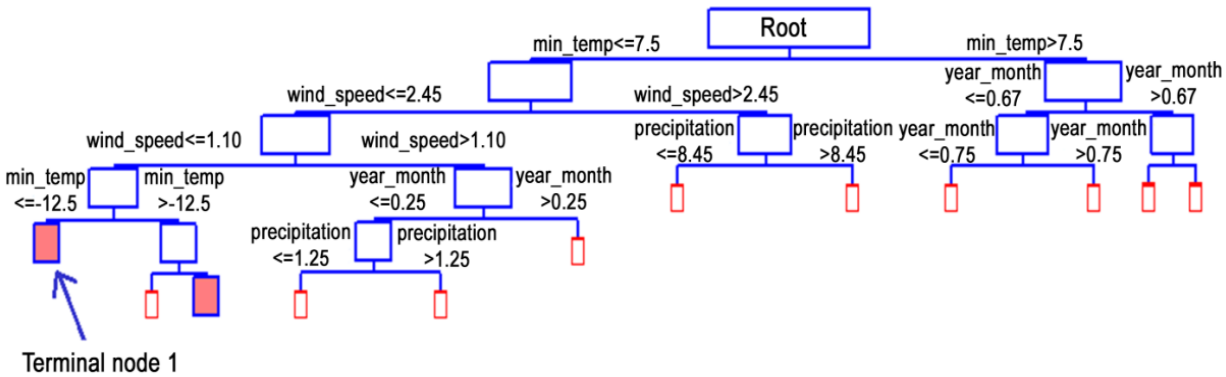
При избора на най-добър модел се ръководим от достигане на най-малката относителна грешка на отклонение на моделните от измерените данни, съгласно [92], [17]. Освен това, тъй като моделът е регресионен, вземаме под внимание и моделите с най-голяма стойност на коефициента на детерминация R^2.

Избрани бяха 5 оптимални модела, отговарящи на поставените условия, като модел M5 е конструиран без отчитане на влиянието на посоката на вятъра и облачността. Основните характеристики на избраните модели са дадени в Таблица 4.3. Общото сравнение показва, че най-добрият модел от избраните е tr_M4, обясняващ до 78% от измерените данни с относителна CART грешка 0.222.

Таблица 4.3 Обобщение на оптималните CART модели за PM10 и tr_PM10

Зависима променлива	**Модел**	**(m_1, m_2)**	**Брой на терминалните възли**	**R^2 Learn**	**Relative CART Error**	**RMSE**
PM10	M1	(20,10)	169	0.609	0.391	18.362
	M2	(10,5)	349	0.727	0.273	15.323
tr_PM10	tr_M3	(20,10)	173	0.677	0.323	18.523
	tr_M4	(10,5)	354	0.778	0.222	15.641
PM10	M5	(10,5)	352	0.718	0.282	15.596

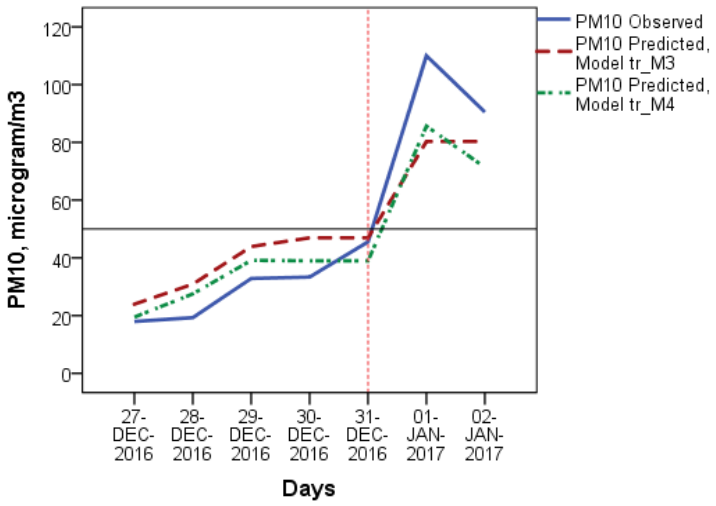
Общата топологична структура на полученото регресионно дърво с модела tr_M4 с 354 крайни възли е показана на Фигура 4.3. с правилата за изграждане на CART дърво. Максималните концентрации на PM10 са класифицирани в терминален възел 1 в лявата част на фигурата. Като се започне от корена на дървото, правилата за класификация на този възел са следните: min_temp <= 7.5, wind_speed <= 2.45, wind_speed <= 1.1, min_temp <= -12.5. По този начин получаваме, че най-високите превишения на PM10 се получават при сравнително ниска скорост на вятъра под 1.1 m/s и минимална температура под -12.5 C^o.



Фигура 4.3 Три нива от топологията на двоичното регресионно CART дърво на модела tr_M4 с 354 крайни възли и 11 предиктора за трансформираната променлива tr_ PM10

Използвайки моделите с данни до 31.12.2016 г., бяха прогнозирани концентрациите на PM10 за 2 дни (1-ви и 2-ри януари 2017 г.). Фигура 4.10 илюстрира прогнозните стойности, за два CART модела - tr_M3 и tr_M4 през последните 5 дни – от 27.12.2016 до 31.12.2016 г.

(от лявата страна на вертикалната линия), а прогнозите за следващите 2 дни – 1-ви и 2-ри януари 2017 г. (от дясната страна на вертикалната линия).



Фигура 4.10 Сравнение на измерените и предсказаните резултати за концентрациите на PM10

4.3.2 CART модели с кросвалидация и лагирани променливи

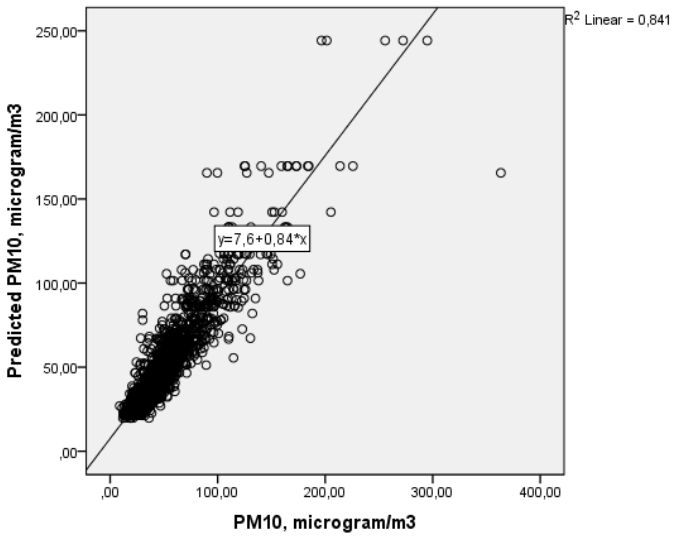
В раздел към променливите по-горе изброените променливи се добавят и две лагирани променливи на PM10, означени с PM10<1>, PM10<2>, и означаващи съответно измерените концентрации от предишния ден и замърсяването от преди два дни. Добавени са още две лагирани променливи- минималната температура и скорост на вятъра от предишния ден, както са означени съответно с min_temp<1>, wind_speed<2> . Резултатите описани в този параграф са получени след прилагането на V = 10% кръстосана валидация.

