



# Възстановяване на липсващи данни за ФПЧ<sub>10</sub> за станции в София с различни статистически методи

Надя Нейкова, Пламен Нейчев

НИМХ, департамент Метеорология, секция Приложна метеорология

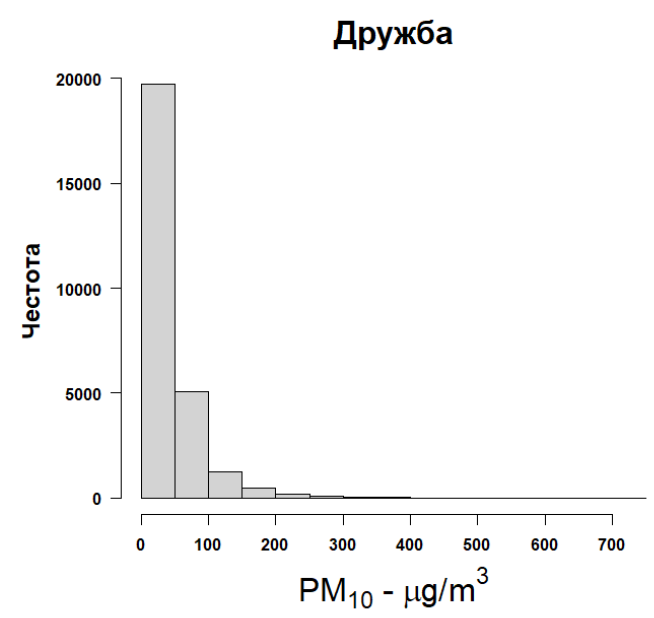
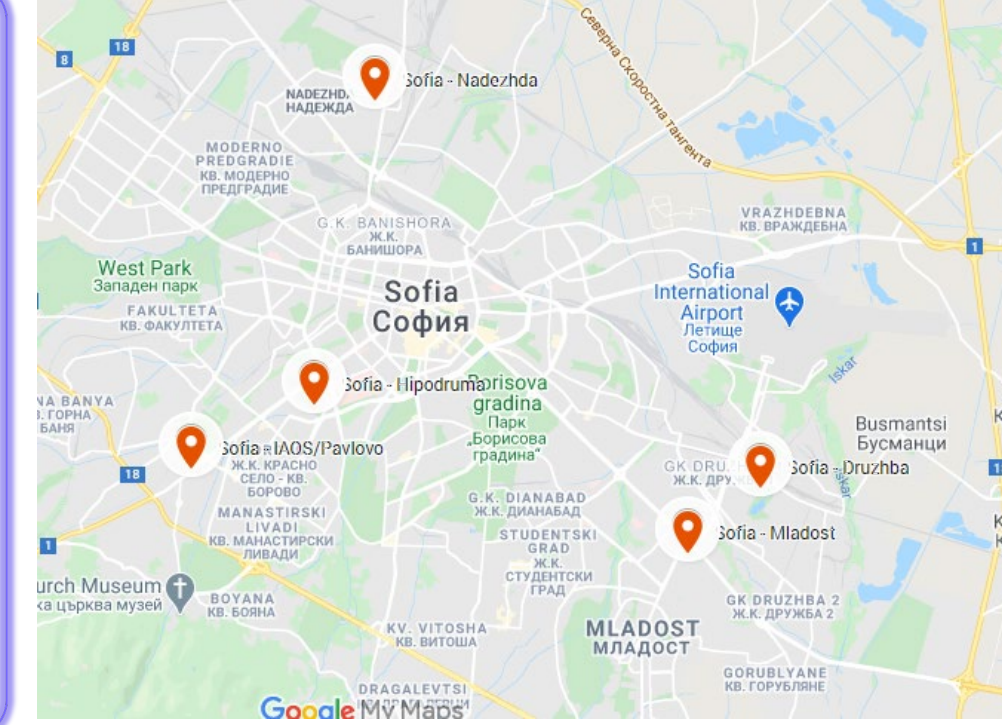
email: nadya.neykova@meteo.bg

## АБСТРАКТ

Разгледани са различни методи за възстановяване на липсващи часови стойности за ФПЧ<sub>10</sub> за 5 станции в град София. Използвани са обобщени адитивни модели (GAMs) с гама и Вейбул разпределения, както и стандартни методи за възстановяване на данни като средна стойност, последна и следваща, линейна интерполация. Предиктори в GAMs моделите са данните за ФПЧ<sub>10</sub> от останалите станции с налични данни. Резултатите показват, че методологията на GAMs осигурява високо качество на възстановените липсващи ФПЧ<sub>10</sub> стойности.

