Environmental Modeling & Assessment (2023) 28:1–14
https://doi.org/10.1007/s10666-022-09854-1

Probabilistic Approach to Modelling, Identification and Prediction of Environmental Pollution

Magdalena Bogalecka[1]

Received: 20 August 2020 / Accepted: 6 September 2022 / Published online: 16 September 2022

Abstract
The probabilistic general model of environmental pollution process based on the semi-Markov one is developed and presented in the paper. The semi-Markov chain model approach is based on using prior information to predict the characteristic of some systems. Now, the semi-Markov process is used for the environmental pollution assessment. The methods and procedures to estimate the environmental pollution process's basic parameters such as the vector of initial probabilities and the matrix of probabilities of transition between the process's states as well as the methods and procedures to identify the process conditional sojourn times' distributions at the particular environmental pollution states and their mean values are proposed and defined. Next, the formulae to predict the main characteristics of the environmental pollution process such as the limit values of transient probabilities and mean total sojourn times in the particular states in the fixed time interval are given. Finally, the application of the presented model and methods for modelling, identification and prediction of the air environmental pollution process generated by sulphur dioxide within the exemplary industrial agglomeration is proposed.

Keywords Pollution · Air pollution · Sulphur dioxide · Semi-Markov process

1 Introduction

The air pollution is a phenomenon defined as a presence of harmful, toxic substances or their mixtures whose high concentration in the atmosphere is detrimental to the quality of life and causes health risks. The transport, industrialisation, agriculture and using chemicals in everyday life have become the main sources of pollution in urban and industrial areas. Some global and regional organisation as well as governments identify the needs for monitoring and assessment the environment quality, established its standards and limits as well as providing information to the public.

According to the World Health Organization (WHO), the air pollution has become the world's largest environmental health risk [1]. The WHO points the air pollution as one of eight reasons of death in the world. Regarding to the air pollution, there are known some air quality assessment methods based on one kind of pollutant concentration in air (commonly such as sulphur dioxide–SO_2, carbon monoxide–CO, nitrogen dioxide–NO_2, ozone–O_3, and particulate matter with a diameter less than 2.5 µm or between 2.5 and 10 µm–$PM_{2.5}$ and PM_{10}, respectively) or all together. There are also some methods for forecasting the air quality. These approaches are usually based on the historical statistical data that are the background for the prediction of the future pollutant concentration. Generally, the statistical forecasting methods, recently reviewed in [2], include linear or nonlinear (e.g. Gaussian) regression [3–6], dispersion [7–9], neural network [10–17], fuzzy logic [18–23] or hybrid systems [24–30].

The new approach, based on the semi-Markov process for the environmental pollution assessment is proposed in this research. The semi-Markov process theory was developed by Lévy [31] and Smith [32]. These processes are used for modelling real systems and are commonly applied in the queuing and reliability theories [33–38]. In this study, modelling, identification and prediction procedures are adopted from [39] and [40, 41] where they are used related to the operation processes of complex technical systems and the critical infrastructure accident consequences for the marine environment, respectively.

✉ Magdalena Bogalecka
m.bogalecka@wznj.umg.edu.pl

[1] Department of Industrial Products Quality and Chemistry, Gdynia Maritime University, 83 Morska Str, Gdynia 81-225, Poland

The semi-Markov model is considered in the paper as the approach more flexible in opposition to the traditional Markov model. In Markov model, it is assumed that the distributions of conditional sojourn times in particular states are only exponential. For the semi-Markov approach, the distributions of sojourn times do not necessarily have to be exponential. Thus, the model is possible to use for any distribution of the operation process sojourn times at the particular operation states. This way the semi-Markov approach is more sensible, giving the better description of reality.

2 Theoretical Background of Modelling, Identification and Prediction of Environmental Pollution Process

2.1 Modelling Environmental Pollution Process

It is assumed that the pollutant's concentration in the environment takes $v, v \in N$ different concentration states $s^1, s^2, \ldots, s^v$. Further, the environmental pollution process $S(t),\ t \in \langle 0, +\infty)$ with the pollutant's concentration states from the set $\{s^1, s^2, \ldots, s^v\}$ is defined. Moreover, a semi-Markov model of the environmental pollution process $S(t)$ is assumed. Its random conditional sojourn time at the pollutant's concentration state s^k while the next transition will be done to the state $s^l,\ k, l = 1, 2, \ldots, v,\ k \neq l$ is denoted by θ^{kl}. In the paper, a state transition means that the environmental pollution process shifts from one state to another. If a state remains the same, there were no transitions, because the process is still at the particular pollutant's concentration state. For example there is no change from state 1 to state 1, but the process is still at state 1, under the condition it can shift to another state, except state 1. Thus, the state can transit to the next one or still stay in the first one.

Hence, the environmental pollution process $S(t)$ is defined by the matrix of probabilities $p^{kl},\ k, l = 1, 2, \ldots, v,\ k \neq l$ of the process $S(t)$ transitions between the pollutant's concentration states s^k and s^l.

$$\left[p^{kl}\right]_{vxv} = \begin{bmatrix} p^{11} & p^{12} & \ldots & p^{1v} \\ p^{21} & p^{22} & \ldots & p^{2v} \\ \vdots & \vdots & \ddots & \vdots \\ p^{v1} & p^{v2} & \ldots & p^{vv} \end{bmatrix}, \tag{1}$$

where by the formal agreement

$$\forall k = 1, 2, \ldots, v,\ p^{kk} = 0.$$

Moreover, the environmental pollution process $S(t)$ is described by the matrix of conditional distribution functions of sojourn times θ^{kl} at the state s^k while its next transition will be done to the state $s^l,\ k, l = 1, 2, \ldots, v,\ k \neq l$

$$\left[H^{kl}(t)\right]_{vxv} = \begin{bmatrix} H^{11}(t) & H^{12}(t) & \ldots & H^{1v}(t) \\ H^{21}(t) & H^{22}(t) & \ldots & H^{2v}(t) \\ \vdots & \vdots & \ddots & \vdots \\ H^{v1}(t) & H^{v2}(t) & \ldots & H^{vv}(t) \end{bmatrix}, \tag{2}$$

where by the formal agreement

$$\forall k = 1, 2, \ldots, v,\ H^{kk}(t) = 0.$$

The matrix (Eq. (2)) corresponds to the matrix of conditional densities of sojourn times θ^{kl} of the environmental pollution process $S(t)$ at the state s^k while its next transition will be done to the state $s^l,\ k, l = 1, 2, \ldots, v,\ k \neq l$

$$\left[h^{kl}(t)\right]_{vxv} = \begin{bmatrix} h^{11}(t) & h^{12}(t) & \ldots & h^{1v}(t) \\ h^{21}(t) & h^{22}(t) & \ldots & h^{2v}(t) \\ \vdots & \vdots & \ddots & \vdots \\ h^{v1}(t) & h^{v2}(t) & \ldots & h^{vv}(t) \end{bmatrix}, \tag{3}$$

where by the formal agreement

$$\forall k = 1, 2, \ldots, v,\ h^{kk}(t) = 0.$$

2.2 Identification of Environmental Pollution Process

Prior to estimating the unknown parameters of the environmental pollution process $S(t)$, its kinds and number v of states $s^1, s^2, \ldots, s^v$ should be fixed and defined. The identification of environmental pollution process $S(t)$ is based on its number of realisations.

The matrix of realisations $n^{kl},\ k, l = 1, 2, \ldots, v$ of the numbers of the process $S(t)$ transitions between the states s^k and s^l during the experimental time is fixed

$$\left[n^{kl}\right]_{vxv} = \begin{bmatrix} n^{11} & n^{12} & \ldots & n^{1v} \\ n^{21} & n^{22} & \ldots & n^{2v} \\ \vdots & \vdots & \ddots & \vdots \\ n^{v1} & n^{v2} & \ldots & n^{vv} \end{bmatrix}. \tag{4}$$

Taking into account the numbers given in the matrix (Eq. (4)), the matrix $\left[p^{kl}\right]_{vxv},\ k, l = 1, 2, \ldots, v$ of realisations of probabilities of the process $S(t)$ transitions between the states s^k and s^l during the experimental time is evaluated where

$$p^{kl} = \frac{n^{kl}}{n^k},\ k, l = 1, 2, \ldots, v,\ k \neq l, \tag{5}$$

and

$$\forall k = 1, 2, \ldots, v,\ p^{kk} = 0,$$

and

$$n^k = \sum_{l \neq k}^{v} n^{kl}, \ k = 1, 2, \ldots, v \tag{6}$$

is the realisation of the total number of transitions of the process $S(t)$ from the state s^k during the experimental time.

Further, the hypotheses on the distribution functions of the process $S(t)$ conditional sojourn times θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ at the state s^k while the next transition is to the state s^l are formulated and verified on the base of their realisations $\theta^{kl}_{\gamma}, \gamma = 1, 2, \ldots, n^{kl}$.

