AIP Conference Proceedings

Markov chains modelling of particulate matter (PM10) air contamination in the city of Ruse, Bulgaria

Cite as: AIP Conference Proceedings **2302**, 060018 (2020); https://doi.org/10.1063/5.0033630
Published Online: 03 December 2020

E. Veleva, I. R. Georgiev, I. Zheleva, and M. Filipova

Markov Chains Modelling of Particulate Matter (PM10) Air Contamination in the City of Ruse, Bulgaria

E. Veleva[1,a)], I.R. Georgiev[1,b)], I. Zheleva[2,c)] and M. Filipova[2,d)]

[1]*Dept of Applied Mathematics and Statistics, Ruse University, 7017 Ruse, 8 Studentska str., Bulgaria*
[2]*Dept of Thermoengineering, Hydraulics, Ecology, Ruse University, 7017 Ruse, 8 Studentska str., Bulgaria*
www.uni-ruse.bg

[a)] Corresponding author: eveleva@uni-ruse.bg
[b)] irgeorgiev@uni-ruse.bg
[c)] izheleva@uni-ruse.bg
[d)] mfilipova@uni-ruse.bg

Abstract. High levels of air pollutants PM10 are a problem of great importance for human health. During the months from April to September of the period 2010 - 2019 the levels in Ruse remain within the norm in 95% of the days. During the other, "cold" months of the year, only 58% of the days have values below the daily norm of 50µg/m^3. When planning their activities, it is useful for people to have forecasts for PM10 levels in the coming days. Markov chains allow such predictions to be given in a tabular form, convenient to use, without the need for calculations.

The data for the "cold" months are modelled using three Markov chains with different degrees of discretization of the original values, respectively with 12, 7 and 3 possible states. The latter, with states: {in the norm}, {slightly above the norm} and {a strong excess of the norm}, can be used without official data on the exact PM10 levels. Determining the condition of the PM10 pollutant today in this case can also be done on the basis of a personal assessment of the purity of the air over the city at the moment.

The measured levels in the period 01.01.2020 - 31.03.2020 were used as test data. They show consistency of the measured levels in 2020 with all three Markov chains considered. The obtained tabular values can be used to predict PM10 levels in the following years, in the months from October to March.

NTRODUCTION

Dust is one of the major atmospheric air pollutants. Its harmful health effect depends mainly on the size and chemical composition of the suspended dust particles. Chemical compounds can be adsorbed on the dust surface, including mutagens, DNA modulators, *etc*. Fine dust particles are emitted into the atmosphere directly (primary emissions) or are formed by gases emitted into the atmosphere - precursors of fine dust particles (secondary emissions). Sulphur dioxide, nitrogen oxides and ammonia are inorganic gaseous substances, precursors of fine dust particles. The dust enters the human body mainly through the respiratory system, where larger particles are retained in the upper respiratory tract, and finer particles reach the lower parts of the respiratory system, leading to its damage. Adults, children, and people with chronic lung disease, asthma or influenza are particularly sensitive to high levels of dust in the air. The harmful effect of dust pollution is more pronounced in the simultaneous presence of sulfur dioxide in the air. Their synergistic effect on the respiratory organs and exposed mucous membranes has been established [1]. The main sources of dust in the atmosphere are industry, transport, energy and domestic heating. During the heating season at the local level the main source of pollution with dust particles is the combustion of solid and liquid fuels in the home. A spatial attention has to be paid to the possibilities for air contamination by industrial units spatially in the state border area [2, 3].

FIGURE 1. Ruse district, Bulgaria

Ruse is the largest Bulgarian city along the Danube river and the fifth largest city in Bulgaria. It is located on the northeastern border of the country with Romania. The city is the administrative center of the municipality of Ruse and the district of Ruse, as well as an economic, transport, cultural and educational centre of regional and national importance. 310km from Sofia and 75km from the Romanian capital Bucharest, the city is a potential strategic intermodal and logistics center of the country and Southeast Europe.

Application of Mathematics in Technical and Natural Sciences
AIP Conf. Proc. 2302, 060018-1–060018-9; https://doi.org/10.1063/5.0033630
Published by AIP Publishing. 978-0-7354-4036-4/$30.00

Ruse has got established role as an important industrial centre. Here is the first in the country road and railway bridge over the river Danube. The river also connects the city to the European transport corridor no 7, creating conditions for commercial and cruise river navigation.

Air pollution by particulate matter PM10 (PM10 - particulate matter with a diameter between 2.5 and 10μm) is going up recently in Ruse region, Bulgaria [4, 5, 6]. PM10 levels for Ruse mark a significant increase during the autumn-winter period compared to the levels during the spring-summer period. Also the specific meteorological conditions during the winter seasons reduce the possibility of dissipation of atmospheric pollutants [7, 8]. Thus the biggest peak of PM10 levels for the autumn-winter period is usually observed in January months. It is usually in January that the number of days with exceedance of the limit values of the PM10 levels is maximum observed compared with the other months in the year [9, 10].

There are several research papers for PM10 measurements modelling and forecasting for different regions in Bulgaria [11-15]. In these publications different statistic methods for PM10 data modelling and studying are used such as ARIMA, SARIMA, GARCH and etc. Here we will use another method – Markov chain to analyse PM10 air contamination in the city of Ruse, Bulgaria.