Получени бяха 6 оптимални модела, които отговарят на поставените изисквания за най–добър модел. Резултатите от изпълнението на CART метода, са посочени в Таблица 4.7. Общото сравнение от Таблица 4.7 показва, че най – добрият модел, описващ данните е CV_M6, като той обяснява над 84% от данните с най – малки относителна CART грешка 0.420 и RMSE 11.694.

Таблица 4.7 Обобщение на оптималните CART с кросвалидация и лагирани променливи модели за PM10 и tr_PM10

Зависима променлива	**Модел**	**(m_1, m_2)**	**Брой терминални възли**	**R^2 Learn**	**R^2 Test**	**Relative Error**	**RMSE**
PM10	CV_M1	(20,10)	92	0.614	0.389	0.631	18.245
	CV_M2	(10,5)	9	0.405	0.347	0.654	22.630
tr_PM10	CV_tr_M3	(20,10)	175	0.681	-	0.559	18.658
	CV_tr_M4	(10,5)	93	0.661	-	0.566	19.271
PM10	CV_M5	(10,5)	70	0.791	-	0.419	13.422
	CV_M6	(10,5)	218	0.841	-	0.420	11.694

Построяването на моделите се прилага за предсказване на стойностите на замърсителя за град Плевен, за сравнение на предсказаните и измерените концентрации на PM10, изследване на точността на моделаи краткосрочни бъдещи прогнози. Фигура 4.19 показва сравнение между предсказаните и измерените стойности на PM10 по метода на линейната регресия, където се вижда, че коефициента на детерминация е 0.841.



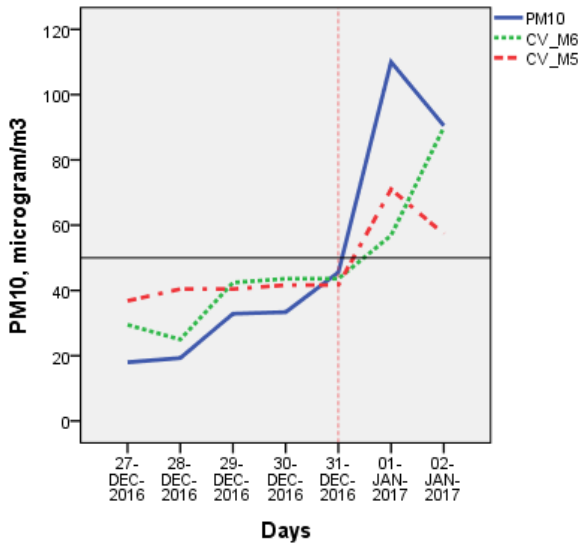
Фигура 4.19 Сравнение на предсказаните и измерените концентрации на PM10 с линейна регресия

Проверка за точността на модела, избран за най – добър, е направена с помощта на контингентна таблица. В Таблица 4.10 са показани резултатите от направената проверка. Избраният модел CV_M6 е познал 1389 или 93% от данните, които са под допустимия праг 50 $\mu g/m^3$ и 86% от концентрациите, които са над среднодневната граница. Общият дял на правилно предсказаните стойности на замърсителя PM10 е 91%. Неправилно прогнозирани под среднодневния лимит са 97, а над 50 $\mu g/m^3$ са 96. От получените резултати след направената проверка с контингентната таблица се установи, че избраният модел проявява много добра точност при предсказването и прогнозирането на концентрациите на PM10.

Таблица 4.10 Контингентна таблица за модел CV_M6 с PM10

		Forecast			
		<50	**>=50**	**Total**	**%O**
Obs	**<50**	1389	97	1486	93
	>=50	96	608	704	86
	Total	1485	705	2190	91
	%P	94	86		

За да се направят бъдещи прогнози за концентрациите на PM10, са използвани данни до 31.12.2016 г. Прогнозите са за два дни напред 1-ви и 2-ри януари 2017 г. Получените от модела прогнозни стойности са сравнени с действително измерените. На Фигура 4.20 е представен отрязък от предсказването с двата модела, отговарящи на изискванията за най-добър модел.

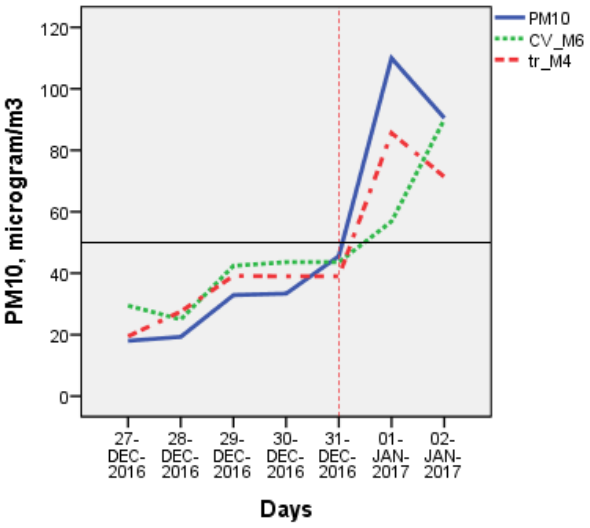


Фигура 4.20 Сравнение на предсказаните и измерените концентрации на PM10

4.3.3 Сравнение на резултатите получени с CART моделите

Сравнението на двата най – добри модела tr_M4 и CV_M6 се извършва на база: най – малка средна квадратична грешка и най – висок коефициент на детерминация. Моделът tr_M4 има коефициент на детерминация 0.778 или моделът описва около 78% от данните със средна квадратична грешка 15.641 и брой крайни възли 354. Моделът CV_M6 е с приближение към реалните данни с приблизително 84%, с грешка RMSE 11.694, а крайните възли са 218. Общото сравнение показва, че модел CV_M6 е по – добър, грешката от конструирането на модела е по – малка, приближението към изходните данни е по – високо и полученото дърво е сравнително по – просто от този на tr_M4.