In order to estimate the distribution parameters of conditional sojourn times θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ of the process $S(t)$ at its particular states, the empirical characteristics of their realisations at these states should be determined as follows:

- the realisations of empirical mean values $\overline{\theta}^{kl}$ of conditional sojourn times θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ at the state s^k while the next transition is to the state s^l on the base of their realisations $\theta^{kl}_{\gamma}, \gamma = 1, 2, \ldots, n^{kl}$, according to the formula

$$\overline{\theta}^{kl} = \frac{1}{n^{kl}} \sum_{\gamma=1}^{n^{kl}} \theta^{kl}_{\gamma}, \ k, l = 1, 2, \ldots, v, \ k \neq l \tag{7}$$

- the number $\overline{r}^{kl}$ of disjoint intervals $I_j = \langle a^{kl}_j, b^{kl}_j \big)$, $j = 1,2, \ldots, \overline{r}^{kl}$, including the realisations $\theta^{kl}_{\gamma}, \gamma = 1,2, \ldots, n^{kl}$ of conditional sojourn times θ^{kl}, $k, l = 1,2, \ldots, v, \ k \neq l$ at the state s^k while the next transition is to the state s^l, according to the formula

$$\overline{r}^{kl} \cong \sqrt{n^{kl}} \tag{8}$$

- the length d^{kl} of intervals $I_j = \langle a^{kl}_j, b^{kl}_j \big)$, $j = 1, 2, \ldots, \overline{r}^{kl}$, according to the formula

$$d^{kl} = \frac{\overline{R}^{kl}}{\overline{r}^{kl} - 1} \tag{9}$$

where

$$\overline{R}^{kl} = \max_{1 \leq \gamma \leq n^{kl}} \theta^{kl}_{\gamma} - \min_{1 \leq \gamma \leq n^{kl}} \theta^{kl}_{\gamma} \tag{10}$$

- the ends a^{kl}_j, b^{kl}_j, $j = 1, 2, \ldots, \overline{r}^{kl}$ of intervals $I_j = \langle a^{kl}_j, b^{kl}_j \big)$, $j = 1,2, \ldots, \overline{r}^{kl}$, according to the formulae

$$\begin{aligned} a^{kl}_1 &= \max\{ \min_{1 \leq \gamma \leq n^{kl}} \theta^{kl}_{\gamma} - \tfrac{d^{kl}}{2}, 0\}, \\ b^{kl}_j &= a^{kl}_1 + j d^{kl}, \ j = 1, 2, \ldots, \overline{r}^{kl}, \\ a^{kl}_j &= b^{kl}_{j-1}, \ j = 2, 3, \ldots, \overline{r}^{kl}, \end{aligned} \tag{11}$$

in such a way that

$$I_1 \cup I_2 \cup \ldots \cup I_{n^{kl}} = \langle a^{kl}_1, b^{kl}_{n^{kl}} \big),$$

and

$$I_i \cap I_j = \emptyset \text{ for all } i \neq j, \ i, j = 1, 2, \ldots, \overline{r}^{kl}$$

- the numbers n^{kl}_j of realisations $\theta^{kl}_{\gamma}, \gamma = 1, 2, \ldots, n^{kl}$ in the intervals $I_j = \langle a^{kl}_j, b^{kl}_j \big)$, $j = 1, 2, \ldots, \overline{r}^{kl}$, according to the formula

$$n^{kl}_j = \#\{\gamma : \theta^{kl}_{\gamma} \in I_j, \gamma \in \{1, 2, \ldots, n^{kl}\}\}, \ j = 1, 2, \ldots, \overline{r}^{kl}, \tag{12}$$

where

$$\sum_{j=1}^{\overline{r}^{kl}} n^{kl}_j = n^{kl}$$

(the number of elements of the set is expressed with the symbol #).

In order to formulate and further to verify the non-parametric hypothesis relating to the distribution form of the environmental pollution process's conditional sojourn time θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ at the state s^k while the next transition is to the state s^l, on the base of its realisations $\theta^{kl}_{\gamma}, \gamma = 1, 2, \ldots, n^{kl}$, the procedure adopted from [39] is applied as follows:

- to construct and to plot the realisation of the histogram of the environmental pollution process's conditional sojourn time θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ at the state s^k, defined by the formula

$$\overline{h}^{n^{kl}}(t) = \frac{n^{kl}_j}{n^{kl}} \ \text{for } t \in I_j \tag{13}$$

- to compare the histogram $\overline{h}^{n^{kl}}(t)$ with the graphs of the density functions given in Chapter 2 in [39], and next to select one of them and to formulate the following hypothesis $\mathcal{H}$, relating to the unknown form of the conditional sojourn time θ^{kl}, $k, l = 1, 2, \ldots, v, \ k \neq l$ distribution: the environmental pollution process's conditional sojourn time θ^{kl} at the state s^k while the next transition is to the state s^l has the distribution expressed with the density function $h^{kl}(t)$
- to join the intervals I_j having the number n^{kl}_j of realisations less than 4 with the neighbour ones I_{j+1} or I_{j-1} to obtain the numbers of realisations not less than 4 in all intervals
- to fix a new number of intervals $\overline{\overline{r}}^{kl}$
- to determine new intervals

$$\overline{I}_j = \langle \overline{a}^{kl}_j, \overline{b}^{kl}_j \big), \ j = 1, 2, \ldots, \overline{\overline{r}}^{kl}$$

- to fix the numbers $\overline{n}_j^{kl}$ of realisations in the new intervals $\overline{I}_j$, $j = 1, 2, \ldots, \overline{\overline{r}}^{kl}$
- under the assumption that the hypothesis $\mathcal{H}$ is true, to calculate the hypothetical probabilities that the conditional sojourn time θ^{kl} takes values from the new interval $\overline{I}_j$ according to the formula

$$p_j = P\left(\theta^{kl} \in \overline{I}_j\right) = P\left(\overline{a}_j^{kl} \le \theta^{kl} < \overline{b}_j^{kl}\right) = H^{kl}\left(\overline{b}_j^{kl}\right) - H^{kl}\left(\overline{a}_j^{kl}\right), j = 1, 2, \ldots, \overline{\overline{r}}^{kl} \quad (14)$$

  where $H^{kl}\left(\overline{b}_j^{kl}\right)$ and $H^{kl}\left(\overline{a}_j^{kl}\right)$ are the values of the distribution function $H^{kl}(t)$ of the random conditional sojourn time θ^{kl} corresponding to the density function $h^{kl}(t)$ assumed in the hypothesis $\mathcal{H}$
- to calculate the realisation of the χ^2- Pearson's statistics $U^{n^{kl}}$, according to the formula

$$u^{n^{kl}} = \sum_{j=1}^{\overline{\overline{r}}^{kl}} \frac{\left(\overline{n}_j^{kl} - n^{kl} p_j\right)^2}{n^{kl} p_j} \quad (15)$$

- to assume the significance level α of the test (for instance $\alpha = 0.05$)
- to fix the number $\overline{\overline{r}}^{kl} - z - 1$ of degrees of freedom where, according to [39], $z = 0$ for the chimney distribution function $H^{kl}(t)$, $z = 1$ for the exponential distribution function $H^{kl}(t)$, and $z = 2$ for Gamma distribution function $H^{kl}(t)$
- to read the value u^α from tables of the χ^2- Pearson's distribution for the fixed values of the significance level α and the number of degrees of freedom $\overline{\overline{r}}^{kl} - z - 1$ such that the equality holds

$$P\left(U^{n^{kl}} > u^\alpha\right) = \alpha \quad (16)$$

- to determine the acceptance and critical domains in the form of the intervals $\langle 0, u^\alpha \rangle$ and $(u^\alpha, +\infty)$, respectively
- to compare the critical value u^α read from tables of the χ^2- Pearson's distribution with the obtained value $u^{n^{kl}}$ of the realisation of the statistics $U^{n^{kl}}$ and to decide on the formulated hypothesis $\mathcal{H}$ in the following way: if the value $u^{n^{kl}}$ does not belong to the critical domain, i.e. when $u^{n^{kl}} \le u^\alpha$ then the hypothesis $\mathcal{H}$ is not rejected, otherwise if the value $u^{n^{kl}}$ belongs to the critical domain, i.e. when $u^{n^{kl}} > u^\alpha$ then the hypothesis $\mathcal{H}$ is rejected

Finally, the mean values M^{kl} of the conditional sojourn times θ^{kl} are determined as follows [39]:

$$M^{kl} = E\left[\theta^{kl}\right] = \int_0^{+\infty} t dH^{kl}(t) = \int_0^{+\infty} t h^{kl}(t) dt, \; k, l = 1, 2, \ldots, v, \; k \ne l. \quad (17)$$

2.3 Prediction of Environmental Pollution Process

It is assumed, that the unknown parameters of the environmental pollution process $S(t)$ are identified using the procedure given in "Sect. 2.2". Now, the main characteristics of the environmental pollution process $S(t)$ can be predicted. Namely, taking into account the formula for the total probability, the unconditional distribution functions of sojourn times θ^k, $k = 1, 2, \ldots, v$ at particular states s^k of the process $S(t)$ are determined by

$$H^k(t) = \sum_{l=1}^{v} p^{kl} H^{kl}(t), \; k = 1, 2, \ldots, v, \quad (18)$$

that are complied with the density functions given by

$$h^k(t) = \sum_{l=1}^{v} p^{kl} h^{kl}(t), \; k = 1, 2, \ldots, v. \quad (19)$$

Hence, the expected values $E\left[\theta^k\right]$ of variables θ^k are given by

$$M^k = E\left[\theta^k\right] = \sum_{l=1}^{v} p^{kl} M^{kl}, \; k = 1, 2, \ldots, v, \quad (20)$$

where p^{kl} are defined by (Eq. (5)) and M^{kl} are defined by (Eq. (17)).