1. DATA DESCRIPTION

For this study official PM10 average daily concentration level measurements by Bulgarian Ministry of environment and water for the last ten years 2010 – 2019 [16] are used. The data contain 55 missing values (1.51% of all observations). Figure 2 shows the average levels of PM10 by months and the average annual rate norm of $40\mu g/m^3$ with red line. Below the norm are only the “warm” months, from April to September, while in the “cold” months, the rest six months the average levels are above the annual norm.

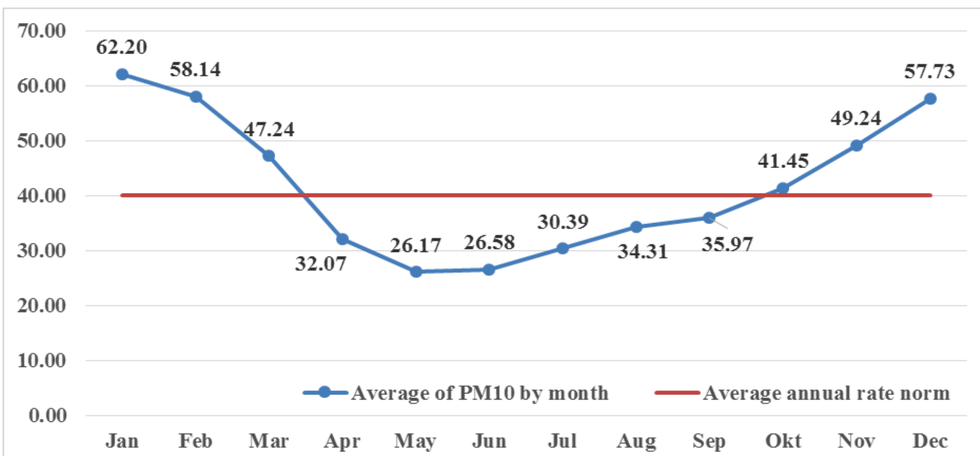

FIGURE 2. The average levels of PM10 by months

Figure 3 shows the observed levels of PM10 in the “cold” and “warm” months separately. Considered individually, the two time series are stationary. Only small number of observations in the “warm” months is above the daily norm of $50\mu g/m^3$.

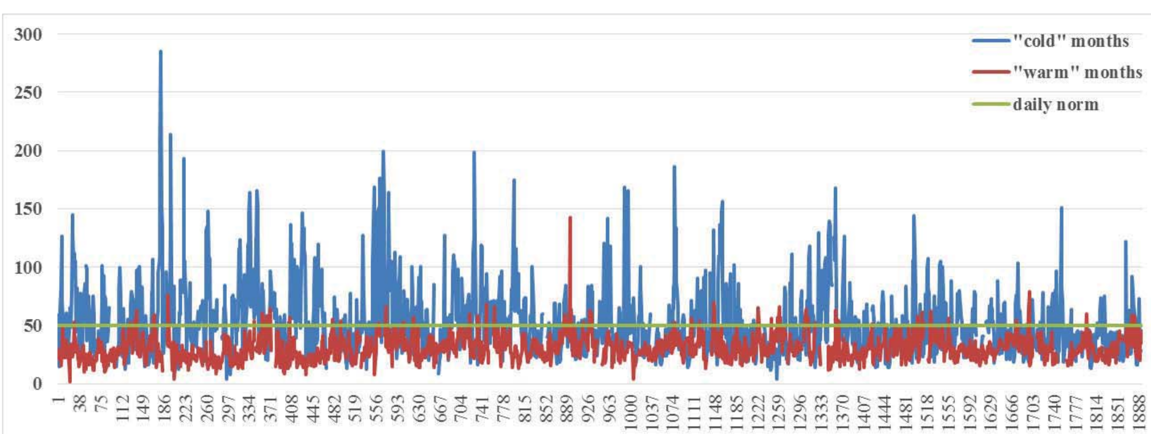

FIGURE 3. The observed levels of PM10 in the “cold” and “warm” months separately

Let us introduce the following discretization of the measured values of PM10: values below 20 μg/m^3 will be denoted by state S1; values in the interval (20, 30] as state S2; in the interval (30, 40] as state S3; in the interval (40, 50] as state S4; in the interval (50, 60] as state S5; in the interval (60, 70] as state S6; in the interval (70, 80] as state S7; in the interval (80, 90] as state S8; in the interval (90, 100] as state S9; in the interval (100, 110] as state S10; in the interval (110, 120] as state S11 and in the interval (120, ∞] as state S12. The first 4 states S1, S2, S3 and S4 include the values within the daily norm.

The distribution of the values of PM10, measured during the "cold" months, by states is given in Figure 4. Above the daily norm (states from S5 to S12) lie 41.63% of all valid observations. On Figure 5 we see the distribution by states during the "warm" months. Here, only 5.03% of valid observations are above the daily norm of 50 μg/m^3.