На Фигура 4.21 графично са представени и сравнени резултатите за двата модела с изходните данни за замърсяването.



Фигура 4.21 Сравнение на предсказаните и измерените концентрации на PM10 с двата най–добри модела tr_M4 и CV_M6

Изводи към Глава 4

В тази Глава се изследват данни от измервания на среднодневни концентрации на PM10, проблемен замърсител на атмосферния въздух в град Плевен. Изследваните данни са

за период от 6 години, между 2011 и 2016. Моделирането е проведено с CART метод. Получени, анализирани и сравнени са 5 оптимални CART модела без кросвалидация и 6 оптимални модела с кросвалидация за моделиране на нивата на PM10 в зависимост от 8 метеорологични променливи и 1 времева променлива. Освен това модели са построени и за трансформация на PM10, която има почти нормално разпределение. Проведен е анализ и диагностика на моделите, най- добрите модели достигат R^2=84% и RMSE=11.694.

Получено е много добро съвпадение с експеримента с до 78% без кросвалидация и с до 84% използвайки кросвалидация. Илюстрирано е приложение на моделите за прогнозиране на замърсяването с 2 дни напред спрямо използваните измервания. Установена е силата на влияние на отделните променливи и най-голямо значение за високите концентрации на PM10 са концентрациите на PM10 от предишния ден, а след него е измерената минимална температура от вчерашния ден в моделите.

От направените анализи и получените резултати, може да направим заключение, че моделите построени с кросвалидация и отчитане на замърсяването от предишните два дни, измерванията за минималната температура и посоката на вятъра от предишния ден, дават по-точни резултати.

Проведеният анализ е алтернатива на официалните доклади на регионалната инспекция по околната среда и водите – Плевен и независимо изследване на измерените концентрации на PM10.

Заключение

Представените изследвания и изводи позволяват да се заключи, че целта и задачите на дисертационния труд са постигнати. Основната част от получените резултати са публикувана в 3 научни публикации. Изнесени са 5 доклада на научни форуми и семинари.

Резюме на получените резултати

В дисертационния труд се изследват и сравняват възможностите на два типа високоефективни методи за моделиране на бързопроменливи времени редове на основните въздушни замърсители (фини прахови частици ФПЧ10 и ФПЧ2.5) – методите АРИМА за стохастично моделиране и дейта майнинг методът на класификационните и регресионни дървета (CART). Методът CART за първи път в литературата се прилага с набор от лагирани променливи на зависимата променлива и на предикторите. Друг нов подход е използването на групи от предиктори и построяване на оптимални и максимални дървета. Заедно с класификацията на случаите и машинното обучение с кросвалидация, е показано, че това повишава гъвкавостта и ефективността на метода.

Научни и научно-приложни приноси, защитавани от автора

1. Построени и анализирани са стохастични едномерни модели по методологията на Бокс-Дженкинс на среднодневните концентрации на фини прахови частици ФПЧ10 във въздуха на град Перник за 5-годишен период. Моделите са приложени за получаване на 7 дневни прогнози и са получени много добри резултати. Моделите описват около 56.6% от данните.

2. За същите данни са построени и изследвани CART модели, като са използвани лагирани променливи на зависимата PM10, три вида замърсители – CO, SO2, NO2, и метеорологични променливи с лагиране. Най-добрите модели съвпадат с измерените стойности на PM10 до 93.7%. Моделите са приложени за прогнозиране на бъдещи замърсявания за 7 дни напред и показват отлични резултати.

3. Построени и изследвани са едномерни и многомерни 2D, 3D и 4D АРИМА модели за среднодневните замърсявания с PM10 и PM2.5 на Пловдив и Асеновград за период от 5 години. Установено е, че получените най-добри едномерни модели са близки по характер. Многомерните модели имат същия тип параметри. Така е показано, че замърсяването е еднакво по сила в целия регион, обхващащ Пловдив и Асеновград. Моделите обясняват от 59 до 61% от данните. Те са приложени за краткосрочни прогнози с много добри показатели.

4. За данни от измервания на среднодневни концентрации на PM10 за град Плевен в период от 6 години са построени и анализирани множество модели с CART метод с групи предиктори, без кросвалидиране и без лагиране. Най-добрите модели обясняват до 78% от наблюдаваните данни и 88% от бъдещите прогнози за 2 дни.

5. Построени и изледвани са CART модели за данните на град Плевен с използване на групи предиктори, лагиране и кросвалидация. Моделите описват до 84% от измерените PM10 концентрации. Моделите са приложени за краткосрочни прогнози за 2 дни напред и показват отлични резултати с 91% познати прогнози спрямо лимита от 50 микрограма на кубичен метър.

Таблица. Връзка на резултатите с целите и публикациите

Принос	Цел	Задача	Параграф	Публикации
1	1	1	2.3, 2.4, 2.5	[P1]
2	1	2	2.6	-
3	1	3	3.3, 3.4, 3.5, 3.6, 3.7	[P2]
4	1	4	4.3, 4.3.1	[P3]
5	1	5	4.3.2	-

Перспективи за бъдеща работа

По-нататъшни изследвания в областта на тематиката на настоящия дисертационен труд могат да бъдат насочени в следните направления:

- Прилагане на разработените методи за подобен вид бързопроменливи времени редове в областта на екологията за замърсители на атмосферния въздух, водите, почвите и др.
- Прилагане на нови методи на дейта майнинг техники като Boosted Regression Tree (BRT), Random Forest (RF) и други за бързопроменливи временни редове от областта на финансовите пазари, маркетинга, бизнес процесите и др.
- Разработка и анализиране на хибриден тип методи за математическо моделиране на временни редове, комбиниращи възможностите на стохастичните и дейта майнинг методи

Списък на публикациите по темата на дисертационния труд

[P1] **M. P. Stoimenova**, „Stochastic Modeling of Problematic Air Pollution with Particulate Matter in the City of Pernik, Bulgaria“, Ecologia Balkanica, vol. 8, issue 2, pp. 33-41, 2016. ISSN: 1314-0213. http://eb.bio.uni-plovdiv.bg/en/archive/2016-vol.8-issue-2