The limit values of the transient probabilities of the process $S(t)$ at its particular states

$$p^k(t) = P\left(S(t) = s^k\right), \; t \in \langle 0, +\infty), \; k = 1, 2, \ldots, v,$$

are calculated according to the formula

$$p^k = \lim_{t \to +\infty} p^k(t) = \frac{\pi^k M^k}{\sum_{l=1}^{v} \pi^l M^l}, \; k = 1, 2, \ldots, v. \quad (21)$$

The probabilities π^k satisfy the system of following equations:

$$\begin{cases} \left[\pi^k\right] = \left[\pi^k\right]\left[p^{kl}\right] \\ \sum_{l=1}^{v} \pi^l = 1 \end{cases} \quad (22)$$

where

$$\left[\pi^k\right] = \left[\pi^1, \pi^2, \ldots, \pi^v\right],$$

and $\left[p^{kl}\right]$ is given by Eq. (1).

The asymptotic distribution of the sojourn total time $\hat{\theta}^k$ at the state s^k, $k = 1, 2, \ldots, v$ of the process $S(t)$ in the time interval $\langle 0, \theta \rangle$, $\theta > 0$ is normal with the expected value

$$\hat{M}^k = E\left[\hat{\theta}^k\right] \cong p^k \theta, \quad (23)$$

where p^k are given by Eq. (21).

Table 1 Air quality according to SO_2 concentration

Pollution level	SO_2 concentration (μg/m^3)	Air quality
1	0–50	Very good
2	50.1–100	Good
3	100.1–200	Moderate
4	200.1–350	Sufficient
5	350.1–500	Bad
6	> 500	Very bad

Based on Main Inspectorate for Environmental Protection (http://powietrze.gios.gov.pl/pjp/current?lang=en)

3 Application–Preliminary Analysis of (Air) Environmental Pollution Process Generated by Sulphur Dioxide

Sulphur dioxide (SO_2) is an invisible gas that has a nasty and pungent odour. It reacts easily with other substances to form harmful compounds, such as sulphuric acid, sulphurous acid and sulphate particles.

Sulphur dioxide is formed in the urbanised and industrial areas by burning coal in domestic fireplaces and the combustion of fossil fuels containing sulphur or sulphur compounds. Then the flue gas is the major anthropogenic source of sulphur dioxide in the air. The fossil fuels have different concentrations of sulphur and sulphur compounds. The coal and oil may contain up to 3% of these substances, whereas a natural gas may be completely free of them. Some chemical reactions in the air transforms SO_2 into sulphuric acid (H_2SO_4) that is condensed into droplets, dissolved in the moisture of air (drops of rain, snow, clouds) and as a so-called acid rain reaches the surface of earth and rivers, seas, oceans and other water areas as well.

Sulphur dioxide affects both health and the environment. It harms the human respiratory system, reduces the lung function and makes breathing difficult. Therefore, people with asthma and chronic lung diseases are more sensitive to these effects than normal individuals. The SO_2 deposition implicates the destruction of vegetation, the degradation of soils and building materials. Due to the harmful properties of SO_2, its limit values for the ambient concentration that correspond to different levels of health concern are distinguished (Table 1). These values are also used as a component of the air quality indicators. The pollution levels presented in Table 1 correspond to Polish ones published by Main Inspectorate for Environmental Protection.

In the experiment, the SO_2 concentration data comes from the monitoring station AM3 located in Gdańsk-Nowy Port (Fig. 1) and free-accessible through https://powietrze.gios.gov.pl/. The AM3 is one of nine stations belonging to the ARMAAG monitoring network of Tri-City (Gdynia, Sopot and Gdańsk) agglomeration in Poland. This agglomeration is situated in Pomerania–the north and seaside part of Poland and has a population of over 1 million people. The area is affected by the pollution coming from industrial sectors as well as transport and domestic sources. Within the air ARMAAG monitoring system continuous measurements (counted every hour) of the air quality are taken at several

Fig. 1 Location of monitoring station AM3 in Gdańsk-Nowy Port (based on ARMAAG)

representative points, where the concentrations of pollutants are the highest ones.

3.1 Modelling (Air) Environmental Pollution Process Generated by Sulphur Dioxide

Under the assumption that the sulphur dioxide concentration in the air is changing in time, taking into account data from Table 1, the following $v = 9$ sulphur dioxide concentration states $s^k,\ k = 1, 2, \dots, 9$ of the environmental pollution process $S(t)$ are arbitrarily distinguished (level 1 from Table 1 is divided into four additional sublevels expressed with state s^1, s^2, s^3 and s^4, respectively):

- a concentration state s^1–SO_2 concentration in the air belongs to the interval $(0, 3.5\rangle$ μg/m^3.
- a concentration state s^2–SO_2 concentration in the air belongs to the interval $(3.6, 17.5\rangle$ μg/m^3.
- a concentration state s^3–SO_2 concentration in the air belongs to the interval $(17.6, 35.0\rangle$ μg/m^3.
- a concentration state s^4–SO_2 concentration in the air belongs to the interval $(35.1, 50.0\rangle$ μg/m^3.
- a concentration state s^5–SO_2 concentration in the air belongs to the interval $(50.1, 100.0\rangle$ μg/m^3.
- a concentration state s^6–SO_2 concentration in the air belongs to the interval $(100.1, 200.0\rangle$ μg/m^3.
- a concentration state s^7–SO_2 concentration in the air belongs to the interval $(200.1, 350.0\rangle$ μg/m^3.
- a concentration state s^8–SO_2 concentration in the air belongs to the interval $(350.1, 500.0\rangle$ μg/m^3.
- a concentration state s^9–SO_2 concentration in the air belongs to the interval $(500.1, +\infty)$ μg/m^3.

Then, according to Eqs. (1)–(3), the environmental pollution process $S(t)$ is expressed by the matrix of probabilities $\left[p^{kl}\right]_{9x9}$ of transitions between the particular states and the matrix of distribution functions $\left[H^{kl}(t)\right]_{9x9}$ or equivalently by the matrix of corresponding to them density functions $\left[h^{kl}(t)\right]_{9x9}$ of conditional sojourn times at the particular states.

3.2 Identification of (Air) Environmental Pollution Process Generated by Sulphur Dioxide

The experiment is performed in Gdańsk-Nowy Port (Fig. 1) during the 120-day period (8th Oct 2019–4th Feb 2020) and the statistical data coming from the real realisation are collected and given in Appendix. Through this experiment, there are not observed realisations in states s^6, s^7, s^8 and s^9, then the matrix of realisations $n^{kl},\ k, l = 1, 2, \dots, 5,\ k \neq l$ of numbers of the process $S(t)$ transitions between the states s^k and s^l during the experimental time are fixed and expressed according to Eq. (4).

$$\left[n^{kl}\right]_{5x5} = \begin{bmatrix} 0 & 142 & 1 & 0 & 1 \\ 143 & 0 & 52 & 15 & 2 \\ 0 & 56 & 0 & 28 & 7 \\ 0 & 15 & 34 & 0 & 12 \\ 0 & 0 & 4 & 18 & 0 \end{bmatrix}.$$

Hence, according to Eq. (6), the realisation of the total numbers of the process $S(t)$ transitions from the state $s^k,\ k = 1, 2, \dots, 5$ during the experimental time is

$$\left[n^{k}\right]_{1x5} = [144, 212, 91, 61, 22].$$

Further, applying Eq. (5), the matrix $\left[p^{kl}\right]_{5x5},\ k, l = 1,2, \dots, 5,\ k \neq l$ of realisations of probabilities of the process $S(t)$ transitions between the states s^k and s^l during the experimental time is fixed as follows:

$$\left[p^{kl}\right]_{5x5} = \begin{bmatrix} 0 & 0.986 & 0.007 & 0 & 0.007 \\ 0.675 & 0 & 0.245 & 0.071 & 0.009 \\ 0 & 0.615 & 0 & 0.308 & 0.077 \\ 0 & 0.246 & 0.557 & 0 & 0.197 \\ 0 & 0 & 0.182 & 0.818 & 0 \end{bmatrix}. \quad (24)$$

Applying the procedure and formulae given in "Sect. 2.2", and based on the data given in Appendix, the empirical parameters of the conditional sojourn times $\theta^{kl},\ k, l = 1, 2, \dots, 5,\ k \neq l$ of the process $S(t)$ can be determined. The conditional sojourn time θ^{21} is an example of this procedure application presented below. The conditional sojourn time θ^{21} is one having sufficient populous set of its realisations, that is it assumed $n = 143$ values presented in Appendix.

