

## СИНТЕЗ НА МАТЕМАТИЧЕСКИ МОДЕЛ ЗА АНАЛИЗ И ОЦЕНКА НА КОНЦЕНТРАЦИИ НА ПРАХОВИ ЧАСТИЦИ В ОТКРИТИ ПРОСТРАНСТВА

Николай Стоянов, Антония Панделова,  
Цанко Георгиев, Юлия Калъпчийска

**Резюме.** Размерите на праховите частици, химичния им състав и адсорбираните на повърхността им елементи влияят пряко върху здравословното състояние на хората. За мониторинг и прогнозиране на нива на прахови замърсявания се прилагат различни модели и системи. В представената работа е разработен линеен множествен регресионен модел за разкриване на причинно-следствена връзка между концентрацията на  $PM_{10}$  и следните независими променливи: средноденонощна температура на въздуха, средноденонощна слънчева радиация, скорост на вятъра, посока на вятъра, средноденонощно атмосферно налягане. За извършване на необходимите анализи и изчисления е използван статистически програмен пакет за обработка на данни **STATGRAPHICS**.

**Ключови думи:** атмосферен въздух,  $PM_{10}$ , математически модел

## SYNTHESIS OF MATHEMATICAL MODEL FOR ANALYSIS AND EVALUATION OF PARTICULATE MATTER CONCENTRATIONS IN OUTDOOR SPACES

Nikolay Stoyanov, Antonia Pandelova,  
Tzanko Georgiev, Julia Kalapchiiska

**Abstract.** The size of the dust particles, their chemical composition and the elements adsorbed on their surface directly affect the health of people. Different models and systems are used for monitoring and forecasting dust pollution levels. In the presented work a linear multiple regression model is developed to reveal a causal relationship between the concentration of  $PM_{10}$  and the following independent variables: average daily air temperature, average daily solar radiation, wind speed, wind direction, average daily atmospheric pressure. **STATGRAPHICS** statistical software package was used to perform the necessary analyzes and calculations.

**Keywords:** atmospheric air,  $PM_{10}$ , mathematical model

## 1. ВЪВЕДЕНИЕ

През последните години в световен мащаб се наблюдава тенденция към увеличаване концентрацията на прахови частици във въздуха. Високите нива на замърсяване на атмосферата от различни антропогенни дейности на територията на Република България са част от този световен проблем. Влошаването на качеството на въздуха е най-осезаемо в населените градски зони [1, 2]. За страната ни района на гр. София е най-силно засегнат от емитираните вредни вещества във въздуха. Измерените концентрации на РМ в столицата превишават системно допустимата СДН, регламентирана в ЗЧАВ. Поддържането на нивата на вредните емисии под допустимите граници е основна задача пред управляващите и живущите в Столична община.

За създаване на експериментален модел и доказване на неговата адекватност се използват методите на математическата статистика, при които се извършва проверка на статистически хипотези за значимост, основаващи се на статистически критерии.

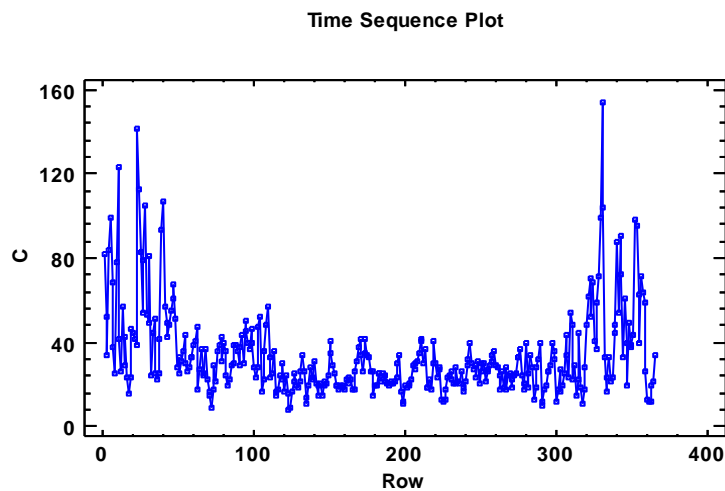
Целта на настоящата разработка е синтезиране на подходящ модел за оценка и контрол на прахови частици във въздуха, като се подберат метеорологичните явления, оказващи най-силно влияние върху процеса на разпространение на РМ<sub>10</sub> в откритите градски зони. За извършване на необходимите анализи и изчисления е използван статистически програмен пакет за обработка на данни **STATGRAPHICS**.