[P2] S. Gocheva-Ilieva, **M. Stoimenova**, A. Ivanov, D. Voynikova, I. Iliev, „Stochastic univariate and multivariate time series analysis of PM2.5 and PM10 air pollution: A comparative case study for Plovdiv and Asenovgrad, Bulgaria“, Eighth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences, Albena, Bulgaria, June 22-27, 2016, Ed. M. Todorov, American Institute of Physics, AIP Conf. Proc. vol.1773, pp. 110004-1 –110004-10, 2016. ISBN: 978-0-7354-1431-0, http://dx.doi.org/10.1063/1.4965008 **SCImagoJR = 0.163**

[P3] **M. Stoimenova**, D. Voynikova, A. Ivanov, S. Gocheva-Ilieva, I. Iliev, Regression trees modeling and forecasting of PM10 air pollution in urban areas, Ninth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences, Albena, Bulgaria, 21–26 June 2017, Ed. M. Todorov, American Institute of Physics, AIP Conf. Proc. vol. 1895, pp. 030005-1–030005-10, 2017. ISBN: 978-0-7354-1579-9, https://doi.org/10.1063/1.5007364 **SCImagoJR = 0.163 (за 2016 г.)**

Апробация на резултатите

А) Доклади, изнесени на научни форуми и семинари

1. M. Stoimenova, Dimitar Fidanov, Picard sequence for solution of first order ordinary differential equation, Sixth International Workshop - 9 July 2015 – 12 July 2015, Miskolc, Hungary
2. Стоименова М., Статистическо моделиране на проблемни замърсители на въздуха на град Перник, Семинар по Изчислителна статистика, ФМИ на ПУ, 30.05.2016.
3. S. Gocheva-Ilieva, M. Stoimenova, A. Ivanov, D. Voynikova and I. Iliev - Stochastic Univariate and Multivariate Time Series Analysis of PM2.5 and PM10 Air Pollution: a Comparative Case Study for Plovdiv and Asenovgrad, Bulgaria, Eighth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences, Albena, Bulgaria, June 22-27, 2016.
4. Стоименова М., Моделиране и прогнозиране на замърсявания с фини прахови частици (PM10) в зависимост от метеорологичните условия с CART метод, Семинар по Изчислителна статистика, ФМИ на ПУ, 14.06.2017.
5. M. Stoimenova, D. Voynikova, A. Ivanov, S. Gocheva-Ilieva, and I. Iliev, Regression trees modeling and forecasting of PM10 air pollution in urban areas, Ninth Conference of the Euro-American Consortium for Promoting the Application of Mathematics in Technical and Natural Sciences, Albena, Bulgaria, June 21-26, 2017.

Б) Участие в проекти

1. Научен проект НИ15-ФМИ-004 (2015-2016), тема: „Иновативни фундаментални и приложни научни изследвания по математика, информатика и педагогика на обучението", Фонд „Научни изследвания" при ПУ „Паисий Хилендарски", 2015/2016 г.
2. СП15-ФМИИТ-007/24.04.2015 „Надграждане на знания и формиране на компетенции у студентите за работа със специализиран математически софтуер за решаване на приложни математически задачи", Фонд „Научни изследвания" при ПУ „Паисий Хилендарски", 2015/2016 г.
3. МУ17-ФМИ-003, 25.04.2017 г. – 30.11.2018 г. , тема: „Дейта майнинг моделиране и проложни софтуерни системи", Фонд „Научни изследвания" при ПУ „Паисий Хилендарски", 2017/2018 г.

В) Преминати специализирани докторантски курсове по докторската програма

1. "Numerical-analytic and constructive methods for boundary-value problems", 29.06.2015 - 12.07.2015, CEEPUS Summer University – 2015, Мишколц, Унгария.
2. "New Aspects of the Time Frequency Analysis Involving Fréchet Frames", 28.09.2016 - 03.10.2016, DAAD Intensive course, Нови Сад, Сърбия.
3. Езиков курс по английски език – А1 и А2.

Декларация за оригиналност

от Мая Пламенова Стоименова,
катедра „Приложна математика и моделиране”,
Факултет по математика и информатика
на Пловдивския университет „Паисий Хилендарски”

Във връзка с провеждането на процедура за придобиване на образователна и научната степен „Доктор“ във ФМИ на ПУ „Паисий Хилендарски” и защита на представения от мен дисертационен труд на тема: „**Моделиране на бързопроменливи временни редове“**, декларирам:

Резултатите и приносите в проведените научни изследвания, включени в дисертационния труд са оригинални и не са заимствани от изследвания и публикации, в които нямам участие.

25.02.2018 г. Декларатор:
гр. Пловдив Мая Стоименова

Благодарности

Изразявам своята сърдечна благодарност и признателност към научния си ръководител проф. д.м.н. Снежана Гочева – Илиева, за пълната й подкрепа, постоянна ангажираност и перфектно отношение към моята работа.

Благодаря на ръководството на Факултета по математика и информатика, за подкрепата и съдействието им. Благодарност изказвам и на членовете на катедра „Приложна математика и моделиране“, които са били съпричастни към работата ми и обучението ми като докторант.

БИБЛИОГРАФИЯ

[1] (WHO), World Health Organization. (2013) Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia. <www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf>

[2] Abdullah S., Ismail M., Fong S. Y., "Multiple linear regression (MLR) models for long term PM10 concentration forecasting during different monsoon seasons", *Journal of Sustainability Science and Management*, vol. 12, no. 1, pp. 60-69, 2017.

[3] *Air Pollution*. World Health Organization. <http://www.who.int/topics/air_pollution/en/>

[4] *Air Quality Guidelines for Europe*, 2nd ed. (World Health Organization (WHO), Regional Office for Europe, 2000. <http://apps.who.int/iris/handle/10665/107335>

[5] *Air quality in Europe - 2014 report*. European Environment Agency, Publications, 19 Nov 2014. <http://www.eea.europa.eu/publications/air-quality-in-europe-2014/at_download/file>

[6] *Air Quality Standards*. European Commission.Environment. <http://ec.europa.eu/environment/air/quality/standards.htm>

[7] Anderson J. O., Thundiyil J. G., Stolbach A., "Clearing the air: A review of the effects of particulate matter air pollution on human health", *Journal of Medical Toxicology*, vol. 8, no. 2, pp. 166-175, 2012.

[8] Anderson D. A., Burnham K. P., *Model selection and inference*, 2nd ed. Colorado State, USA, 2003.

[9] Biancofiore F., Busilacchio M., Verdecchia M., Tomassetti B., Aruffo E., Bianco S., Di Tommaso S. , Colangeli C., Rosatelli G. , Di Carlo P., "Recursive neural network model for analysis and forecast of PM10 and PM2.5", *Atmospheric Pollution Research*, vol. 8, no. 4, pp. 652-659, 2017.