The results for the conditional sojourn time θ^{21} are:

- the realisation $\bar{\theta}^{21}$ of the defined by Eq. (7) mean value of the conditional sojourn time θ^{21} of the environmental pollution process's state s^2 when the next transition is to the state s^1

$$\bar{\theta}^{21} = \frac{1}{143} \sum_{\gamma=1}^{143} \theta_{\gamma}^{21} = 5.503 \quad (25)$$

- the number $\bar{r}^{21}$ of disjoint intervals $I_j = \langle a_j^{21}, b_j^{21}\big)$, $j = 1,2, \dots, \bar{r}^{21}$ including the realisations θ_{γ}^{21}, $\gamma = 1, 2, \dots, 143$ of the conditional sojourn time θ^{21} at the environmental pollution process's state s^2 when the next transition is to the state s^1, defined by Eq. (8)

$$\bar{r}^{21} \cong \sqrt{143} \cong 12$$

- the length d^{12} of intervals $I_j = \langle a_j^{21}, b_j^{21}\big)$, $j = 1, 2, \dots, 12$, defined by Eq. (9), after considering Eq. (10)

$$\overline{R}^{21} = \max_{1 \le \gamma \le 143} \theta_\gamma^{21} - \min_{1 \le \gamma \le 143} \theta_\gamma^{21} = 42 - 1 = 41,$$

is

$$d^{21} = \frac{\overline{R}^{21}}{\overline{r}^{21} - 1} = \frac{41}{11} = 3.7$$

- the ends a_j^{21}, b_j^{21} of intervals $I_j = \langle a_j^{21}, b_j^{21} \big)$, $j = 1,2, \dots, 12$, defined by Eq. (11), after considering

$$\min_{1 \le \gamma \le 143} \theta_\gamma^{21} - \frac{d^{21}}{2} = 1 - \frac{3.7}{2} = -0.8636,$$

are

$$\begin{array}{ll}
a_1^{21} = \max\{-0.8636, 0\} = 0 & b_1^{21} = a_1^{21} + 1d^{21} = 0 + 3.7 = 3.7 \\
a_2^{21} = b_1^{21} = 3.7 & b_2^{21} = a_1^{21} + 2d^{21} = 0 + 2 \cdot 3.7 = 7.4 \\
a_3^{21} = b_2^{21} = 7.4 & b_3^{21} = a_1^{21} + 3d^{21} = 0 + 3 \cdot 3.7 = 11.1 \\
a_4^{21} = b_3^{21} = 11.1 & b_4^{21} = a_1^{21} + 4d^{21} = 0 + 4 \cdot 3.7 = 14.8 \\
a_5^{21} = b_4^{21} = 14.8 & b_5^{21} = a_1^{21} + 5d^{21} = 0 + 5 \cdot 3.7 = 18.5 \\
a_6^{21} = b_5^{21} = 18.5 & b_6^{21} = a_1^{21} + 6d^{21} = 0 + 6 \cdot 3.7 = 22.2 \\
a_7^{21} = b_6^{21} = 22.2 & b_7^{21} = a_1^{21} + 7d^{21} = 0 + 7 \cdot 3.7 = 25.9 \\
a_8^{21} = b_7^{21} = 25.9 & b_8^{21} = a_1^{21} + 8d^{21} = 0 + 8 \cdot 3.7 = 29.6 \\
a_9^{21} = b_8^{21} = 29.6 & b_9^{21} = a_1^{21} + 9d^{21} = 0 + 9 \cdot 3.7 = 33.3 \\
a_{10}^{21} = b_9^{21} = 33.3 & b_{10}^{21} = a_1^{21} + 10d^{21} = 0 + 10 \cdot 3.7 = 37.0 \\
a_{11}^{21} = b_{10}^{21} = 37.0 & b_{11}^{21} = a_1^{21} + 11d^{21} = 0 + 11 \cdot 3.7 = 40.7 \\
a_{12}^{21} = b_{11}^{21} = 40.7 & b_{12}^{21} = a_1^{21} + 12d^{21} = 0 + 12 \cdot 3.7 = 44.4
\end{array} \tag{26}$$

- the numbers n_j^{21} of realisations θ_γ^{21} in particular intervals $I_j = \langle a_j^{21}, b_j^{21} \big)$, $j = 1, 2, \dots, 12$ defined by Eq. (12) are as follows:

$$\begin{gathered} n_1^{21} = 87, n_2^{21} = 25, n_3^{21} = 14, n_4^{21} = 7, n_5^{21} = 4, n_6^{21} = 1, \\ n_7^{21} = 2, n_8^{21} = 0, n_9^{21} = 1, n_{10}^{21} = 0, n_{11}^{21} = 0, n_{12}^{21} = 2. \end{gathered} \tag{27}$$

Using the procedure given in "Sect. 2.2" as well as the data given in Appendix and the above results, the hypotheses concerning the distribution forms of the environmental pollution process's conditional sojourn times θ^{kl}, $k, l = 1, 2, \dots, 5$, $k \neq l$ at the particular states may be verified. To do this, a sufficiently numerous set of these variables realisations is needed. It means that the sets of particular realisations coming from the experiment should contain at least 30 ones (see Appendix). The conditional sojourn time θ^{21} is the one having the most numerous set of its realisations and preliminarily analysed above in this section.

The histogram $\overline{h}^{21}(t)$ of the environmental pollution process's conditional sojourn time θ^{21} realisation defined by Eq. (13) is presented and illustrated in Table 2 and Fig. 2, respectively.

After analysing and comparing the realisation of histogram $\overline{h}^{21}(t)$ with the graphs of the density function of distributions distinguished in Chapter 2 in [39], the following hypothesis $\mathcal{H}$ is formulated: the environmental pollution process's conditional sojourn time θ^{21} at the state s^2 when the next transition is to the state s^1 has the exponential distribution expressed with the density function of the form

$$h^{21}(t) = \begin{cases} 0 & t < x^{21} \\ \alpha^{21} \exp\left[-\alpha^{21}\left(t - x^{21}\right)\right] & t \ge x^{21}. \end{cases} \tag{28}$$

The unknown parameters of the hypothetical density function Eq. (28) are estimated using (4.13) in [39], and the results are as follows:

$$\begin{gathered} x^{21} = a_1^{21} = 0, \\ \alpha^{21} = \frac{1}{\overline{\theta}^{21} - x^{21}} = \frac{1}{5.2028 - 0} = 0.192204. \end{gathered} \tag{29}$$

Next, substituting Eq. (29) into Eq. (28), the hypothetical density function takes the form

$$\begin{aligned} h^{21}(t) &= \begin{cases} 0 & t < 0 \\ 0.192204 \exp[-0.192204(t - 0)] & t \ge 0 \end{cases} \\ &= \begin{cases} 0 & t < 0 \\ 0.192204 \exp(-0.192204t) & t \ge 0. \end{cases} \end{aligned} \tag{30}$$

Taking the integral of the hypothetical density function $h^{21}(t)$ of the conditional sojourn time θ^{21} expressed by Eq. (30), the hypothetical distribution function $H^{21}(t)$ takes the form.

$$H^{21}(t) = \int_0^t h^{21}(t)dt = \begin{cases} 0 & t < 0 \\ 1 - \exp(-0.192204t) & t \ge 0. \end{cases} \tag{31}$$

Next, the intervals of the histogram $\overline{h}^{21}(t)$ having the numbers n_j^{21} of realisations less than 4 are jointed into new ones and the following steps are performed:

- the new number of intervals are fixed

$$\overline{\overline{r}}^{21} = 6$$

Table 2 The realisation of the histogram of the environmental pollution process's conditional sojourn time θ^{21}

$I_j = \langle a_j^{21}, b_j^{21} \big)$	0–3.7	3.7–7.4	7.4–11.1	11.1–14.8	14.8–18.5	18.5–22.2	22.2–25.9	25.9–29.6	29.6–33.3	33.3–37	37–40.7	40.7–44.4
n_j^{21}	87	25	14	7	4	1	2	0	1	0	0	2
$\overline{h}^{21}(t) = \frac{n_j^{21}}{n^{21}}$	$\frac{87}{143}$	$\frac{25}{143}$	$\frac{14}{143}$	$\frac{7}{143}$	$\frac{4}{143}$	$\frac{1}{143}$	$\frac{2}{143}$	$\frac{0}{143}$	$\frac{1}{143}$	$\frac{0}{143}$	$\frac{0}{143}$	$\frac{2}{143}$