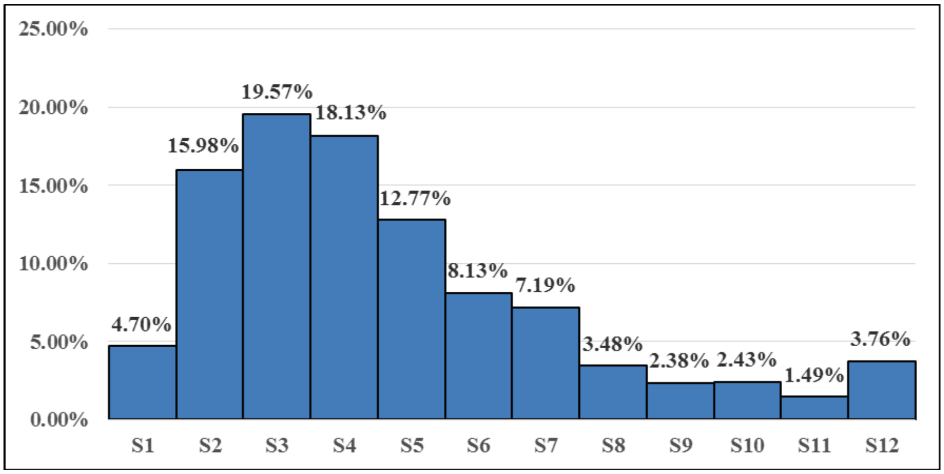

FIGURE 4. The distribution of the values of PM10, measured during the "cold" months, by states

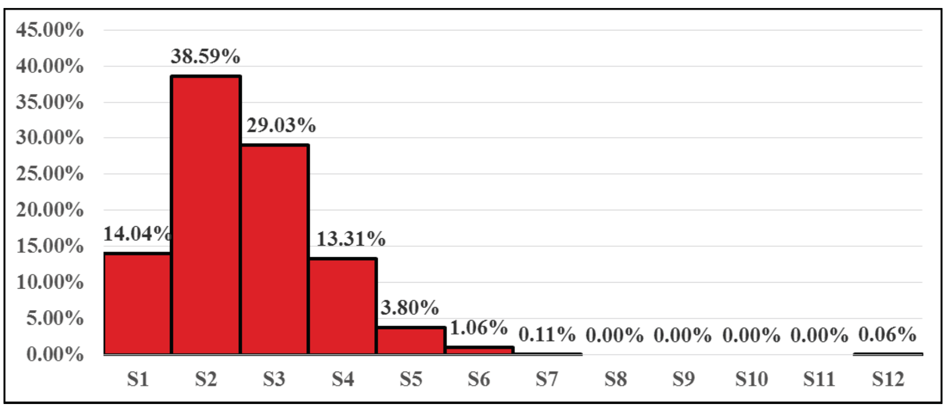

FIGURE 5. The distribution of the values of PM10, measured during the "warm" months, by states

Therefore, only the levels of PM10, measured during the "cold" months of the year, are interesting for research. Modelling using Markov chains is not sensitive to missing values in the data. When using this approach, it is important to estimate the probabilities $p_{i,j}$ of transition from state Si to state Sj, taking into account the number $n_{i,j}$ of the observations in state Si, followed by state Sj at the next moment. Then the maximum likelihood estimates for $p_{i,j}$ are (see [17,18])

$$\hat{p}_{i,j} = n_{i,j} / n_i^*, \quad n_i^* = \sum_j n_{i,j} . \tag{1}$$

Markov chains provide an opportunity to answer questions such as: If today we are in state S12 (*i.e.*, level of PM10 above 120μg/m^3), after how many days can we expect with a probability above 50% a value below 50μg/m^3 (*i.e.*, within the norm, states S1, S2, S3 or S4); If today the value of PM10 is over 70μg/m^3, what is the expected number of days before falling below the norm; If a value between 110 and 120 is measured today, what is the most probable condition for tomorrow, for two or more days; what are the future states that we can expect with a probability close to 1 and so on.

Markov chains have found considerable application in meteorology (see [19]). A stochastic process $Y = \{Y_n, n = 1, 2, \ldots\}$ with r states (S1, S2, …, Sr) is said to be a first-order Markov chain, if

$$P(Y_{n+1} = Sj \mid Y_1 = Si_1, \ldots, Y_{n-1} = Si_{n-1}, Y_n = Si) = P(Y_{n+1} = Sj \mid Y_n = Si) = p_{i,j} \qquad (2)$$

for all states Si_1, …, Si_{n-1}, Si, Sj and for $n = 1, 2, \ldots$. The condition (2) is referred to as the "Markovian property": the future state of the process depends only on the present state.

Not all time series are suitable for modelling with first-order Markov chains, not for all the "Markovian property" it is valid. Figure 6 gives the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the time series of the levels of PM10 during the "cold" months. We can see that a first-order autoregressive model of the form

$$Y_{n+1} = c + \phi_1 Y_n + \varepsilon_{n+1},$$

where ε_n is white noise, c and ϕ_1 are coefficients, would be appropriate in describing the data (see [20]). Such a model is also based on the assumption that, despite the entire previous history of the process, its future value depends only on the current one.