## 2. ИЗСЛЕДВАНА ОБЛАСТ И ИЗПОЛЗВАНИ ДАННИ

Град София е разположен в Югозападна България, в централната част на Софийското котловинно поле, на територия от 1311 km<sup>2</sup>. Градът е заобиколен от няколко планини - Витоша, Люлин, Лозен, Средна гора и Стара планина. Релефът на земната повърхност е разнообразен и се разделя на следните геоморфоложки единици: котловинно дъно, планински подножия и склонове на прилежащите планински системи. Софийското поле е затворено, с лоша вентилация. Разположението на града има решаващо значение за климатичните особености на региона. Образуването на мъгли, температурните инверсии, безветрието и слабото разсейване на замърсителите са често срещани явления на територията на столицата. Климатичните условия в комбинация с големия брой антропогенни източници на замърсяване на въздуха са причина за рязкото повишаване на концентрациите им за кратък период.

За синтеза на математически модел се използват данните за стойностите на РМ<sub>10</sub>, измерени на територията на квартал „Младост”, София за период от една година, от 01.01.2017 г. до 31.12.2017 г. Необходимите данни са измерени с автоматична измервателна станция Thermo Sharp 5030. Избрана е измервателната станция в квартал „Младост”, тъй като това е единствената АИС на територията на Столична община, за която високите нива на замърсяване с РМ<sub>10</sub> се формират предимно от градския трафик. Данните са предоставени от Изпълнителна агенция по околна среда.

Анализирани са 364 стойности за  $PM_{10}$  (една стойност е цензурирана) за периода от 01.01.2017 г. до 31.12.2017 г. Експерименталните данни, вариращи в интервала от  $7.89 \mu\text{g}/\text{m}^3$  до  $153.76 \mu\text{g}/\text{m}^3$ , са представени на фиг. 1.



Фиг.1. Средноденонощен ход на експерименталните данни за  $PM_{10}$

### 3. СИНТЕЗ НА ЛИНЕЕН МНОЖЕСТВЕН РЕГРЕСИОНЕН МОДЕЛ

Линейният множествен регресионен модел е разработен с цел разкриване на взаимовръзка между  $PM_{10}$  и пет независими променливи: средноденонощна температура на въздуха, средноденонощна слънчева радиация, скорост на вятъра, посока на вятъра, средноденонощно атмосферно налягане (означени съответно с T, R, S, D и P). Наличието на променлива C (с C е означена величината  $PM_{10}$ ), зависеща от множество независими променливи, определя използването на множествена линейна регресия. Общият вид на тази зависимост се представя с уравнението:

$$C = a_0 + a_1 T + a_2 R + a_3 S + a_4 D + a_5 P \quad (1)$$

Всяка стойност на зависимата  $C_i$  може да бъде получена като сума от стойността, зададена от модела  $f_i$  и съответния остатък  $e_i$ . Коефициентите  $a_0$ ,  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$  и  $a_5$  трябва да бъдат така определени, че векторът с остатъците  $e_i$  да има минимална стойност.

След направени изчисления е избрана трансформацията  $\text{LOG}(C)$  като най-подходяща такава за стойностите на C. Уравнението на синтезираният множествен линейен регресионен модел, описващ взаимовръзката между  $PM_{10}$  и петте независими променливи добива следния вид:

$$\begin{aligned} \text{LOG}(C) = & -0.0034 * T - 0.1312 * \text{LOG}(R) + 1.2022 * (1/S) - 0.1672 * \text{LOG}(D) \\ & + 0.0000047 * P^2 \end{aligned} \quad (2)$$

В таблици 1 и 2 са представени резултатите от изчисленията за P-Value за факторните променливи и показателите за отклонението между модела и остатъците. Както може да се види получените стойности са ниски, с изключение на тази за средноденонощната температура на въздуха T. Това е предпоставка, според която е възможно да се търси и опростяване на модела.