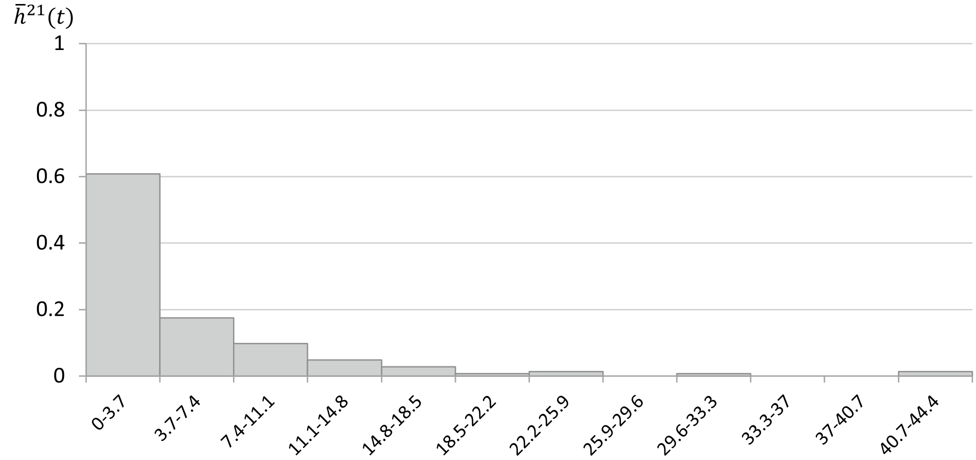

Fig. 2 The graph of the histogram of the environmental pollution process's conditional sojourn time θ^{21}

- the new intervals are determined

$$\bar{I}_1 = \langle 0, 3.7), \bar{I}_2 = \langle 3.7, 7.4), \bar{I}_3 = \langle 7.4, 11.1), \bar{I}_4 = \langle 11.1, 14.8), \bar{I}_5 = \langle 14.8, 18.5), \bar{I}_6 = \langle 18.5, +\infty)$$

- the numbers of realisations in the new intervals are fixed

$$\bar{n}_1^{21} = 87, \bar{n}_2^{21} = 25, \bar{n}_3^{21} = 14, \bar{n}_4^{21} = 7, \bar{n}_5^{21} = 4, \bar{n}_6^{21} = 6$$

- the hypothetical probabilities that the conditional sojourn time θ^{21} takes values from the new intervals are calculated using Eq. (14)

$$p_1 = P\left(\theta^{21} \in \bar{I}_1\right) = P\left(0 \le \theta^{21} < 3.7\right) = H^{21}(3.7) - H^{21}(0) \cong 0.5089 - 0 = 0.5089,$$

$$p_2 = P\left(\theta^{21} \in \bar{I}_2\right) = P\left(3.7 \le \theta^{21} < 7.4\right) = H^{21}(7.4) - H^{21}(3.7) \cong 0.7588 - 0.5089 = 0.2499,$$

$$p_3 = P\left(\theta^{21} \in \bar{I}_3\right) = P\left(7.4 \le \theta^{21} < 11.1\right) = H^{21}(11.1) - H^{21}(7.4) \cong 0.8816 - 0.7588 = 0.1227,$$

$$p_4 = P\left(\theta^{21} \in \bar{I}_4\right) = P\left(11.1 \le \theta^{21} < 14.8\right) = H^{21}(14.8) - H^{21}(11.1) \cong 0.9418 - 0.8816 = 0.0603,$$

$$p_5 = P\left(\theta^{21} \in \bar{I}_5\right) = P\left(14.8 \le \theta^{21} < 18.5\right) = H^{21}(18.5) - H^{21}(14.8) \cong 0.9714 - 0.9418 = 0.0296,$$

$$p_6 = P\left(\theta^{21} \in \bar{I}_6\right) = P\left(18.5 \le \theta^{21} < +\infty\right) = 1 - H^{21}(18.5) \cong 1 - 0.9714 = 0.0286$$

- the realisation of the χ^2- Pearson's statistics are calculated using Eq. (15)

$$u^{21} = \sum_{j=1}^{6} \frac{\left(\bar{n}_j^{21} - n^{21}p_j\right)^2}{n^{21}p_j} \cong \frac{(87-143 \cdot 0.5089)^2}{143 \cdot 0.5089} + \frac{(25-143 \cdot 0.2499)^2}{143 \cdot 0.2499} + \frac{(14-143 \cdot 0.1227)^2}{143 \cdot 0.1227} + \frac{(7-143 \cdot 0.0603)^2}{143 \cdot 0.0603} + \frac{(4-143 \cdot 0.0296)^2}{143 \cdot 0.0296} + \frac{(6-143 \cdot 0.0286)^2}{143 \cdot 0.0286}$$
$$\cong 2.78 + 3.23 + 0.72 + 0.30 + 0.01 + 0.90 = 7.94$$

- the significance level $\alpha = 0.05$ is assumed
- the number of degrees of freedom for the hypothetical exponential distribution ($z = 1$) is fixed

$$\bar{\bar{r}}^{21} - z - 1 = 6 - 1 - 1 = 4$$

- the value u^α for the fixed values of the significance level $\alpha = 0.05$ and the number of degrees of freedom $\bar{\bar{r}}^{21} = 4$ is read from tables of the χ^2- Pearson's distribution, that according to Eq. (16), amounts $u^\alpha = 9.49$, therefore the acceptance domain in the form of the interval $\langle 0, 9.49\rangle$ and the critical domain in the form of the interval $(9.49, +\infty)$ are determined as presented in Fig. 3
- the obtained value $u^{21} = 7.94$ of the realisation of the statistics U^{21} and the critical value $u^\alpha = 9.49$ read from tables of χ^2- Pearson's distribution are compared and the hypothesis $\mathcal{H}$ is not rejected since the value $u^{21} = 7.94$ belongs to the acceptance domain, i.e.

$$u^{21} = 7.94 \le u^\alpha = 9.49.$$

Proceeding in the analogous way, based on the data given in Appendix, the forms of the particular density function $h^{kl}(t)$ of the environmental pollution process's conditional sojourn times θ^{kl}, $k, l = 1, 2, \ldots, 5$, $k \ne l$ that have a sufficient number of their realisation at the particular states are identified. The results are as follows:

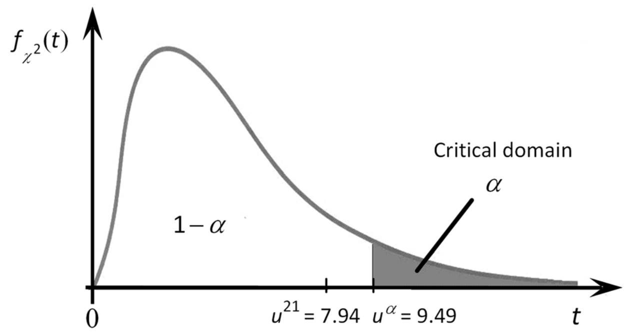

Fig. 3 The graphical interpretation of acceptance and critical intervals for the chi-square goodness-of-fit test

- the conditional sojourn time θ^{12} has Gamma distribution expressed by the density function

$$h^{12}(t) = \begin{cases} 0 & t < 0 \\ \frac{t^{-0.5092}\exp(-\frac{t}{19.5997})}{4.3076\cdot\Gamma(0.4908)} & t \ge 0 \end{cases} \tag{32}$$

- the conditional sojourn time θ^{23} has the chimney distribution expressed by the density function

$$h^{23}(t) = \begin{cases} 0 & t < 0 \\ 0.0723077 & 0 \le t < 12.5 \\ 0.0012821 & 12.5 < 87.5 \\ 0 & t \ge 87.5 \end{cases} \tag{33}$$

- the conditional sojourn time θ^{32} has the exponential distribution expressed by the density function

$$h^{32}(t) = \begin{cases} 0 & t < 0.1667 \\ 0.5853659\exp(-0.5853659t + 0.0975610) & t \ge 0.1667 \end{cases} \tag{34}$$

- the conditional sojourn time θ^{34} has the exponential distribution expressed by the density function

$$h^{34}(t) = \begin{cases} 0 & t < 0.375 \\ 0.6153846\exp(-0.6153846t + 0.2307692) & t \ge 0.375 \end{cases} \tag{35}$$

- the conditional sojourn time θ^{43} has the exponential distribution expressed by the density function

$$h^{43}(t) = \begin{cases} 0 & t < 0.5 \\ 0.9189189\exp(-0.9189189t + 0.4594594) & t \ge 0.5. \end{cases} \tag{36}$$

When there are less than 30 realisations of the environmental pollution process $S(t)$, it is assumed that such conditional sojourn time θ^{kl}, $k, l = 1, 2, \dots, 5$, $k \ne l$ has the empirical density function given by

$$h^{kl}(t) = \frac{1}{n^{kl}}\#\left\{\gamma : \theta_\gamma^{kl} \in I_j, \gamma \in \left\{1, 2, \dots, n^{kl}\right\}\right\},\ j = 1, 2, \dots, \bar{r}^{kl} \tag{37}$$

that complies with the following distribution function:

$$H^{kl}(t) = \frac{1}{n^{kl}}\#\left\{\gamma : \theta_\gamma^{kl} < t, \gamma \in \left\{1, 2, \dots, n^{kl}\right\}\right\} \tag{38}$$

for $t \ge 0$, $k, l = 1, 2, \dots, 5$, $k \ne l$ (the number of elements of the set is expressed with the symbol #).