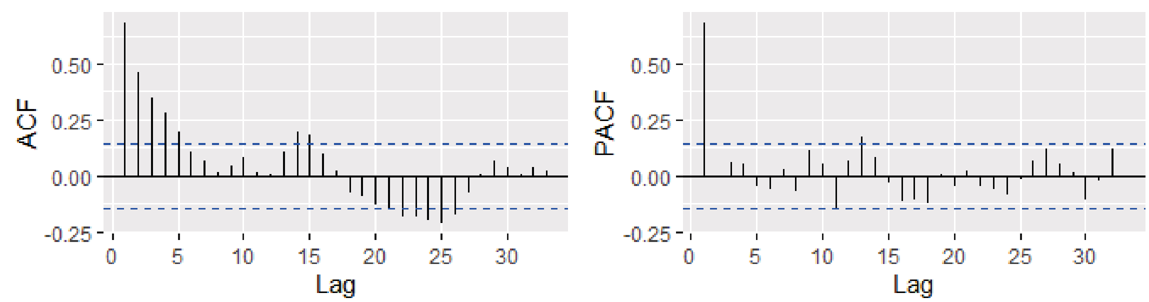

FIGURE 6. The ACF and PACF plots of the levels of PM10 during the "cold" months

2. MODELLING BY A FIRST-ORDER MARKOV CHAIN

In this section, the time series of the PM10 levels, measured in the "cold" months in the period 2010 – 2019, is modelled through Markov chains, at different number of process states, i.e. different discretization of the observed values. Calculations are done by R programming language.

2.1. Modelling Assuming 12 States

In this subsection we assume that every day the process is in one of the 12 states S1, …, S12, defined in Section 1. The probabilities $p_{i,j}$ of transition from today state Si to tomorrow state Sj, estimated by formulas (1), are given in Table 1.

TABLE 1. The estimated transition probabilities

Today \ Tomorrow	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0.2619	0.4524	0.1905	0.0357	0.0357	0.0119	0	0	0	0.0119	0	0
S2	0.0979	0.3077	0.3322	0.1503	0.0455	0.0385	0.0175	0.0105	0	0	0	0
S3	0.0283	0.2096	0.3031	0.2776	0.0850	0.0368	0.0312	0.0227	0	0.0028	0.0028	0
S4	0.0370	0.1481	0.2006	0.2346	0.1852	0.0772	0.0617	0.0247	0.0093	0.0123	0	0.0093
S5	0.0349	0.0786	0.1310	0.2358	0.2227	0.1135	0.0830	0.0306	0.0262	0.0087	0.0087	0.0262
S6	0.0069	0.0621	0.1034	0.1379	0.1862	0.1862	0.1655	0.0345	0.0552	0.0414	0.0138	0.0069
S7	0.0077	0.0308	0.0923	0.1231	0.2077	0.1308	0.1385	0.0769	0.0769	0.0692	0.0154	0.0308
S8	0	0.0339	0.0678	0.1356	0.1017	0.2034	0.1525	0.0847	0.0508	0.0339	0.0169	0.1186
S9	0.0238	0.0476	0.0952	0.0476	0.1190	0.1667	0.2143	0.0714	0.0238	0.1190	0.0238	0.0476
S10	0	0.0227	0.0455	0.0682	0.0455	0.0455	0.1364	0.0909	0.1591	0.1364	0.0909	0.1591
S11	0	0	0	0.0370	0.0741	0.1111	0.0741	0.1111	0.0370	0.0741	0.2222	0.2593
S12	0	0.0294	0.0294	0.0147	0.0441	0.0294	0.0882	0.0882	0.0588	0.0735	0.1176	0.4265

Only 4.70% of all observations are in state S1 (Figure 4), however if today we are in state S1, then the probability that tomorrow we will be again in state S1 is 0.2619 (Table 1). If today we are in state S11, then most likely, with probability 0.2593 tomorrow we will be in state S12; with probability 0.4815 (0.2222 + 0.2593) the average daily level of PM10 tomorrow will be above 110μg/m3 (state S11 or S12); with probability 0.9630 the average daily level of PM10 tomorrow will be above the norm of 50μg/m^3 (states from S5 to S12). The largest from the diagonal probabilities, is the one corresponding to the state S12, *i.e.*, it is most inert, followed by states S2 and S3.

Colored probabilities are the largest values in Table 1 whose sum exceeds close to one number γ, the same for all rows. The sum of the red values on each row exceeds 0.77, so they can be used as a 77% confidence set for the tomorrow process state provided the today state. For example, if the current state is S2, then with a probability of at least 0.77 the next will be S2, S3 or S4, *i.e.*, the average daily concentration of PM10 for the next day will be in the range (20, 50]. The addition of the blue values further expands the confidence set for the future state of the process, provided its current state and increases the confidence probability γ to at least 0.84 (84%). For example, if the current value of PM10 is in the interval (20, 30] (state S2), then with a probability of at least 0.84 tomorrow's value will be at most 50 (state S1, S2, S3 or S4), *i.e.*, in the norm.

When the transition matrix $P = (p_{i,j})$ is raised to the *n*th degree, the conditional probabilities for the state of the system after *n* days are obtained, provided the current state (see [21]). Let us divide the 12 states into three groups of four states: the first {S1, S2, S3, S4} will correspond to values of PM10 in the norm, the second {S5, S6, S7, S8} will include values in the range (50, 90], *i.e.*, moderately exceeding the norm and the last {S9, S10, S11, S12} contains values above 90 μg/m^3, *i.e.*, strongly exceeding the norm. Because these three groups include all 12 states, the sum of their probabilities is one. Table 2 shows the conditional probabilities that the average daily level of PM10 will be in the first or third group separately, respectively after one, two, three, four, five, six or seven days, depending on the current state of the process. After more than 15 days, these probabilities, up to the second decimal place, no longer depend on the current state of the process.