## ОПИСАНИЕ НА ДАННИТЕ

Използвани са часови данни за ФПЧ<sub>10</sub> за периода 01.01.2017 - 29.02.2020 г. от АИС в гр. София, разположени в кв. Дружба, Надежда, Хиподрума, Павлово и Младост. Времевите редици от измервани стойности за ФПЧ<sub>10</sub> от АИС се характеризират с дясно скосено разпределение.



## МЕТОДИ

### Възстановяване с „последна и следваща“ (last & next)

Този подход се използва за възстановяване на липсващи стойности както следва:

$$\hat{x}_{t_m} = \begin{cases} x_{t_{\text{последна}}}, & t_m \leq t_{\text{последна}} + \frac{t_{\text{следваща}} - t_{\text{последна}}}{2} \\ x_{t_{\text{следваща}}}, & t_m > t_{\text{последна}} + \frac{t_{\text{следваща}} - t_{\text{последна}}}{2} \end{cases}$$

където  $t_m$  е моментът на възстановяване на съответната липсваща стойност, удовлетворяващ условието  $t_{\text{последна}} < t_m < t_{\text{следваща}}$ ,  $t_{\text{последна}}$  и  $t_{\text{следваща}}$  са съответните моменти на последната и следващата налични стойности, ограждащи последователността от липсващи стойности.

### Възстановяване с линейна интерполация

Този подход се използва за възстановяване на последователност от липсващи стойности като се построява права линия през последната  $x_{t_1}$  и първата  $x_{t_2}$  измерени стойности, ограждащи последователността от липсващи стойности в моментите  $t_1$  и  $t_2$ :

$$\hat{x}_{t_m} = x_{t_1} + \theta(t_m - t_1), \text{ където } \theta = \frac{x_{t_2} - x_{t_1}}{t_2 - t_1} \text{ и } t_1 < t_m < t_2.$$

### Възстановяване с GAMs регресионни модели с гама и Вейбул разпределения

Използвани са плътностите на гама и Вейбул разпределенията:

$$g(y; a, b) = \begin{cases} \frac{b^a y^{(a-1)} \exp(-by)}{\Gamma(a)}, & y > 0 \\ 0, & y = 0, \end{cases}$$

$$w(y; a, b) = \begin{cases} \frac{a y^{(a-1)} \exp(-(y/b)^a)}{b^a}, & y > 0 \\ 0, & y = 0, \end{cases}$$

където  $a > 0$  и  $b > 0$  са неизвестни константи (коэффициенти).

GAMs регресионните модели се дефинират както следва:

$$a_t = \exp\left(\alpha_0 + \alpha_1 x_1(t) + \dots + \alpha_p x_p(t) + s_1(x_{p+1}(t)) + \dots + s_u(x_{p+u}(t))\right)$$

$$b_t = \exp\left(\beta_0 + \beta_1 z_1(t) + \dots + \beta_q z_q(t) + s_1(z_{q+1}(t)) + \dots + s_v(z_{q+v}(t))\right)$$