For instance, the environmental pollution process's conditional sojourn time θ^{24} assumed $n = 15$ values given in Appendix. The order sample realisations θ^{24} is 1, 1, 1, 1, 1, 2, 2, 3, 5, 6, 9, 14, 18, 21, 21. Thus the conditional sojourn time θ^{24} has the empirical density function and the distribution function respectively given by

$$h^{24}(t) = \begin{cases} 0 & t < 1 \\ 5/15 & 1 \le t < 2 \\ 2/15 & 2 \le t < 3 \\ 1/15 & 3 \le t < 5 \\ 1/15 & 5 \le t < 6 \\ 1/15 & 6 \le t < 9 \\ 1/15 & 9 \le t < 14 \\ 1/15 & 14 \le t < 18 \\ 1/15 & 18 \le t < 21 \\ 2/15 & t \ge 21 \end{cases}$$

$$H^{24}(t) = \int_0^t h^{24}(t)dt = \begin{cases} 0 & t < 1 \\ 5/15 & t < 2 \\ 7/15 & t < 3 \\ 8/15 & t < 5 \\ 9/15 & t < 6 \\ 10/15 & t < 9 \\ 11/15 & t < 14 \\ 12/15 & t < 18 \\ 13/15 & t < 21 \\ 1 & t \ge 21. \end{cases} \tag{39}$$

Proceeding in the analogous way, based on the data given in Appendix, it is assumed that the conditional sojourn times $\theta^{42}, \theta^{45}, \theta^{53}$ and θ^{54} have also the empirical density function in the following forms:

$$h^{42}(t) = \begin{cases} 0 & t < 1 \\ 13/15 & 1 \le t < 2 \\ 1/15 & 2 \le t < 3 \\ 1/15 & t \ge 3 \end{cases} \tag{40}$$

$$h^{45}(t) = \begin{cases} 0 & t < 1 \\ 7/12 & 1 \le t < 2 \\ 3/12 & 2 \le t < 3 \\ 1/12 & 3 \le t < 4 \\ 1/12 & t \ge 4 \end{cases} \tag{41}$$

$$h^{53}(t) = \begin{cases} 0 & t < 1 \\ 2/4 & 1 \le t < 2 \\ 1/4 & 2 \le t < 3 \\ 1/4 & t \ge 3 \end{cases} \tag{42}$$

$$h^{54}(t) = \begin{cases} 0 & t < 1 \\ 12/18 & 1 \le t < 2 \\ 3/18 & 2 \le t < 3 \\ 2/18 & 3 \le t < 5 \\ 1/18 & t \ge 5. \end{cases} \tag{43}$$

When only the number of realisations of process $S(t)$ is known and all these realisations are equal to an approximate value, it is assumed that such conditional sojourn time $\theta^{kl},\ k, l = 1, 2, \dots, 5,\ k \ne l$ has the uniform distribution in the interval between this value minus to this value plus its half (Bogalecka, 2020). For instance, the environmental pollution process's conditional time θ^{35} assumed $n^{35} = 7$ values given in Appendix. All of them equal to 1. Thus the conditional sojourn time θ^{35} has the uniform density function and the distribution function respectively given by

$$h^{35}(t) = \begin{cases} 0 & t < 0.5 \\ 1 & 0.5 \le t < 1.5 \\ 0 & t \ge 1.5 \end{cases}$$
$$H^{35}(t) = \int_0^t h^{35}(t)dt = \begin{cases} 0 & t < 0.5 \\ t & 0.5 \le t < 1.5 \\ 1 & t \ge 1.5. \end{cases} \tag{44}$$

Proceeding in the analogous way, based on the data given in Appendix, it is assumed that the conditional sojourn times θ^{13}, θ^{15} and θ^{25} have also the uniform density function in the following forms:

$$h^{13}(t) = \begin{cases} 0 & t < 6.5 \\ 1 & 6.5 \le t < 19.5 \\ 0 & t \ge 19.5 \end{cases} \tag{45}$$

$$h^{15}(t) = \begin{cases} 0 & t < 2 \\ 1 & 2 \le t < 6 \\ 0 & t \ge 6 \end{cases} \tag{46}$$

$$h^{25}(t) = \begin{cases} 0 & t < 0.5 \\ 1 & 0.5 \le t < 1.5 \\ 0 & t \ge 1.5. \end{cases} \tag{47}$$

After accepting the density functions of the particular conditional sojourn times $\theta^{kl},\ k, l = 1, 2, \dots, 5,\ k \ne l$ of the environmental pollution process $S(t)$ given by Eq. (30) and Eqs. (32)–(36), the general formula Eq. (17) is applied to find their mean value $M^{kl} = E\left[\theta^{kl}\right]$. In other cases, when the statistical identification of the environmental pollution process's conditional sojourn times distributions at the particular states is not possible because of the lack of sufficient numbers of their realisations, the approximate empirical values of mean values $M^{kl} = E\left[\theta^{kl}\right]$ of the conditional sojourn times at the particular states are calculated using the formula (7). The results are given in the matrix below

$$\left[M^{kl}\right]_{5x5} = \begin{bmatrix} 0 & 9.620 & 13 & 0 & 4 \\ 5.203 & 0 & 10.456 & 7.067 & 1 \\ 0 & 1.875 & 0 & 2 & 1 \\ 0 & 1.2 & 1.588 & 0 & 1.667 \\ 0 & 0 & 2.25 & 1.556 & 0 \end{bmatrix}. \tag{48}$$

3.3 Prediction of (Air) Environmental Pollution Process Generated by Sulphur Dioxide

The process the environmental pollution process $S(t)$ is identified in "Sect. 3.2". Now, its main characteristics may be predicted using the procedure presented in "Sect. 2.3".

Applying Eq. (20) and considering Eqs. (24) and (48), the approximate mean values $M^k,\ k = 1, 2, \dots, 5$ of unconditional sojourn times of variables $\theta^k,\ k = 1, 2, \dots, 5$ can be evaluated. The values that are not equal to 0 are presented only, and they are as follows:

$$\begin{aligned} M^1 &= M^{12}p^{12} + M^{13}p^{13} + M^{15}p^{15} = 9.620 \cdot 0.986 + 13 \\ &\quad \cdot 0.007 + 4 \cdot 0.007 = 9.486 + 0.090 + 0.028 = 9.604, \\ M^2 &= M^{21}p^{21} + M^{23}p^{23} + M^{24}p^{24} + M^{25}p^{25} = 5.203 \\ &\quad \cdot 0.675 + 10.456 \cdot 0.245 + 7.067 \cdot 0.071 + 1 \cdot 0.009 \\ &\quad = 3.509 + 2.565 + 0.500 + 0.009 = 6.583, \\ M^3 &= M^{32}p^{32} + M^{34}p^{34} + M^{35}p^{35} = 1.875 \\ &\quad \cdot 0.615 + 2 \cdot 0.308 + 1 \cdot 0.077 \\ &\quad = 1.154 + 0.615 + 0.077 = 1.846, \\ M^4 &= M^{42}p^{42} + M^{43}p^{43} + M^{45}p^{45} = 1.200 \\ &\quad \cdot 0.246 + 1.588 \cdot 0.557 + 1.667 \cdot 0.197 \\ &\quad = 0.295 + 0.885 + 0.328 = 1.508, \\ M^5 &= M^{53}p^{53} + M^{54}p^{54} = 2.250 \\ &\quad \cdot 0.182 + 1.556 \cdot 0.818 \\ &\quad = 0.409 + 1.272 = 1.681. \end{aligned} \tag{49}$$

To find the limit values of the transient probabilities $p^k,\ k = 1, 2, \dots, 5$ at particular states of the process $S(t)$, the system of equations (Eq. (22)) has to be solved that here it takes the following form:

$$\begin{cases} \pi^1 = 0.675\pi^2 \\ \pi^2 = 0.986\pi^1 + 0.615\pi^3 + 0.246\pi^4 \\ \pi^3 = 0.007\pi^1 + 0.245\pi^2 + 0.557\pi^4 + 0.182\pi^5 \\ \pi^4 = 0.071\pi^2 + 0.308\pi^3 + 0.818\pi^5 \\ \pi^5 = 0.007\pi^1 + 0.009\pi^2 + 0.077\pi^3 + 0.197\pi^4 \\ \pi^1 + \pi^2 + \pi^3 + \pi^4 + \pi^5 = 1. \end{cases} \quad (50)$$