If the system is in the S6 state today, a fall below the norm can be expected with a probability of more than 50% from the third day onwards; if it is in the S12 state today, a fall below the norm with a probability higher than 0.50 can be expected from the seventh day onwards.

TABLE 2. The conditional probabilities for PM10 to be in the norm or strongly exceeding the norm after 1, 2, 3, 4, 5, 6, 7 or more days, depending on the current state

After / Today	1 day		2 days		3 days		4 days		5 days		6 days		7 days		> 15 days	
	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)	(0,50]	(90,∞)
S1	0.94	0.01	0.85	0.02	0.75	0.03	0.69	0.05	0.65	0.06	0.63	0.08	0.61	0.08	0.58	0.10
S2	0.89	0.00	0.77	0.02	0.70	0.04	0.65	0.06	0.63	0.07	0.61	0.08	0.60	0.09	0.58	0.10
S3	0.82	0.01	0.70	0.04	0.65	0.06	0.63	0.07	0.61	0.08	0.60	0.09	0.60	0.09	0.58	0.10
S4	0.62	0.03	0.61	0.06	0.61	0.08	0.60	0.09	0.60	0.09	0.59	0.09	0.59	0.10	0.58	0.10
S5	0.48	0.07	0.54	0.09	0.56	0.10	0.57	0.10	0.58	0.10	0.58	0.10	0.58	0.10	0.58	0.10
S6	0.31	0.12	0.45	0.13	0.51	0.12	0.54	0.12	0.56	0.11	0.57	0.11	0.57	0.10	0.58	0.10
S7	0.25	0.19	0.41	0.16	0.48	0.14	0.52	0.13	0.54	0.12	0.56	0.11	0.57	0.11	0.58	0.10
S8	0.24	0.22	0.37	0.20	0.45	0.17	0.50	0.15	0.53	0.13	0.55	0.12	0.56	0.11	0.58	0.10
S9	0.21	0.21	0.38	0.20	0.46	0.16	0.51	0.14	0.53	0.13	0.55	0.12	0.56	0.11	0.58	0.10
S10	0.14	0.55	0.26	0.33	0.37	0.24	0.44	0.19	0.49	0.16	0.52	0.14	0.54	0.12	0.58	0.10
S11	0.04	0.59	0.18	0.41	0.31	0.30	0.40	0.23	0.46	0.18	0.50	0.15	0.53	0.13	0.58	0.10
S12	0.07	0.68	0.19	0.45	0.30	0.32	0.39	0.24	0.45	0.19	0.50	0.16	0.53	0.14	0.58	0.10

The predicted probabilities for after 22 days or more, up to the fourth decimal place, regardless of the current state of the process, are given in the row 2 of Table 3. This is the unique fixed probability vector for $P = (p_{i,j})$ (see [21]). These values are close to the estimated probabilities for the states given in row 1 of the table and shown in Figure 4.

TABLE 3. The estimated probabilities for the states (row 1) and the unique fixed probability vector for *P* (row 2)

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**
1	0.0470	0.1598	0.1957	0.1813	0.1277	0.0813	0.0719	0.0348	0.0238	0.0243	0.0149	0.0376
2	0.0462	0.1594	0.1965	0.1818	0.1283	0.0821	0.0723	0.0346	0.0239	0.0239	0.0148	0.0361

We use the values of PM10, measured in the period 01.01.2020 – 31.03.2020 as test data for the considered Markov chains in this and the following subsections. The data have one missing value. Using a chi-square based test, described in [18], we first check whether the empirical transition matrix for the test data is statistically compatible with the transition matrix in Table 1. The chi-square statistic is 79.14126, the degrees of freedom are

114 and the corresponding p-value is 0.9946464. Therefore, the described Markov chain can be used to model future levels of PM10. When using the 77% confidence sets, shown in Table 1 with red color, to forecast the values in 2020 one step ahead we guess correctly in 78.41% of cases. When using the 84% confidence sets, shown in Table 1 with red and blue colors on each row, the forecast contains the actual observed state in 86.36% of cases.

2.2. Modelling Assuming 7 States

Looking at the forecasts for 1 day in Table 2, we notice that some consecutive rows have similar values. Therefore, we could make the discretization rougher by combining the corresponding consecutive states. Thus we will get a Markov chain with 7 states: {S1, S2}, S3, S4, S5, S6, {S7, S8, S9}, {S10, S11, S12}. Using for it the chi-square test developed in [18] for checking the "Markovian property" (2), we obtain: Chi - square statistic is 236.3615; degrees of freedom are 254 and corresponding p-value is 0.7798561. Therefore we can use a first-order Markov chain to model the levels of PM10, falling into these 7 states. The transition probabilities, estimated according to (1) are given in Table 4.