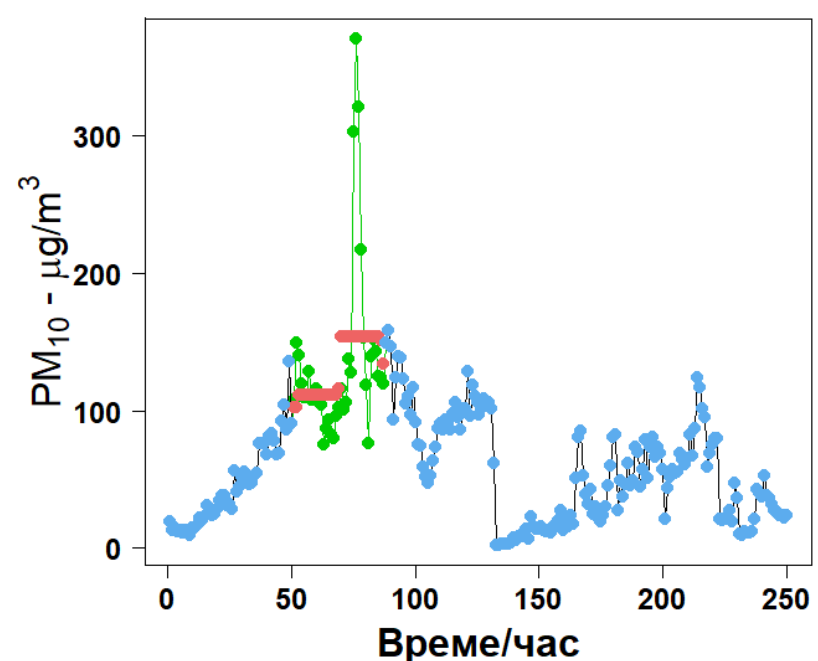
$$s_l(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \sum_{l=1}^k \theta_j (x - \tau_l)_+^3,$$

$x_1(t), \dots, x_p(t), z_1(t), \dots, z_q(t)$  са предиктори.  $a_t$  и  $b_t$  са неизвестни коэффициенти на гама и Вейбул разпределенията в момента  $t$ ,  $\alpha_0, \alpha_1, \dots, \alpha_p, \beta_0, \beta_1, \dots, \beta_q, \theta_0, \dots, \theta_k$  са неизвестни регресионни коэффициенти, а  $s_1, s_2, \dots$  са сплайн функции, които се оценяват по метода на максималното правдоподобие:

$$\max_{\alpha, \beta, \theta} L(\alpha, \beta, \theta) = \max_{\alpha, \beta, \theta} \prod_{i=1}^T f(y(t), a_t, b_t)$$