**Таблица 1. ANOVA характеристики за независимите променливи**

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
T	-0,00342006	0,00383364	-0,892119	0,3729
LOG(R)	-0,131164	0,0406289	-3,22834	0,0014
(1/S)	1,20217	0,118726	10,1256	0,0000
LOG(D)	-0,167195	0,0541342	-3,08853	0,0022
P^2	0,00000472586	4,05049E-7	11,6674	0,0000

**Таблица 2. Сравнение на модела с остатъците**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	4200,42	5	840,083	5305,34	0,0000
Residual	57,0049	360	0,158347		
Total	4257,42	365			

R-squared = 98,661 percent R-squared (adjusted for d.f.) = 98,6462 percent

Standard Error of Est. = 0,397928

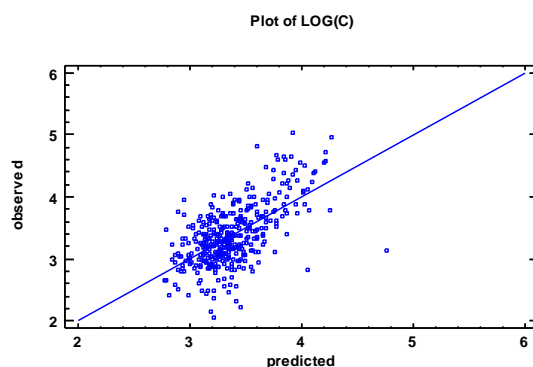
Mean absolute error = 0,304431

Durbin-Watson statistic = 0,894187

Lag 1 residual autocorrelation = 0,551824

Статистиката R-Squared показва, че моделът описва 98,661% от данните на петте независими променливи, което е много висока стойност. Стандартната грешка, показваща стандартното отклонение на остатъците е Standard Error of Est. = 0,397928, което може да се използва за изграждане на граници на прогнозиране за нови наблюдения. Средната абсолютна грешка (Mean absolute error) = 0,304431, представлява средната стойност на остатъците. Статистиката на Дърбин-Уотсън (Durbin-Watson statistic) тества остатъците за определяне на значима корелация на база на последователността, в която те се появяват в данните. Стойностите на тази статистика трябва да попаднат в интервала между 0 и 4. Стойности по-малки от 2 показват положителна автокорелация, а стойности по-големи от 2 показват отрицателна автокорелация между зависимата и независимите променливи.

На фиг. 2 в графичен вид е представено съответствието между измерените стойности на зависимата променлива LOG(C) и изчислените, определени от правата линия. Вижда се отчетливо групиране на данните с тези предсказани от модела. В таблица 3 са представени изчислените характеристики ANOVA на синтезирания регресионен модел.



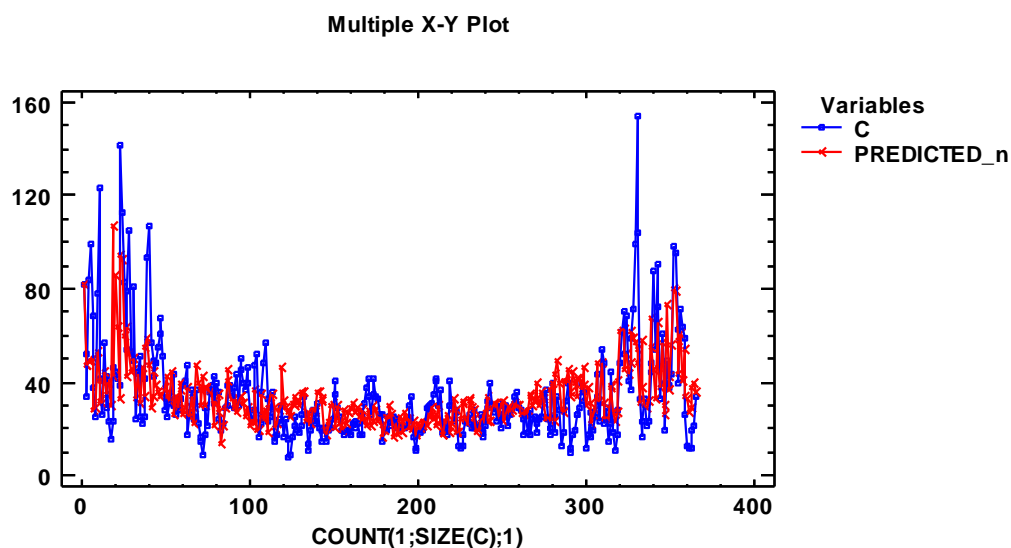
**Фиг. 2. Връзка между изчислени и измерени стойности за LOG(C)**

**Таблица 3. ANOVA характеристики за независимите променливи**

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
T	2484,84	1	2484,84	15692,36	0,0000
LOG(R)	1605,85	1	1605,85	10141,34	0,0000
(1/S)	61,7496	1	61,7496	389,96	0,0000
LOG(D)	26,427	1	26,427	166,89	0,0000
P^2	21,5554	1	21,5554	136,13	0,0000
Model	4200,42	5			

Таблицата показва значимостта на независимите променливи. Съгласно получените стойности в нея, може да се определи, че всяка една от тях е статистически значима за модела.