TABLE 4. The estimated transition probabilities for the 7 states Markov chain

Today \ Tomorrow	**(0, 30]**	**(30, 40]**	**(40, 50]**	**(50, 60]**	**(60, 70]**	**(70, 100]**	**(100, ∞)**
(0, 30]	0.4757	0.3000	0.1243	0.0432	0.0324	0.0216	0.0027
(30, 40]	0.2380	0.3031	0.2776	0.0850	0.0368	0.0538	0.0057
(40, 50]	0.1852	0.2006	0.2346	0.1852	0.0772	0.0957	0.0216
(50, 60]	0.1135	0.1310	0.2358	0.2227	0.1135	0.1397	0.0437
(60, 70]	0.0690	0.1034	0.1379	0.1862	0.1862	0.2552	0.0621
(70, 100]	0.0433	0.0866	0.1126	0.1645	0.1558	0.2944	0.1429
(100, ∞)	0.0216	0.0288	0.0360	0.0504	0.0504	0.2806	0.5324

The last, seventh state is the most inert, followed by the first. If the value of PM10 is above 100μg/m^3 today, the probability that a value above 100μg/m^3 will be observed again tomorrow is 0.5324.

The sum of the red values on each row exceeds 0.76. They can be used as a 76% confidence set for the tomorrow process state provided the today state. For example, if the current value is above 100μg/m^3, then with a probability of at least 0.76 tomorrow we will have a value above 70μg/m^3 (the last two states).

The addition of blue values further expands the confidence set for the future state of the process, provided its current state and increases the confidence probability to at least 0.84 (84%). For example, if the current value is above 100μg/m^3, then with a probability of at least 0.84 tomorrow we will have a value above 60μg/m^3 (the last three states); if the current value is below or equal to 30 μg/m^3, then with a probability of at least 0.84 tomorrow we will have a value within the norm (the first three states).

Table 5 shows the conditional probabilities for PM10 to be in the norm or more than 40% above the norm after 1, 2, 3, 4, 5, 6, 7 or more days, depending on the current state. After more than 14 days, these probabilities, up to the second decimal place, no longer depend on the current state of the process. If the current value is above 100μg/m^3, a fall below the norm can be expected with a probability of more than 50% from the sixth day onwards; if the current value is between 50 and 60μg/m^3, a fall below the norm with a probability higher than 0.50 can be expected from the second day onwards.

Probability predictions for after 2, 3, and so on days for each of the states can be obtained by raising the transition probability matrix given in Table 4 to the power 2, 3, and so on.

TABLE 5. The conditional probabilities for PM10 to be in the norm or more than 40% above the norm after 1, 2, 3, 4, 5, 6, 7 or more days, depending on the current state

Today \ After	**1 day**		**2 days**		**3 days**		**4 days**		**5 days**		**6 days**		**7 days**		**> 14 days**	
	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)
(0, 30]	0.90	0.02	0.79	0.07	0.71	0.12	0.66	0.15	0.63	0.17	0.61	0.18	0.60	0.19	0.58	0.21
(30, 40]	0.82	0.06	0.70	0.11	0.65	0.15	0.63	0.17	0.61	0.18	0.60	0.19	0.60	0.20	0.58	0.21
(40, 50]	0.62	0.12	0.61	0.16	0.61	0.18	0.60	0.19	0.59	0.20	0.59	0.20	0.59	0.20	0.58	0.21
(50, 60]	0.48	0.18	0.54	0.21	0.56	0.21	0.57	0.21	0.58	0.21	0.58	0.21	0.58	0.21	0.58	0.21
(60, 70]	0.31	0.32	0.45	0.28	0.51	0.25	0.54	0.24	0.56	0.23	0.57	0.22	0.57	0.21	0.58	0.21
(70,100]	0.24	0.44	0.39	0.34	0.47	0.29	0.51	0.26	0.54	0.24	0.56	0.23	0.57	0.22	0.58	0.21
(100, ∞)	0.09	0.81	0.22	0.59	0.33	0.44	0.42	0.36	0.48	0.30	0.51	0.27	0.54	0.24	0.58	0.21

The transition probabilities in Table 4 are estimated using the data from 2010 till 2019. Analogous to the previous subsection, we can test the hypothesis of consistency of these probabilities with the empirical transition probabilities, calculated on the basis of test data in 2020. The chi-square statistic is 39.99571, the degrees of freedom are 42 and the corresponding p-value is 0.5592833. Therefore, the hypothesis does not contradict the experimental data and there are no significant differences in the behaviour of the process in 2020. When using the 76% confidence sets, shown in Table 4 with red color, to forecast the values in 2020 one step ahead we guess correctly in 80.68% of cases. When using the 84% confidence sets, shown in Table 4 with red and blue colors on each row, the forecast contains the actual observed state in 93.18% of cases.

2.3. Modelling Assuming Only 3 States

In the days when information on the exact meaning of PM10 is not available, it is still possible to make a forecast for the next few days, based on a very simple Markov chain, having only three states that we could determine without measuring instruments, based on personal judgment only: {in the norm, i.e. PM10 $\in$ (0, 50]}, {slightly above the norm, PM10 $\in$ (50, 70]} and {a strong excess of the norm, PM10 $\in$ (70, ∞)}.

This Markov chain has the "Markovian property" (2): chi - square statistic is 34.17224, degrees of freedom are 58 and corresponding p-value is 0.9946752. The estimated transition probabilities are given in Table 6.