## РЕЗУЛТАТИ

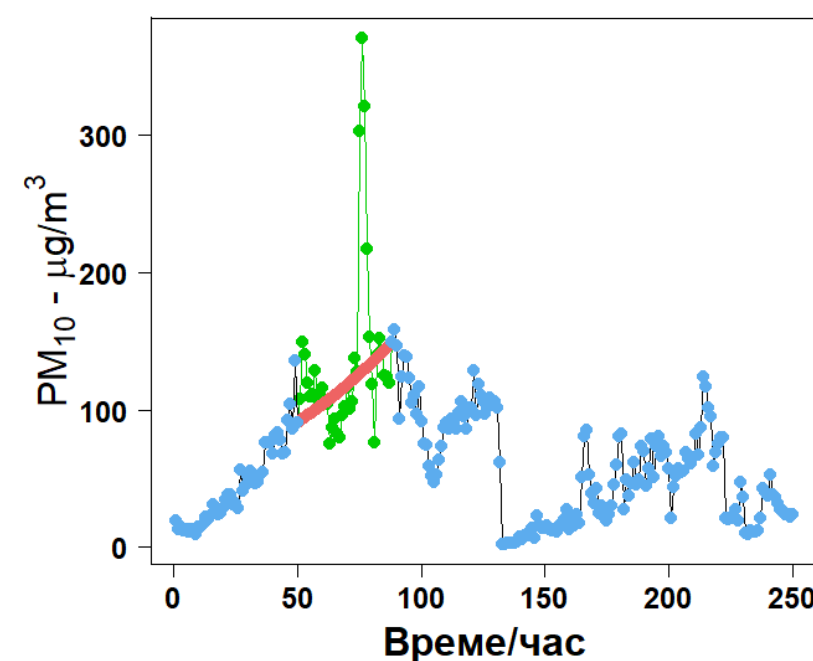
Дружба - възстановени стойности



• възстановени • измерени • измерени

а) „Последна и следваща“

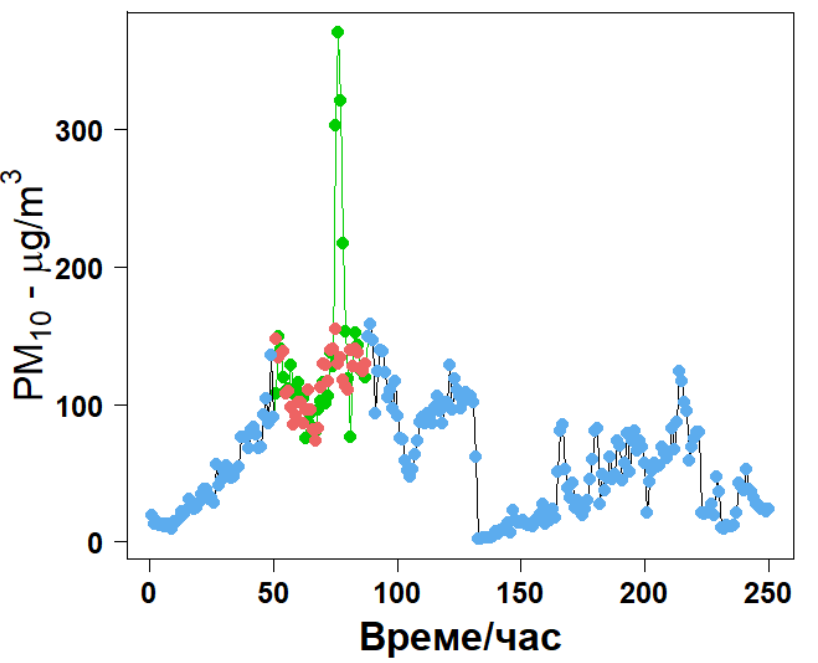
Дружба - възстановени стойности



• възстановени • измерени • измерени

б) Линейна интерполация

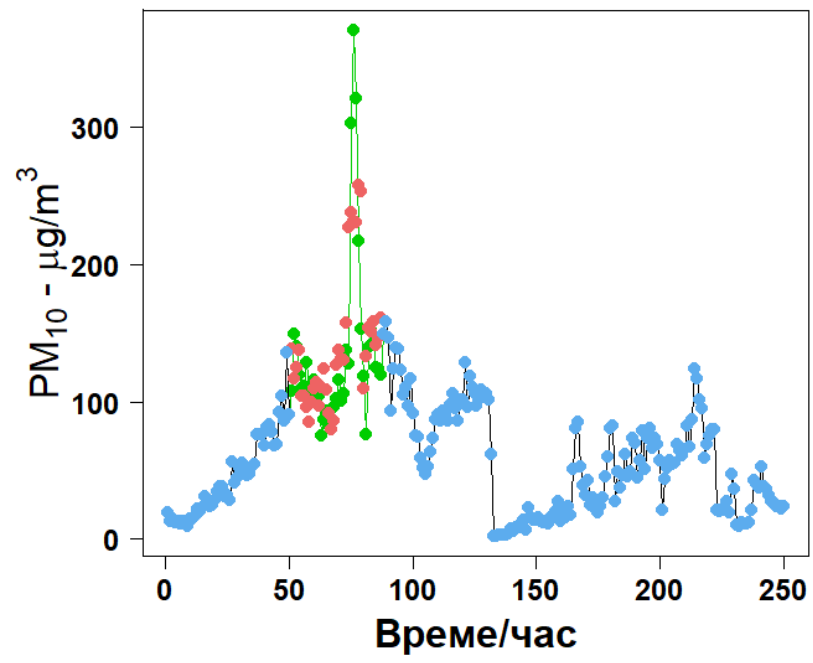
Дружба - възстановени стойности



• възстановени • измерени • измерени

в) GAMs модел с гама

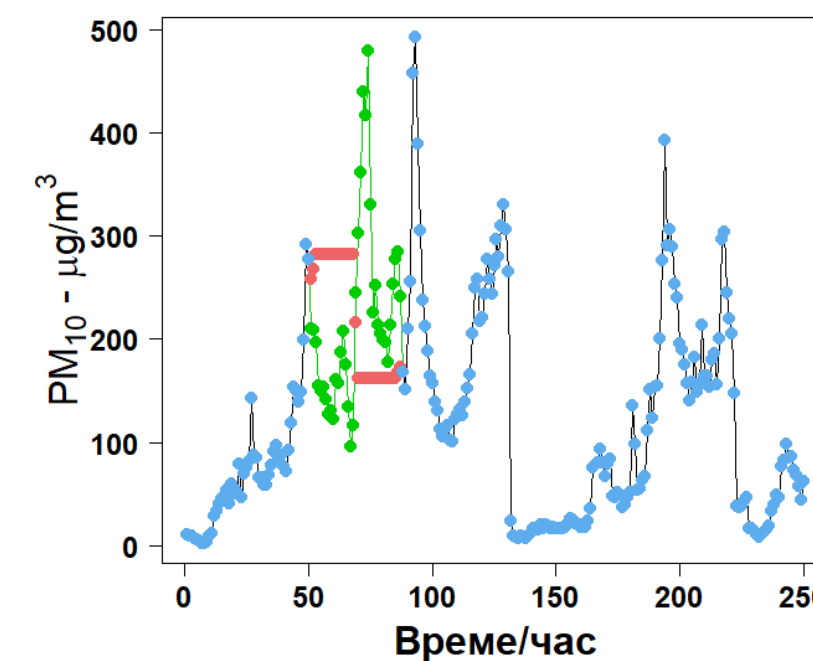
Дружба - възстановени стойности



• възстановени • измерени • измерени

д) GAMs модел с Вейбул

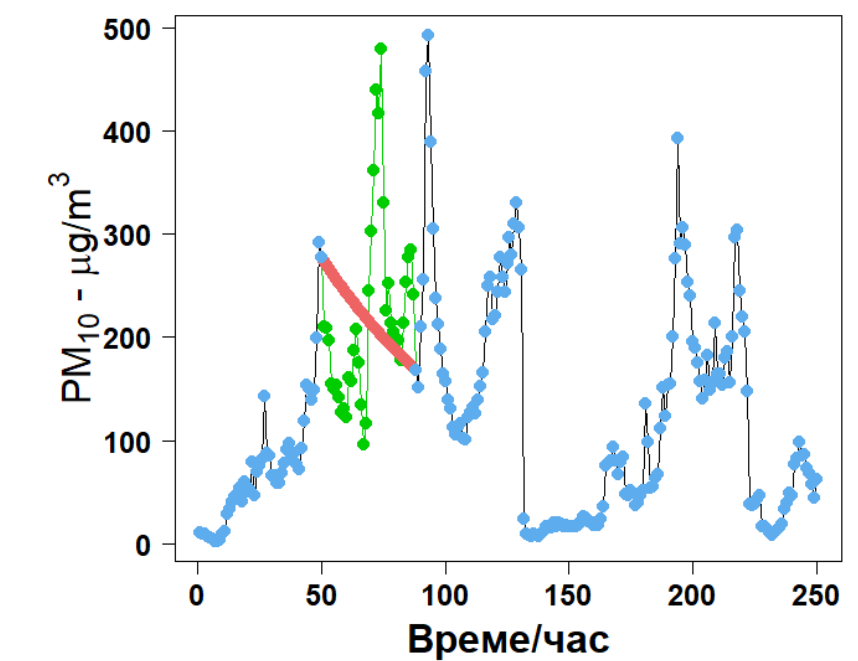
Надежда - възстановени стойности