[10] Bøhler T., Karatzas K., Peinel G., Rose T., Jose R. S., "Providing multi-modal access to environmental data -customizable information services for disseminating urban air quality information in APNEE", *Computers, Environment and Urban Systems*, vol. 26, no. 1, pp. 39-61, 2002.

[11] Box G. E. P., Jenkins G. M., Reinsel G. S., *Time series analysis, forecasting and control*. 3rd ed., New Jersey, Prentice-Hall, Inc., 1994.

[12] Boylan J., Odman M., Wilkinson J., Russell A., "Integrated assessment modeling of atmospheric pollutants in the southern appalachian mountains: Part II. Fine particulate matter and visibility", *Journal of the Air & Waste Management Association*, vol. 56, pp. 12-22, 2012.

[13] Breiman L., Friedman J., Olshen R., Stone C., *Classification and Regression Trees*, 1st ed., Belmont, Wadsworth Advanced Books and Software, Canada, 1984.

[14] Burrows W. R., Benjamin M., Beauchamp S., Lord E. R., McCollor D., Thomson B., "CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada", *American Metodological Society*, vol. 34, pp. 1848-1862, 1995.

[15] Cambra-López M., Hermosilla T., Lai H. T. L., Aarnink A. J. A., Ogink N. W. M., "Particulate matter emitted from poultry and pig houses source indentification and quantification", *American Society of Agricultural and Biological Engineers*, vol. 54, no. 2, pp. 629-642, 2011.

[16] Cambra-López M., Hermosilla T., Lai H. T. L., "Source identification and quantification of particulate matter emitted from livestock houses", International Symposium on Air Quality and Manure Management for Agriculture Conference Proceedings, 13-16 September 2010, Dallas, Texas, pp. 41, 2010.

[17] CART® *Classification and Regression Trees*, 2012. <http://www.salford-systems.com/en/products/cart>

[18] Choi W., Paulson S. E., Casmassi J., Winer A. M., "Evaluating meteorological comparability in

air quality studies: Classification and regression trees for primary pollutants in California's South coast air basin", *Atmospheric Environment*, vol. 64, pp. 150-159, 2013.

[19] Choi J., Fuentes M., Reich B. J., "Spatial–temporal association between fine particulate matter and daily mortality", *Computational Statistics and Data Analysis*, vol. 53, no. 8, pp. 2989-3000, 2009.

[20] Cox T., Popken D., Ricci P. F., "Temperature, not fine particulate matter (PM2.5), is causally", *International Dose-Response Society*, vol. 11, pp. 319-343, 2013.

[21] Cutler D. R., Edwards T. C., Beard K. H., Cutler A., Hess K. T., "Random Forest for classification in ecology", *Ecology - Ecological Society of America*, vol. 88, pp. 2783-2792, 2007.

[22] De'Ath G., "Boosted Trees for ecological modeling and prediction", *Ecology- Ecological Society of America*, vol. 88, pp. 243-251, 2007.

[23] Directive 2008/50/EC of the European Parliament and of the council of 21 May 2008 on ambient air quality and cleaner air for Europe, *Official Journal of the European Union* L 152/1, 2008. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044>

[24] Dong-jun L., Shenzhen L.L., "Application study of comprehensive forecasting Model based on entropy weighting method on trend of PM2.5 concentration in Guangzhou, China", *Int. J. Environ. Res. Public Health*, vol. 12, pp. 7085-7099, 2015.

[25] EEA Daily Bulletin for air quality in the country. EEA - Executive environment agency, National system for realtime air quality control in Bulgaria. <http://pdbase.government.bg/airq/bulletin-en.jsp>

[26] Ehsanzadeh A., Nejadkoorki F., Khodadoostan S., "A study on the most important factors affecting the concentration of particulate matter smaller than 10 microns (PM10) using principal component regression", *Journal of Research in Environmental Health*, vol. 2, no. 2, pp. 154-164, 2016. <http://eprints.mums.ac.ir/id/eprint/7622>

[27] European Environment Agency. <http://www.eea.europa.eu/themes/air/air-quality-index/index>

[28] Executive Environment Agency (ExEA), Bulgaria. <http://eea.government.bg/en>

[29] Eynaud Y., Nerini D., Baklouti M., Poggiale J. C., "Towards a simplification of models using regression trees", *Journal of the Royal Society Interface*, vol. 9, 1-14, 2012.

[30] Ganesh S. S., Arulmozhivarman P., Rao Tatavarti, "Forecasting air quality index using an ensemble of artificial neural networks and regression models", *Journal of Intelligent Systems*, vol. 34, no. 1, 2017.

[31] Gardner M. W., Dorling S. R., "Statistical surface ozone models: an improved methodology to account for non-linear behaviour", *Atmospheric Environment*, vol. 34, no. 1, pp. 21-34, 2000.

[32] Gass K., Klein M., Chang H. H., Flanders W. D., Strickland M. J., "Classification and regression trees for epidemiologic research: an air pollution example", *Environmental Health*, vol. 13, no. 1, pp. 17-26, 2014.

[33] Georgieva E., Syrakov D., Prodanova M., Etropolska I., Slavov K., "Evaluating the performance of WRF-CMAQ air quality modelling system in Bulgaria by means of the DELTA tool", *International Journal of Environment and Pollution*, vol. 57, no. 3-4, pp. 272-284, 2015.

[34] Glymour C., Madigan D., Pregibon D., "Statistical themes and lessons for data mining", *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 11-28, 1997.

[35] Gmur S., Vogt D., Zabowski D., Moskal L. M., "Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees", vol. 12, no. 8, pp. 10639-10658, 2012.

[36] Gocheva-Ilieva S. G., Ivanov A. V., Iliev I. P., "Modeling of air pollutants and ozone concentration by using multivariate analysis: Case study Dimitrovgrad, Bulgaria", *British Journal of Applied Science & Technology*, vol. 7, no. 10, pp. 1-18, 2016.

[37] Gocheva-Ilieva S. G., Ivanov A. V., Voynikova D. S., Boyadzhiev D. T., "Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach",

Stoch. Environ. Res. Risk Assess., vol. 28, no. 4, pp. 1045-1060, 2014.