TABLE 6. The estimated transition probabilities for the 3 states Markov chain

Today \ Tomorrow	**(0, 50]**	**(50, 70]**	**(70, ∞)**
(0, 50]	0.7861	0.1490	0.0649
(50, 70]	0.4144	0.3503	0.2353
(70, ∞)	0.1838	0.2378	0.5784

If we have a strong excess of the norm today, the probability that a strong excess of the norm will be observed again tomorrow is 0.5784; if the value of PM10 today is in the norm, with probability 0.7861 tomorrow we will have again a value within the norm.

The sum of the red values on each row exceeds 0.76. They can be used as a 76% confidence set for the tomorrow process state provided the today state. For example, if the current value is above 70μg/m^3, then with a probability of at least 0.76 tomorrow we will have a value above 50μg/m^3 (the last two states), *i.e.*, above the norm.

Table 7 shows the conditional probabilities for PM10 to be in the norm or well above the norm after 1, 2, 3, 4, 5, 6, 7 or more days, depending on the current state. After more than 9 days, these probabilities, up to the second decimal place, no longer depend on the current state of the process. If the current value is strongly exceeding the norm, a fall below the norm can be expected with a probability of more than 50% from the fourth day onwards; if the current value is between 50 and 70μg/m^3, a fall below the norm with a probability higher than 0.50 can be expected from the second day onwards.

TABLE 7. The conditional probabilities for PM10 to be in the norm or more than 40% above the norm after 1, 2, 3, 4, 5, 6, 7 or more days, depending on the current state

Today \ After	**1 day**		**2 days**		**3 days**		**4 days**		**5 days**		**6 days**		**7 days**		**> 9 days**	
	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)	(0,50]	(70,∞)
(0, 50]	0.79	0.06	0.69	0.12	0.64	0.16	0.62	0.18	0.60	0.19	0.59	0.20	0.59	0.20	0.58	0.21
(50, 70]	0.41	0.24	0.51	0.25	0.55	0.23	0.56	0.22	0.57	0.22	0.58	0.21	0.58	0.21	0.58	0.21
(70, ∞)	0.18	0.58	0.35	0.40	0.45	0.31	0.51	0.27	0.54	0.24	0.56	0.23	0.57	0.22	0.58	0.21

Table 8 gives the expected number of days to reach for the first time any of the states, starting on a given state (see [21]). If today the value of PM10 is within the norm, then: the expected number of days to observe a value strongly exceeding the norm is 10.3 days; the expected number of days to obtain again a value within the norm is 1.7 days.

TABLE 8. The expected number of days to reach for the first time any of the states, starting on a given state

start \ reach	**(0, 50]**	**(50, 70]**	**(70, ∞)**
(0, 50]	1.7	6.2	10.3
(50, 70]	3.0	4.8	8.1
(70, ∞)	4.1	5.1	4.8

There are no significant differences in the behaviour of the process in 2020: χ^2-statistic is 3.099558, the degrees of freedom are 6 and the corresponding p-value is 0.7962516. When using the 76% confidence sets, shown in Table 6 with red color, to forecast the values in 2020 one step ahead we guess correctly in 80.68% of cases.

CONCLUSIONS

High levels of air pollutants PM10 are a problem of great importance for human health. During the months from April to September of the period 2010 - 2019 the levels remain within the norm in 95% of the days. During the other, “cold” months of the year, only 58% of the days have values below the norm 50 $\mu g/m^3$. When planning their activities, it is useful for people to have forecasts for PM10 levels in the coming days. Markov chains allow such predictions to be given in a tabular form, convenient to use, without the need for calculations.

The data for the “cold” months are modelled using three Markov chains with different degrees of discretization of the original values, respectively with 12, 7 and 3 possible states. The latter, with states: {in the norm}, {slightly above the norm} and {a strong excess of the norm}, can be used without official data on the exact PM10 levels. Determining the condition of the PM10 pollutant today in this case can also be done on the basis of a personal assessment of the purity of the air over the city at the moment.

The measured levels in the period 01.01.2020 - 31.03.2020 were used as test data. They show consistency of the measured levels in 2020 with all three Markov chains considered. The obtained confidence sets for forecasting the future level of PM10 under the current state based on each of the three models were tested on the 2020 data. The percentage of correct forecasts is consistent and even slightly exceeds the expected percentage. This means that the obtained tabular values can be used to predict PM10 levels in the following years, in the months from October to March. This is due to the relative stability in the behaviour of the PM10 pollutant in the “cold” months. Indeed, in recent years no such high values have been observed as at the beginning of the period under review, but due to the discretization in Markov chain modelling, for each of the years the highest values fall into the same state - S12, including PM10 values above 120 $\mu g/m^3$.

ACKNOWLEDGEMENTS

This paper contains results of the work on project No 2020-FNSE–04, financed by “Scientific Research Fund” of Ruse University.