На базата на разработеният модел е направено сравнение между предсказаните стойности от регресионното уравнение и експерименталните данни. На фиг. 3. може да се добие визуална представа за степента на близост между тях.



**Фиг. 3. Сравнение между измерени и предсказани от модела стойности**

Проведен е анализ на вероятностното разпределение на остатъците. Резултатите са показани в таблица 4. От таблицата ясно може да се определи, че остатъците могат да се апроксимират с нормално разпределение с 95% вероятност. Това доказва адекватността на предложения линеен множествен регресионен модел.

**Таблица 4. Тест за нормалност на остатъците**

Test	Statistic	P-Value
Chi-Square	41,9041	0,266489
Shapiro-Wilk W	0,98731	0,788394
Skewness Z-score	1,21808	0,223194
Kurtosis Z-score	2,91682	0,00353631

#### 4. ЗАКЛЮЧЕНИЕ

Разработен е линеен множествен регресионен модел за разкриване на причинно-следствена връзка между  $PM_{10}$  и петте променливи средноденонощна температура на въздуха, средноденонощна слънчева радиация, скорост на вятъра, посока на вятъра, средноденонощно атмосферно налягане. По метода на най-малките квадрати е определена функционалната зависимост между зависимата величина и предикторите. Получената висока стойност на статистиката  $R^2 = 98,661\%$  показва много силна причинно-следствена връзка между концентрацията на прахови частици и независимите променливи. Направено е сравнение между предказаните стойности на  $PM$  от регресионното уравнение и експерименталните данни, показваща високата степен на близост между тях. Изчислени са ANOVA характеристики на синтезирания регресионен модел, които показват, че всеки един от независимите променливи е статистически значим за модела.

#### ЛИТЕРАТУРА

- [1] Liping Xia, Yaping Shao, Modelling of traffic flow and air pollution emission with application to Hong Kong Island, Environmental Modelling & Software 20, pp. 1175–1188, 2005
- [2] Maya 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

**Автори:** *Николай Стоянов*, доц. д-р, Технически Университет – София, Факултет Автоматика, катедра „Електроизмервателна техника”,  
*e-mail: [n\\_stoyanov@tu-sofia.bg](mailto:n_stoyanov@tu-sofia.bg)*

*Антония Панделова*, гл. ас. д-р, Технически Университет – София, Факултет Автоматика, катедра „Електроизмервателна техника”,  
*e-mail: [apandelova@tu-sofia.bg](mailto:apandelova@tu-sofia.bg)*

*Цанко Георгиев*, ас., Технически Университет – София, Факултет Автоматика, катедра „Автоматизация на непрекъснатите производства“,  
*e-mail: [tzg@tu-sofia.bg](mailto:tzg@tu-sofia.bg)*

*Юлия Калъпчийска*, редовен докторант, Технически Университет – София, Факултет Автоматика, катедра „Електроизмервателна техника”,  
*e-mail: [kalapchiiska@abv.bg](mailto:kalapchiiska@abv.bg)*

**Authors:** *Assoc. prof. Dr. Nikolay Stoyanov*, Technical University of Sofia, Faculty of Automatics, dept. Electrical Measurement Systems  
*e-mail: [n\\_stoyanov@tu-sofia.bg](mailto:n_stoyanov@tu-sofia.bg)*

*Assist prof. Dr. Antonia Pandelova*, Technical University of Sofia, Faculty of Automatics, dept. Electrical Measurement Systems  
*e-mail: [apandelova@tu-sofia.bg](mailto:apandelova@tu-sofia.bg)*

*Assist prof. Dr. Tzanko Georgiev*, Technical University of Sofia, Faculty of Automatics, dept. Industrial automation  
*e-mail: [tzg@tu-sofia.bg](mailto:tzg@tu-sofia.bg)*

*Julia Kalapchiiska*, Technical University of Sofia, Faculty of Automatics, dept. Electrical Measurement Systems  
*e-mail: [kalapchiiska@abv.bg](mailto:kalapchiiska@abv.bg)*