[38] He Y., Novac J., Glosemeyer D., *Mathematica Time Series: A Fully Integrated Environment for Time-Dependent Data Analysis (Wolfram Research, Inc., Illinois, 2007)*., 2007. <http://media.wolfram.com/documents/TimeSeriesDocumentation.pdf>

[39] Henry R., Park E.-S., Spiegelmanc C. H., "Comparing a new algorithm with the classic methods for estimating the number of factors", *Chemometrics and Intelligent Laboratory Systems*, vol. 48, no. 1, pp. 91-97, 1999.

[40] Hu W., Mengersen K., McMichael A., Tong S., "Temperature, air pollution and total mortality during summers in Sydney, 1994–2004", *International Journal of Biometeorology*, vol. 52, no. 7, pp. 689-696, 2008.

[41] Ivanov A., Gocheva-Ilieva S., "Short-time particulate matter PM10 forecasts using predictive modeling techniques", *AIP Conf. Proc.*, Todorov M. (ed), vol. 156, pp. 209-218, 2013.

[42] Ivanov A., Voynikova D., Gocheva-Ilieva S., Kulina H., Iliev I., "Using principal component analysis and general path seeker regression for investigation of air pollution and CO modeling", *AIP Conf. Proc.*, Todorov M. (ed.), vol. 1684, 100004, pp. 1-11, 2015.

[43] Izenman A. J., *Modern Multivariate Statistical Techniques Regression Classification, and Manifold Learning.* Springer, New York, 2008.

[44] Jacobson M. Z., *Fundamentals of Atmospheric Modeling, 2nd edn.* Cambridge Univ. Press, Cambridge, 2005.

[45] Jerrett M., Burnett R. T., Beckerman B. S., "Spatial analysis of air pollution and mortality in California", *American Journal of Respiratory and Critical Care Medicine*, vol. 188, no. 5, pp. 593-599, 2013.

[46] Jian L., Zhao Y., Zhu Y. P., Zhang M. B., Bertolatti D., "An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China", *Science of The Total Environment*, vol. 426, pp. 336-345, 2012.

[47] Junmin Li., Wang L., "The research of PM2.5 concentrations model based on regression calculation model", *AIP Conf. Proc.*, vol. 1794, 030005, 2017.

[48] Kadiyala A., Kumar A., "Vector time series-based radial basis function neural network modeling of air quality inside a public transportation bus using available software", *Environmental Progress & Sustainable Energy*, vol. 36, pp. 4-10, 2017.

[49] Kubošová K., Komprda J., Jarkovský J., "Spatially resolved distribution models of POP concentrations in soil: A stochastic approach using regression trees", *Environmental Science & Technology*, vol. 43, no. 24, pp. 9230-9236, 2009.

[50] Kumar R., Joseph A. E., "Air pollution concentrations of PM2.5, PM10 and NO2 at ambient and Kerbsite and their correlation in Metro City-Mumbai," *Environmental Monitoring and Assessment*, vol. 119, no. 1-3, pp. 191-199, 2006.

[51] Lima E. A. P., Guimaraes E. C., Pozza S. A., Barrozo M. A. S., Coury J. R., "A study of atmospheric particulate matter in a city of the central region of Brazil using time-series analysis", *International Journal of Environmental Engineering*, vol. 1, pp. 80-94, 2009.

[52] Liu P. W. G., "Simulation of the daily average PM10 concentrations at Ta-Liao with Box-Jenkins time series models and multivariate analysis", *Atmospheric Environment*, vol. 43, no. 13, pp. 2104-2113, 2009.

[53] Maji J. K., Dikshit A. K., Deshpande A., "Disability - adjusted life years and economic cost assessment of the health effets related to PM2.5 and PM10 pollution in Mumbai and Delhi, in India from 1991 to 2015", *Environ. Sci. Pollut. Res.*, vol. 24, no. 5, pp. 4709-4730, 2017.

[54] *MARS 3.0, Technical guide*. Salford Systems, San Diego, 2011.

[55] Martın M. L., Turias I. J., Gonzalez F. J., Galindo P. L., Trujillo F. J., Puntonet C. G., Gorriz J. M., "Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks", *Chemosphere*, vol. 70, no. 7, pp. 1190-1195, 2008.

[56] Mas S., Juan A., Tauler R., Olivieri A. C, "Application of chemometric methods to environmental analysis of organic pollutants: A review", *Talanta*, vol. 80, no. 3, pp. 1052-1067, 2010.

[57] McBerthouex P., Brown L. C., *Statistics for Environmental Engineers*, 2nd ed. Lewis Publishers, CRC Press LLS, Boca Raton, 2002.

[58] Mehtal, S., Smith K. R., Balakrishnan K., Sankar S., Padmavath R., Kumar S., Akbar S., "Using household characteristics topredict respirable particulate levels in rural households of Andhra Pradesh, India". In: *9th International Conference on Indoor Air Quality and Climate*, pp. 596-601, 2002.

[59] Michaelides Silas, Paronis D., Retalis A., Tymvios F., "Monitoring and forecasting air pollution levels by exploiting satellite, ground-based, and synoptic data, elaborated with regression models", *Advances in Meteorology*, vol. 2017, p. 17, 2017.

[60] Milionis A. E., Davies T. D., "Regression and stochastic models for air pollution. I. Review, comments and suggestions.", *Atmospheric Environment*, vol. 28, no. 17, pp. 2801-2810, 1994.

[61] Moisan S., Herra R., Clements A., "A Dynamic Multiple Equation Approach for Forecasting PM2.5 Pollution in Santiago, Chile", NCER Working Paper Series, pp. 1-37, 2017. <http://www.ncer.edu.au/papers/documents/WP117.pdf>

[62] Moldovan O. T., Meleg I. N., Levei E., Terente M., "A simple method for assessing biotic indicators and predicting biodiversity in the hyporheic zone of a river polluted with metals", *Ecological Indicators*, vol. 24, pp. 412-420, 2013.

[63] Murtaugh P. A., "Performance of several variable-selection methods applied to real ecological data", *Ecology Letters*, vol. 12, no. 10, pp. 1061-1068, 2009.

[64] Nastos P. T., Paliatsos A. G., Anthracopoulos M. B., Roma E. S., Priftis K. N., "Outdoor particulate matter and childhood asthma admissions in Athens, Greece: a time-series study", *Environmental Health*, vol. 9, pp. 45-53, 2010.

[65] Nazif A., Mohammed N. I., Malakahmad A., Abualqumboz M. S., "Application of step wise regression analysis in predicting future particulate matter concentration episode", *Water Air Soil Pollution*, vol. 227, no. 4, p. 117-128, 2016.