REFFERENCES

1. EEA (2019). Air quality in Europe — 2019 report, ISBN: 978-92-9480-088-6 https://www.eea.europa.eu/publications/air-quality-in-europe-2019.
2. A. Stefanova, I. Zheleva, and M. Filipova, “Deviation analysis using applicable methodologies for impact assessment on the ambient air component from industrial site in Razgrad, Bulgaria,” in *AMiTaNS’19*, edited by M. Todorov, AIP CP2025 (American Institute of Physics, Melville, NY, 2019), #040014 (2018); https://doi.org/10.1063/1.5064898.
3. A. Stefanova and M. Filipova, “Approaches to predict the impact of air atmospheric component in environmental impact assessment reports in assessing extensions to active industrial objects,” in *International Symposium “The Environment and the Industry”, SIMI 2018*, Proceedings book, doi: http://doi.org/10.21698/simi.2018.fp22.
4. M. Filipova, I. Zheleva, and P. Roussev (2013) Characteristics of PM air pollution along Bulgaria - Romania Danube region, *ECOLOGICA* **71**, 215-217.
5. M. Filipova, I. Zheleva, I. Tsvetanova, D. Stefanova, and P. Rusev, “Analysis of the state of ambient air in the border region Bulgaria-Romania,” in *International Symposium “The Environmental and The Industry” SIMI 2016, Bucharest*, Symposium Proceedings, pp.440-450, http://www.simiecoind.ro/doi10-21698simi-2016-0062-analysis-of-the-state-of-ambient-air-in-the-border-region-bulgaria-romania/.
6. I. Zheleva, E. Veleva, and M. Filipova, “Analysis and modeling of daily air pollutants in the city of Ruse, Bulgaria,” in *AMiTaNS’17*, edited by M.Todorov, AIP CP1895 (American Institute of Physics, Melville, NY, 2017), #030007; http://dx.doi.org/10.1063/1.5007366.
7. A. Stefanova, I. Zheleva, M. Filipova, and I. Tsvetanova (2018) Examination of the possibility of transborder pollution in the days with registered exceedances of pollutant dust in the city of Ruse, Bulgaria, *MATEC Web of Conferences* **145**, 01004, NCTAM 2017, https://doi.org/10.1051/matecconf/201814501004.

8. I. Zheleva and M. Filipova, "Atmospheric characteristics statistic study of Ruse region, Bulgaria," in *AMiTaNS'16*, edited by M. Todorov, AIP CP1773 (American Institute of Physics, Melville, NY, 2016), #110019; doi: 10.1063/1.4965023, http://scitation.aip.org/content/aip/proceeding/aipcp/1773.
9. I. Tsvetanova, I. Zheleva, M. Filipova, and A. Stefanova (2018) "Statistical analysis of ambient air PM10 contamination during winter periods for Ruse region, Bulgaria," *MATEC Web of Conferences* **145**, 01007, NCTAM 2017, https://doi.org/10.1051/matecconf/201814501007.
10. I. Tsvetanova, I. Zheleva, and M. Filipova, "Statistical study of the influence of some atmospheric characteristics upon the particulate matter (PM10) air pollutant in the city of Ruse, Bulgaria," in AMiTaNS'18, AIP CP2025, edited by M. Todorov (American Institute of Physics, Melville, NY, 2018) #110006 (2018); https://doi.org/10.1063/1.5064949.
11. A. Ivanov, S. Gocheva-Ilieva, " Short-time particulate matter PM10 forecasts using predictive modeling techniques," in *AMiTaNS'13*, AIP CP1561, edited by M. Todorov (American Institute of Physics, Melville, NY, 2013), pp. 209-218, doi: 10.1063/1.4827230.
12. S. Gocheva-Ilieva and A. Ivanov (2019) Assaying stochastic SARIMA and generalized regularized regression for particulate matter PM10 modeling and forecasting, *International Journal of Environment and Pollution* **66**: 41-62; doi: 10.1504/IJEP.2019.104520.
13. S. Gocheva-Ilieva, A. Ivanov,and I. Iliev, "Exploring key air pollutants and forecasting particulate matter PM10 by a two-step SARIMA approach," in AIP CP2106 (American Institute of Physics, Melville, NY, 2019) , #020004, 10p, https://doi.org/10.1063/1.5109327.
14. E. Veleva and I. Zheleva, "GARCH models for particulate matter PM10 air pollutant in the city of Ruse, Bulgaria," in *AMiTaNS'18*, AIP CP2025, edited by M. Todorov (American Institute of Physics, Melville, NY, 2018) #040016, https://doi.org/10.1063/1.5064900.
15. I. Tsvetanova, I. Zheleva, and M. Filipova, "Statistical study of the influence of the atmospheric characteristics upon the particulate matter (PM10) air pollutant in the city of Silistra, Bulgaria," in *AMiTaNS'19*, AIP CP2164, edited by M. Todorov (American Institute of Physics, Melville, NY, 2019) #120014, https://doi.org/10.1063/1.5130874.
16. http://www.riosv-ruse.org/mesechna-spravka-za-nivata-na-fpch10.
17. D.L. Isaacson and R.W. Madsen, *Markov Chains: Theory and Applications*. 1st edn (R.E. Krieger Pub. Co., New York, 1985) ISBN-13:9780898748345, 256p.
18. T.W. Anderson and L.A. Goodman (1957) Statistical Inference about Markov Chains, *Ann. Math. Statist.* **28**(1), 89-110.
19. A. Murphy and R. Katz, *Probability, Statistics, and Decision Making In the Atmospheric Sciences* (CRC Press, Boca Raton, 1985), https://doi.org/10.1201/9780429303081.
20. R.J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd edn (OTexts: Melbourne, Australia, 2018) OTexts.com/fpp2.
21. C. M. Grinstead and J. L. Snell, *Introduction to Probability* (AMS, 2012).