This system of equations solution is

$$\pi^1 = 0.271, \pi^2 = 0.401, \pi^3 = 0.172, \pi^4 = 0.115, \pi^5 = 0.041.$$

Hence, according to Eq. (21) and considering Eq. (49), the approximate limit values of the transient probabilities p^k, $k = 1, 2, \ldots, 5$ at the particular states s^k of the process $S(t)$ are

$$p^1 = 0.448, p^2 = 0.455, p^3 = 0.055, p^4 = 0.030, p^5 = 0.012. \quad (51)$$

Further, by Eq. (23) and considering Eq. (51), the approximate mean values of the sojourn total time $\widehat{\theta}^k$ of the process $S(t)$ in the fixed time interval $\theta = 1$ month = 720 h at the particular states s^k, $k = 1, 2, \ldots, 5$ expressed in hours are

$$\widehat{M}^1 = E\left[\widehat{\theta}^1\right] = 322.7, \widehat{M}^2 = E\left[\widehat{\theta}^2\right] = 327.7, \widehat{M}^3 = E\left[\widehat{\theta}^3\right] = 39.4, \widehat{M}^4 = E\left[\widehat{\theta}^4\right] = 21.6, \widehat{M}^5 = E\left[\widehat{\theta}^5\right] = 8.6. \quad (52)$$

According to Eq. (51), states s^2 and s^1 reach the highest value of the transient probabilities equal to $p^2 = 0.455$ and $p^1 = 0.448$, respectively. Similarly, according to Eq. (52), states s^2 and s^1 reach the highest value of the sojourn total times equal to $\widehat{M}^2 = 327.7$ h and $\widehat{M}^1 = 322.7$ h per 720 h of the fixed time, respectively. The states s^6, s^7, s^8 and s^9 have never occurred during the experimental time; thus, their values of transient probabilities and the values of sojourn total times equal to zero. Nevertheless, the values Eqs. (51)–(52) are evaluated based on the experiment and the real statistical data; therefore, the values (Eqs. (51)–(52)) may change and being more precise if the experiment duration is longer.

Moreover, the last results (Eqs. (51)–(52)) can play a practically role in the minimisation of air pollution caused by sulphur dioxide and its losses mitigation what is the subject of future research.

4 Conclusion

The model of the environmental pollution process based on the semi-Markov process designed and presented in the paper is a novel approach. The procedure of its practical application is illustrated in the modelling, identification and prediction of the environmental pollution process caused by the air pollutant, i.e. sulphur dioxide. The proposed method provides to establish the limit values of transient probabilities and the mean values of sojourn total times staying at particular pollution states indicating the concentration of pollutant (Table 3). There is the first approach to usage of this method; therefore, the obtained results should be treated just as an illustration of the proposed method.

The developed general model of the environmental pollution process is a universal tool. It can be used successfully in regard to other environmental pollutants existing in air or water and soil [42]. Moreover, the model allows to consider two or more pollutants in parallel. It means that the next stage of research will consider the air environmental pollution process generated jointly by SO_2, CO, NO_2, O_3, $PM_{2.5}$ and PM_{10} commonly used in determination of air quality index (AQI) that is based on these pollutants concentration and describes the air pollution levels.

Table 3 Main characteristics of the environmental pollution process–final results of application to sulphur dioxide

Environmental pollution state $s^k, k = 1, 2, \ldots, 9$	SO_2 concentration (μg/m^3)	Transient probability	Sojourn time at state $s^k, k = 1, 2, \ldots, 9$ (h/month)
s^1	0–3.5	0.448	322.7
s^2	3.6–17.5	0.455	327.7
s^3	17.6–35	0.055	39.4
s^4	35.1–50	0.030	21.6
s^5	50.1–100	0.012	8.6
s^6	100.1–200	0	0
s^7	200.1–350	0	0
s^8	350.1–500	0	0
s^9	> 500	0	0

Appendix. Realisations of conditional sojourn times θ^{kl} at states of the environmental pollution process $S(t)$

Transitions $s^k \to s^l$	Transition time θ^{kl} (min)	Number of transitions n^{kl}
$s^1 \to s^2$	7, 21, 3, 32, 2, 4, 34, 44, 1, 2, 1, 1, 11, 2, 1, 5, 1, 10, 7, 1, 6, 5, 15, 10, 2, 1, 7, 1, 1, 1, 1, 1, 15, 3, 23, 26, 5, 2, 4, 3, 5, 10, 1, 3, 1, 7, 9, 1, 45, 1, 6, 1, 1, 3, 6, 1, 2, 9, 1, 4, 1, 40, 67, 2, 32, 1, 2, 1, 8, 1, 1, 22, 2, 17, 18, 2, 2, 1, 1, 1, 6, 9, 1, 2, 1, 2, 10, 29, 18, 47, 30, 7, 1, 1, 1, 2, 1, 1, 1, 35, 26, 5, 3, 5, 1, 40, 6, 3, 17, 7, 2, 2, 1, 1, 18, 1, 15, 8, 1, 4, 1, 6, 1, 10, 16, 2, 7, 1, 4, 4, 16, 59, 7, 70, 12, 5, 20, 1, 16, 1, 45, 30	142
$s^1 \to s^3$	13	1
$s^1 \to s^5$	4	1
$s^2 \to s^1$	2, 5, 1, 2, 5, 2, 2, 2, 8, 2, 1, 20, 1, 9, 10, 15, 1, 17, 1, 1, 8, 3, 6, 32, 8, 2, 1, 9, 1, 1, 6, 11, 1, 3, 4, 1, 6, 1, 2, 2, 2, 9, 2, 3, 23, 3, 1, 1, 12, 1, 6, 1, 1, 2, 24, 1, 1, 42, 1, 4, 4, 1, 1, 1, 1, 1, 13, 8, 4, 2, 8, 6, 3, 12, 11, 1, 3, 5, 4, 2, 10, 1, 4, 16, 3, 2, 10, 1, 1, 1, 5, 1, 6, 2, 3, 2, 5, 2, 2, 2, 2, 2, 2, 1, 14, 1, 6, 1, 2, 2, 3, 1, 1, 2, 4, 5, 2, 18, 3, 1, 2, 2, 1, 1, 2, 1, 42, 2, 7, 11, 3, 13, 3, 2, 1, 7, 1, 7, 14, 1, 13, 6, 4	143
$s^2 \to s^3$	6, 5, 2, 7, 11, 3, 3, 4, 2, 1, 2, 7, 46, 25, 4, 76, 2, 1, 1, 1, 1, 1, 1, 6, 5, 1, 22, 11, 3, 6, 5, 3, 3, 18, 4, 1, 9, 4, 10, 2, 1, 1, 3, 12, 9, 3, 2, 1, 3, 1, 4, 5	52
$s^2 \to s^4$	6, 21, 1, 2, 3, 9, 14, 1, 18, 5, 1, 1, 1, 2, 21	15
$s^2 \to s^5$	1, 1	2
$s^3 \to s^2$	1, 1, 1, 7, 4, 1, 1, 3, 1, 1, 1, 3, 1, 3, 1, 2, 1, 2, 2, 1, 1, 1, 2, 1, 1, 1, 5, 2, 3, 1, 2, 1, 2, 1, 2, 3, 11, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 4, 1, 2, 2, 2, 1, 1, 1	56
$s^3 \to s^4$	1, 2, 1, 3, 6, 4, 4, 1, 3, 1, 1, 2, 4, 1, 2, 1, 1, 2, 1, 1, 1, 2, 1, 4, 1, 2, 1, 2	28
$s^3 \to s^5$	1, 1, 1, 1, 1, 1, 1	7
$s^4 \to s^2$	2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1	15
$s^4 \to s^3$	1, 1, 2, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 2, 1, 2, 3, 1, 1, 6, 3, 1, 1, 1, 1, 1, 2, 4, 2, 1	34
$s^4 \to s^5$	2, 2, 3, 1, 1, 1, 1, 1, 4, 1, 1, 2	12
$s^5 \to s^3$	3, 1, 1, 4	4
$s^5 \to s^4$	1, 1, 1, 1, 1, 1, 2, 4, 2, 3, 3, 1, 1, 1, 1, 2, 1, 1	18

Acknowledgements Acknowledgment is made to the Chief Inspectorate for Environmental Protection, Poland, for free access to sulphur dioxide concentration data used in the paper.

Funding This work was supported by the Gdynia Maritime University ("Monitoring and analysis of the impact of selected substances and materials in terms of environmental protection"–project grant no. WZNJ/2022/PZ/10).

Availability of Data and Material The data that support the findings of this study are available free of charge and remain the property of the Chief Inspectorate for Environmental Protection, Poland (https://powietrze.gios.gov.pl/).

Declarations

Ethics Approval The author declares no ethical violation during the preparation of this manuscript.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author declares no competing interests.