[66] Neto, J., Ferreira F., Torres P. M., Boavida F., "Lisbon air quality forecast using statistical methods", *International Journal of Environment and Pollution*, vol. 39, no. 3-4, pp. 333-339, 2009.

[67] Nisbet R., Elder J., Miner G., *Handbook of Statistical Analysis and Data Mining Applications*, 2nd ed. Burlington, MA, Academic Press Elsevier Inc., 2009.

[68] Pearson K., *Mathematical Contributions to the Theory of Evolution*. Drapers' Company Research Memoirs. London, vol. 13-17, 1904. <https://archive.org/stream/b24397933#page/n11/mode/2up>

[69] Peng R. D., Chang H. H., Bell M. L., McDermott A., Zeger S. L., Samet J. M., Dominici F., "Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among medicare patients", *JAMA*, vol. 299, no. 18, pp. 2172-2179, 2008.

[70] Pohoata A., Lungu E., "A complex analysis employing ARIMA model and statistical methods on air pollutants recorded in Ploiesti, Romania", *Revista de Chimie*, vol. 68, pp. 818-823, 2017. <http://www.revistadechimie.ro>

[71] Polat K., Durduran S. S., "Usage of output dependent data scaling in modeling and prediction of air pollution daily concentration values (PM10) in city of Konya"", *Neural Computing and Applications*, vol. 8, pp. 2153-2162, 2012.

[72] Potdar K., Pardawala T. S., "Forecasting ambient air quality in mumbai using neural networks", In: *5th National Conf. on Role of Engineers in Nation Building*, 2017.

[73] Prakash A., Kumar U., Kumar K., Jain V., "A wavelet-based neural network model to predict ambient air pollutants' concentration", *Environmental Modeling & Assessment*, vol. 16, no. 5,

pp. 503-517, 2011.

[74] Prasad A., Iverson L. R., Liaw A., "Newer classification and regression tree techniques: bagging and random forests for ecological prediction", *Ecosystems*, vol. 9, pp. 181-199, 2006.

[75] Regional Inspectorate of Environment and Water - Pleven, Reports on the state of the environment, 2011-2016. <http://riew-pleven.eu/doc/docladiOS/>

[76] Reid C.E., Jerrett M., Petersen M. L., Pfister G. G., "Spatiotemporal prediction of fine particulate matter during the 2008 Northern California Wildfires using machine learning", *Environ. Sci. and Technol.*, vol. 49, no. 6, pp. 3887-3896, 2015.

[77] *Report on the state of air quality*. 2015. RIOSV Pernik. (2015) <http://pk.riosv-pernik.com/index.php?option=com_content&view=category&id=74:revisheniq&Itemid=28&layout=default>

[78] Roy A., Georgopoulos P. G., Ouyang M., Freeman N., Lioy P. J., "Environmental, dietary, demographic, and activity variables associated with biomarkers of exposure for benzene and lead", *Journal of Exposure Science & Enviromental Epidemiology*, vol. 13, no. 6, pp. 417-426, 2003.

[79] Saeed, S., Hussain L., Awan I. A., Idris A., "Comparative analysis of different statistical methods for prediction of PM2.5 and PM10 concentrations in advance for several hours", *International Journal of Computer Science and Network Security*, vol. 17, pp. 45-52, 2017. <http://paper.ijcsns.org/07_book/201711/20171106.pdf>

[80] *Salford Systems Data Mining and Predictive Analytics Software Modeler, SPM* Version 8.0. Salford Systems, San Diego, 2016.

[81] Sayegh A., Tate J. E., Ropkins K., "Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees", *Atmospheric Environment*, vol. 127, pp. 163-175, 2016.

[82] Sayegh A. S., Munir S., Habeebullah T. M., "Comparing the performance of statistical models for predicting PM10", *Aerosol and Air Quality Research*, vol. 14, no. 3, pp. 653-665, 2014.

[83] Shahraiyni H. T., Sodoudi S., Kerschbaumer A., Cubasch U., "Re-construction of the shut-down PM10 monitoring stations for the reliable assessment of PM10 in Berlin using fuzzy modelling and data transformation", *Environ Monit Assess*, vol. 189, no. 3, pp. 133-145, 2017.

[84] Sharma P., Chandra A., Kaushik S. C., "Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City", *Environmental Monitoring and Assessment*, vol. 157, no. 1-4, pp. 105-112, 2009.

[85] Singh K., Gupta S., Rai P., "Identifying pollution sources and predicting urban air quality using ensemble learning methods", *Atmospheric Environment*, vol. 80, pp. 426-437, 2013.

[86] Slini T., Kaprara A., Karatzas K., Moussiopoulos N., "PM10 Forecasting for Thessaloniki, Greece", *Environ Modell Softw.*, vol. 21, no. 4, pp. 559-565, 2006.

[87] Slini Th., Karatzas K., Moussiopoulos N., "Statistical analysis of environmental data as the basis of forecasting: an air quality application", *Sci. Tot. Environ.*, vol. 288, no. 3, pp. 227-237, 2002.

[88] Spruill T. B., Showers W. J., "Application of classification-tree methods to identify nitrate sources in ground water", *Journal of Environmental Qualit*, vol. 31, pp. 1538-1549, 2002.

[89] Srinivas K., Raghavendra Rao G., Govardhan A., "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", In: *The 5th International Conference on Computer Science & Education Hefei, China. August 24-27*, pp. 1344 - 1349, 2010.

[90] Stadlober E., Hubnerova Z., Michalek J., Kolar M., "Forecasting of daily PM10 concentrations in Brno and Graz by different regression approaches", *Austrian Journal of Statistics*, vol. 41, pp. 287-310, 2012.

[91] Statistics, SPSS IBM. < http://www-01.ibm.com/software/analytics/spss/ >

[92] Steinberg D., Colla P., *CART: Tree-Structured Non-Parametric Data Analysis.*, Salford systems,

1995.

[93] Suleiman A., Tight M. R., Quinn A. D, "Hybrid neural networks and boosted regression tree models for predicting roadside particulate matter", *Environmental Modeling & Assessment*, vol. 21, no. 6, pp. 731-750, 2016.

[94] Sun, Z., Tao Y., Li S., Ferguson K. K., Meeker J. D., Park S. K., Batterman S. A., Mukherjee B., "Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons", *Environmental Health*, vol. 12, no. 1, pp. 85-104, 2013.

[95] Timofeev R. , *Classification and Regression Trees (CART) Theory and Applications*. Berlin, 2005.