• възстановени • измерени • измерени

а) „Последна и следваща“

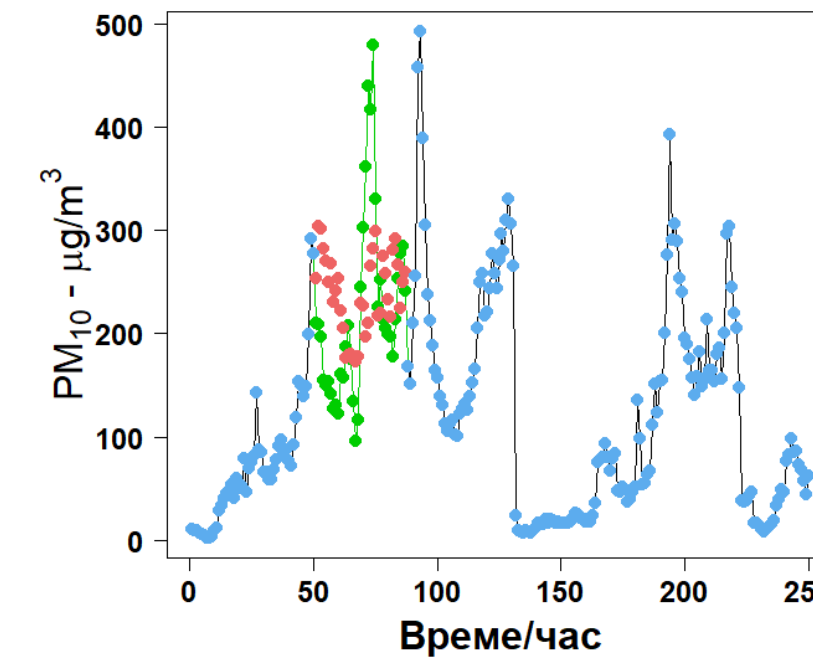
Надежда - възстановени стойности



• възстановени • измерени • измерени

б) Линейна интерполация

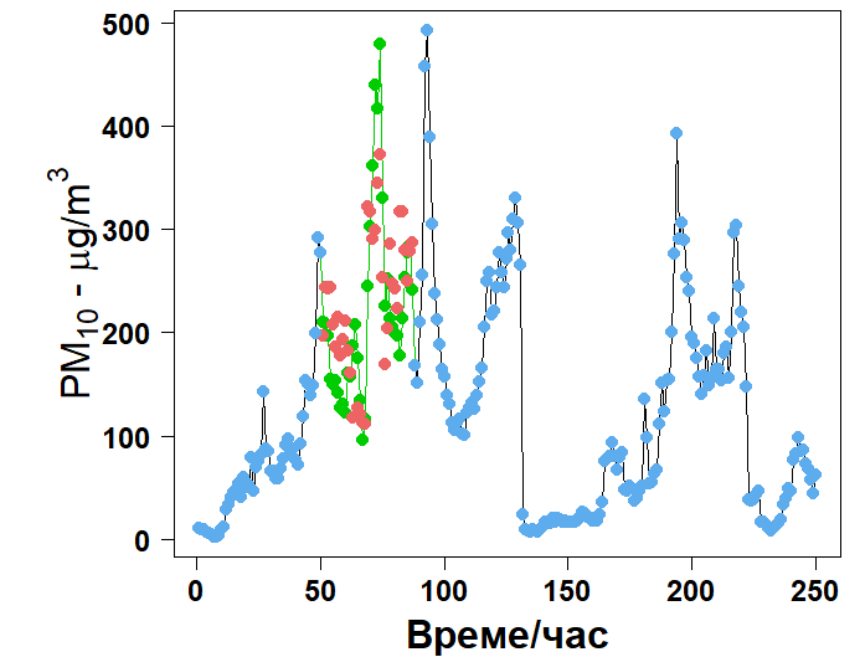
Надежда - възстановени стойности



• възстановени • измерени • измерени

в) GAMs модел с гама

Надежда - възстановени стойности



• възстановени • измерени • измерени

д) GAMs модел с Вейбул

## ЗАКЛЮЧЕНИЕ

Чрез GAMs методологията успешно се възстановяват по-дълги последователности от липсващите ФПЧ<sub>10</sub> и се подобрява качеството на данните. Възстановяването с „последна и следваща“ и линейна интерполация са лесни за използване и са подходящи при последователности от липсващи стойности в кратки поредни срокове.

## ЛИТЕРАТУРА

Нейкова, Н., Нейчев, П. (2020). Възстановяване на липсващи часови данни за ФПЧ<sub>10</sub> за станции в София. Bulgarian Journal of Meteorology and Hydrology (BJMH), 24(1), 18-39.  
Hastie, T.J. and Tibshirani, R.J. (1990), Generalized Additive Models. Boca Raton, Florida: Chapman & Hall/CRC.

**БЛАГОДАРНОСТИ:** Това изследване е финансирано от Национална програма „Млади учени и постдокторанти“ на МОН, договор №: ЧР-04-14/31.03.2020 г.

Ден на отворените врати в НИМХ - 23.03.2023 г.