References

1. WHO. (2014). 7 million premature deaths annually linked to air pollution. *Air Quality & Climate Change, 22*(1), 53–59.
2. Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air pollutants forecasts: An overview. *International Journal of Environmental Research and Public Health, 15*(4), 780. https://doi.org/10.3390/ijerph15040780
3. Dalal, P. (2015). Modeling of air quality index. *International Journal of Advanced Research in Engineering and Applied Sciences, 4*(9), 1–11.
4. Huebnerova, Z., & Michalek, J. (2014). Analysis of daily average PM_{10} predictions by generalized linear models in Brno, Czech Republic. *Atmospheric Pollution Research, 5*, 471–476. https://doi.org/10.5094/APR.2014.055

5. Kaboodvandpour, S., Amanollahi, J., Qhavami, S., & Mohammadi, B. (2015). Assessing the accuracy of multiple regressions, ANFIS, and ANN models in predicting dust storm occurrences in Sanandaj, Iran. *Natural Hazards, 78*, 879–893. https://doi.org/10.1007/s11069-015-1748-0
6. Shafabakhsh, G. A., Taghizadeh, S. A., & Kooshki, S. M. (2018). Investigation and sensitivity analysis of air pollution caused by road transportation at signalized intersections using IVE model in Iran. *European Transport Research Review, 10*, 7. https://doi.org/10.1007/s12544-017-0275-3
7. PriyaDarshini, S., Sharma, M., & Singh, D. (2016). Synergy of receptor and dispersion modelling: Quantification of PM_{10} emissions from road and soil dust not included in the inventory. *Atmospheric Pollution Research, 7*(3), 403–411. https://doi.org/10.1016/j.apr.2015.10.015
8. Shadab, A., Farhan, A. K., & Kafeel, A. (2019). Evaluating traffic-related near-road CO dispersions on an urban road during summer season: A model inter-comparison. *Asian Journal of Water, Environment and Pollution, 16*(1), 69–79. https://doi.org/10.3233/AJW190008
9. Sivacoumar, R., Bhanarkar, A. D., Goyal, S. K., Gadkarib, S. K., & Aggarwal, A. L. (2001). Air pollution modeling for an industrial complex and model performance evaluation. *Environmental Pollution, 111*(3), 471–477. https://doi.org/10.1016/S0269-7491(00)00083-X
10. Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research, 7*, 557–566. https://doi.org/10.1016/j.apr.2016.01.004
11. Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment, 107*, 118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030
12. Fu, M., Wang, W., Le, Z., & Khorram, M. S. (2015). Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. *Neural Computing and Applications, 26*, 1789–1797. https://doi.org/10.1007/s00521-015-1853-8
13. Park, Y., Kwon, B., Heo, J., Hu, X., Liu, Y., & Moon, T. (2020). Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. *Environmental Pollution, 256*, 113395. https://doi.org/10.1016/j.envpol.2019.113395
14. Rahman, N. H. A., Lee, M. H., Suhartono, & Latif, M. T. (2015). Artificial neural networks and fuzzy time series forecasting: An application to air quality. *Quality and Quantity, 49*, 2633–2647. https://doi.org/10.1007/s11135-014-0132-6
15. Sarwat, E., & El-Shanshoury, G. I. (2018). Estimation of air quality index by merging neural network with principal component analysis. *International Journal of Computer Application, 1*(8), 2250–1797. https://doi.org/10.26808/rs.ca.i8v1.01
16. Wongsathan, R., & Seedadan, I. (2016). A hybrid ARIMA and neural networks model for PM-10 pollution estimation: The case of Chiang Mai City Moat Area 9. *Procedia Computer Science, 86*, 273–276. https://doi.org/10.1016/j.procs.2016.05.057
17. Yan, L., Wu, Y., Yan, L., & Zhou, M. (2018). Encoder-decoder model for forecast of PM2.5 concentration per hour. *Proceedings of 1st international cognitive cities conference (IC3)* (pp. 45–50). https://doi.org/10.1109/IC3.2018.00020
18. Bouharati, S., Benzidane, C., Braham-Chaouch, W., & Boumaïza, S. (2014). Air quality index and public health: Modelling using fuzzy inference system. *American Journal of Environmental Engineering and Science, 1*(4), 85–89.
19. Dunea, D., Pohoata, A. A., & Lungu, E. (2011). Fuzzy inference systems for estimation of air quality index. *Romanian Society of Industrial and Applied Mathematics, 7*(2), 63–70.
20. Olvera-Garcia, M. A., Carbajal-Hernandez, J. J., Sanchez-Fernandez, L. P., & Hernandez-Bautista, L. (2016). Air quality assessment using a weighted fuzzy inference system. *Ecological Informatics, 33*, 57–74. https://doi.org/10.1016/j.ecoinf.2016.04.005
21. Xu, Q., & Xu, K. (2018). Assessment of air quality using a cloud model method. *Royal Society Open Science, 5*(9), 171580. https://doi.org/10.1098/rsos.171580
22. Yadav, J., Kharat, V., & Deshpande, A. (2015). Fuzzy-GA modeling in air quality assessment. *Environmental Monitoring and Assessment, 187*, 1–14. https://doi.org/10.1007/s10661-015-4351-7
23. Yang, H., Zhu, Z., Li, Ch., & Li, R. (2020). A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Applied Soft Computing, 87*, 105972. https://doi.org/10.1016/j.asoc.2019.105972
24. Chen, D., Xu, T., Li, Y., Zhou, Y., Lang, J., Liu, X., & Shi, H. (2015). A hybrid approach to forecast air quality during high-PM concentration pollution period. *Aerosol and Air Quality Research, 15*, 1325–1337. https://doi.org/10.4209/aaqr.2014.10.0253
25. Qin, S., Liu, F., Wang, J., & Sun, B. (2014). Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmospheric Environment, 98*, 665–675. https://doi.org/10.1016/j.atmosenv.2014.09.046
26. Wang, P., Liu, Y., Qin, Z., & Zhang, G. (2015). A novel hybrid forecasting model for PM_{10} and SO_2 daily concentrations. *Science of the Total Environment, 505*, 1202–1212. https://doi.org/10.1016/j.scitotenv.2014.10.078
27. Wu, Q., & Lin, H. (2019). A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Science of the Total Environment, 683*, 808–821. https://doi.org/10.1016/j.scitotenv.2019.05.288
28. Yang, H., Jiang, Z., & Lu, H. (2017). A hybrid wind speed forecasting system based on a "decomposition and ensemble" strategy and fuzzy time series. *Energies, 10*(9), 1422. https://doi.org/10.3390/en10091422
29. Zhou, Q., Jiang, H., Wang, J., & Zhou, J. (2014). A hybrid model for $PM_{2.5}$ forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Science of the Total Environment, 496*, 264–274. https://doi.org/10.1016/j.scitotenv.2014.07.051
30. Zhu, S., Yang, L., Wang, W., Liu, X., Lu, M., & Shena, X. (2018). Optimal-combined model for air quality index forecasting: 5 cities in North China. *Environmental Pollution, 243B*, 842–850. https://doi.org/10.1016/j.envpol.2018.09.025
31. Lev'y, P. (1954). Proceesus semi-markoviens. *Proceedings of International Congress of Mathematicians* (pp. 416–426). Amsterdam.
32. Smith, W. L. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society of London, Series A, 232*, 631. https://doi.org/10.1098/rspa.1955.0198
33. Grabski, F. (2015). *Semi-Markov processes: Applications in system reliability and maintenance*. Elsevier. https://doi.org/10.1016/C2013-0-14260-2
34. Iosifescu, M. (1980). *Finite Markov processes and their applications*. John Wiley & Sons Ltd.
35. Kołowrocki, K. (2004). *Reliability of large systems*. Elsevier. https://doi.org/10.1016/B978-0-08-044429-1.X5000-4
36. Kołowrocki, K. (2014). *Reliability of large and complex systems*. Elsevier. https://doi.org/10.1016/C2013-0-12769-9

37. Korolyuk, V. S., Brodi, S. M., & Turbin, A. F. (1975). Semi-Markov processes and their applications. *Journal of Soviet Mathematics, 4*(3), 244–280. https://doi.org/10.1007/BF01097184
38. Limnios, N., & Oprisan, G. (2001). *Semi-Markov processes and reliability*. Birkhauser. https://doi.org/10.1007/978-1-4612-0161-8
39. Kołowrocki, K., & Soszyńska-Budny, J. (2011). *Reliability and safety of complex technical systems and processes: Modeling – identification – prediction – optimization*. Springer. https://doi.org/10.1007/978-0-85729-694-8
40. Bogalecka, M. (2020). *Consequences of maritime critical infrastructure accidents. Environmental impacts. Modeling – identification – prediction – optimization – mitigation*. Elsevier. https://doi.org/10.1016/C2019-0-00396-2
41. Dąbrowska, E., & Soszyńska-Budny, J. (2018). Monte Carlo simulation forecasting of maritime ferry safety and resilience. *IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 376–380). https://doi.org/10.1109/IEEM.2018.8607464
42. Bogalecka, M. (2021). Semi-markovian approach to modelling air pollution. In K. Kołowrocki, et al. (Eds.), *Safety and reliability of systems and processes, summer safety and reliability seminar 2021* (pp. 31–44). Gdynia: Gdynia Maritime University. https://doi.org/10.26408/srsp-2021-03