[96] UCLA Newsroom. <http://newsroom.ucla.edu/releases/Urban-Air-Pollution-Linked-to-Birth-2932>

[97] Ul-Saufie A. Z., Yahaya A. S., Ramli N. A., Rosaida N., Hamid H. A., "Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis", *Atmospheric Environment*, vol. 77, pp. 621-630, 2013.

[98] Vesely V., Tonner J., Hrdlivckova Z., Michalek J., Kolar M., "Analysis of PM10 air pollution in Brno based on generalized linear model with strongly rank-deficient design matrix", *Environmetrics.*, vol. 20, no. 6, pp. 676–698, 2009.

[99] Vinceti M., Malagoli C., Malavolti M., Cherubini A., Maffeis G., Rodolfi R., Heck J. E., Astolfi G., Calzolari E., Nicolini et al., "Does maternal exposure to benzene and PM10 during pregnancy increase the risk of congenital anomalies? A population-based case–control study", *Science of The Total Environment*, vol. 541, pp. 444-450, 2016.

[100] Viscarra R. R. A., Behrens T., "Using data mining to model and interpret soil diffuse reflectance spectra", *Geoderma*, vol. 158, no. 1, pp. 46-54, 2010.

[101] Vlachogianni A., Kassomenos P., Karppinen A., Karakitsios S., Kukkonen J., "Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki", *Science of The Total Environment*, vol. 409, no. 8, pp. 1559-1571, 2011.

[102] Voukantsis D., Karatzas K., Kukkonen J., "Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki", *Sci. Tot. Environ.*, vol. 409, no. 7, pp. 1266-1276, 2011.

[103] Voynikova D. S., Gocheva-Ilieva S. G., Ivanov A. V., Iliev I. P., "Studying the effect of meteorological factors on the SO2 and PM10 pollution levels with refined versions of the SARIMA model", In: *AIP Conf. Proc.*, Todorov M. (ed), vol. 1684, 100005, pp. 1-12, 2015.

[104] Wang C., Xiaodan Zhou, Renjie Chen,Xiaoli Duan, Xingya Kuang, Haidong Kan, "Estimation of the effects of ambient air pollution on life expectancy of urban residents in China", *Atmospheric Environment*, vol. 80, pp. 347-351, 2013.

[105] Wei C. L., Rowe G. T., Escobar-Briones E., Boetius A., "Global patterns and predictions of seafloor biomass using random forests", *Plos one Tenth Anniversary*, vol. 5, no. 12, p. e15323, 2010.

[106] Whalley J., Zandi S., "Particulate matter sampling techniques and data modelling methods", In: *Air Quality - Measurement and Modeling*, Sallis P. (ed.), INTECH, ch. 2, pp. 29-54, 2016.

[107] Wikipedia. <https://bg.wikipedia.org/wiki/%D0%92%D1%8A%D0%B3%D0%BB%D0%B5%D1%80%D0%BE%D0%B4%D0%B5%D0%BD_%D0%BE%D0%BA%D1%81%D0%B8%D0%B4>

[108] Wilks D. S., *Statistical methods in the atmospheric sciences*, 3rd ed., 2011.

[109] Witten I. H., Frank E., Hall M. A., Pal C. J., *Data Mining: Practical machine learning tools and techniques*, 4th ed. India: Todd Green, 2016.

[110] WolframMathematica,. <http://www.wolfram.com/mathematica/>

[111] Wongsathan Rati, Seedadan I., Wanasri S., "Hybrid forecast models for PM-10 prediction: A case study of Chiang Mai city of Thailand during high season", *KKU Engineering Journal*, vol. 43, no. S2, pp. 203-206, 2016.

[112] Xu Gang, Jiao L., Zhang B., Zhao S., Yuan M., Gu Y., Liu J., Tang X., "Spatial and temporal variability of the PM2.5/PM10 ratio in Wuhan Central China", *Aerosol and Air Quality Research*, vol. 17, pp. 741-751, 2017.

[113] Yeo I. K., Johnson R. A., "A new family of power transformations to improve normality or symmetry", *Biometrika*, vol. 87, pp. 954-959, 2000.

[114] Yim S. H. L., Barrett S. R. H., "Public health impacts of combustion emissions in the United Kingdom", *Environmental Science and Technology*, vol. 46, no. 8, pp. 4291-4296, 2012.

[115] Yuanyuan C., Runhe S., Shijie S., Wei G., "Ensemble and enhanced PM10 concentration forecast model based on stepwise regression and wavelet analysis", *Atmospheric Environment*, vol. 74, pp. 346-359, 2013.

[116] Yuen K. K. F., "Towards multiple regression analyses for relationships of air quality and weather", *Journal Adv. Inform. Technol.*, vol. 8, pp. 135-140, 2017.

[117] Zhao M., Li X., "An application of spatial decision tree for classification of air pollution index", In: *19th Intern. Conf. on Geoinformatics*, pp. 1-6, 2011.

[118] Zhao C., Song G., "Application of data mining to the analysis of meteorological data for air quality prediction: A case study in Shenyang", *Earth and Environmental Science*, vol. 81, 012070, 2017.

[119] Zheleva I., Veleva E., Filipova M., "Analysis and modeling of daily air pollutants in the city of Ruse, Bulgaria", In: *AIP Conference Proceedings*, Todorov M. (ed.), vol. 1895, 030007, 2017.

[120] Zickus M., Greig A. J., Niranjan M., "Comparison of four machine learning methods for predicting PM10 concentrations in Helsinki, Finland", *Water, Air, & Soil Pollution: Focus*, vol. 2, no. 5, pp. 717-729, 2002.

[121] Zwozdziak A., Samek L., Sowka I., Furman L., Skrętowicz M., "Aerosol pollution from small combustors in a village", *The Scientific World Journal*, vol. 2012, pp. 1-8, 2012.

[122] Георгиев О., Велева Б. , Христова Е., Бъварова Е., Коларова М., "Връзка между метеорологичните характеристики и замърсяването на въздуха в София", *3rd National Congress on Physical Sciences, 29 Sep. – 2 Oct. 2016, Sofia*, 2016.

[123] Наредба № 9 от 3 май 1999 г. за норми за серен диоксид, азотен диоксид, фини прахови частици и олово в атмосферния въздух. <http://econ.bg/Нормативни-актове/Наредба-9-от-3-май-1999-г-за-норми-за-серен-диоксид-азотен-диоксид-фини-прахови-частици-и-_l.l_i.128774_at.5.html>