PROBABILITY THEORY IN TRANSPORT MODELLING PROCESS

ĽUBOMÍR VOLNER

ISSN 2235-2007
DOI: https://doi.org/10.59505/IRHNS.22352007.2023.2.06

Abstract:

Today's world is accelerating and the need to transport goods and people is growing enormously. This leads to the need to think about the possibilities of transport systems with special attention to transport processes. The article presents a basic view of the possibility of modelling transport as such, with an emphasis on the basis of the theoretical possibility of a solution. Many of these external and internal forces produce fluctuations so complex that we are forced to classify them as random and express them by the theory of probability. Therefore, it is important to create a dynamic theory that can predict the behaviour of the system under the influence of these fluctuating forces, and the prediction can be given by probabilistic expressions. This paper describes some basics for the probability features of transport models and also provides an application example of a practical application of probabilistic theory in traffic modelling.

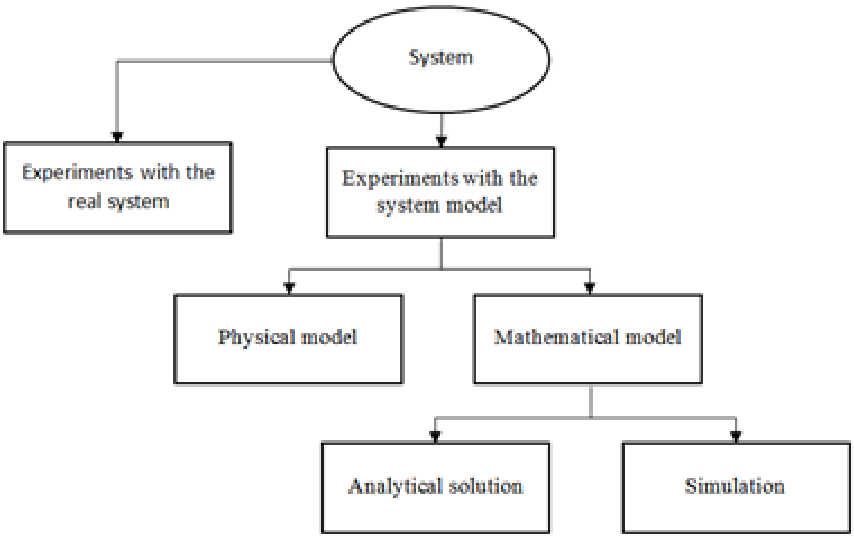
Keywords: transport; transport modelling; probability theory

Introduction

The system is defined as a collection of entities, such as people or machines, who act and work together to achieve a common logical goal (Law A., 2007). You can create a very diverse range of systems. From relatively simple ones, which can be different abstract systems (family system or literature system) or physical systems (house, car, production line) to very complex, complex systems such as natural systems (weather system, universe) (Siebers, P-O 2011), shown in the

Figure 1.

Figure 1 Methods of examining a system



Source: Author, modified by Law A., 2007

The model means a collection of objects, concepts or ideas, arranged in some form and with some bonds other than the model object itself (system) (Shannon, R., 1998). The model thus represents a system under examination (object, problem). It is necessary to specify the important elements of the system that affect its state. Other elements that are not so important for running and describing the system can be ignored in the created model. Such simplification is called abstraction.

Because the world is very large and contains a plethora of different objects, it is possible to create a huge number of different models representing these objects. For this reason, there must be a large number of criteria for their proper distribution, like for example symbolic, dynamic or continuous models. Of course, there are many goals and criteria by which decisions are evaluated in real life. However, until these criteria are reduced to a single objective function, the goals are not well defined for decision-makers (Bukvič L., et al., 2021).

Symbolic models they are the opposite of physical models. While it is possible to carry out physics experiments with physical models (measurement, weighing,...), symbolic models are not possible. Symbolic (abstract) models are based on the principle of a formalized description of the model system. As a rule, a mathematical description is chosen using different types of equations. However, the abstract model may not be described only mathematically; different diagrams can also be used to describe it (Borshchev A, 2020). The main difference between abstract models and physical models is that physical models can directly obtain information about the model system, but the abstract must be solved in some way, either mathematically or logically.

Dynamic models consider output values of the model dependent not only on inputs but also on time, that means these models monitor system development over time.

To predict the future state of the model it is necessary to perform simulation. Various experiments are performed with the model, from changes in input variables to model configuration changes. By observing these changes, we are able to identify possible future states of the model. In railway transport, simulation modelling is one of the tools for assessing capacity (throughput performance), traffic efficiency, and reliability on railway lines (Široký J., et. al., 2021).

Continuous models changes state variables continuously, respecting the time, so that state variables change with the evolution of time.

2. Probability theory in transport modelling

A large number of different models emerged in the first half of the 20th century (Křivda, V., et. al., 2011). Some of these models are even used today to simplify the status of traffic. All of these models are based on the same fundamental relationship between velocity and density of traffic flow, resulting in the intensity of traffic flow. The speeds indicate in kilometres per hour, the density as the number of vehicles per one lane of the road and the intensity as the number of vehicles passing through the lane in one hour. The models in the order in question were created over time as they were re-examined to improve some of their predecessors' deficiency. Therefore, if we want to create a fairly accurate model of traffic flow today for the broadest possible spectrum of conditions, we use so-called multi-mode combined models (Křivda, V., et. al., 2011) that combine different models, i.e. their suitability for certain situations, together.

Types of models by abstraction level - The basic purpose of creating simulation transport models is in particular modelling of the movement and mutual interaction of the participants. The main criterion in the development of transport models is primarily the extent of the model environment, while the scope of the model covers the level of detail the model should represent. Depending on the scale of the model environment, the models can be divided into macroscopic, mesoscopic and microscopic (Křivda, V., et. al., 2011).

The systems were designed to be used by humans. In this use, the resulting human activity and device is referred to as an operation. The dynamic properties of the operating system are only partly defined by the physical properties of the devices that the system contains, but are also, and often above, the human characteristics of the operators, the directives of the administration concerned, and the requirements imposed on the operation by external influences. For example necessary input data for the construction of timetable and, for microsimulation of railway traffic, are, among other things, data on railway infrastructure (Vávra, R., et. al., 2021). Also the majority of related research in the field of mixed and freight railway timetabling and operation can be characterized by quantitative approach (mathematical modelling, network timetable optimization, various optimization of partial problems, etc.) (Drábek, M., et. al., 2021).

Many of these external and internal forces produce fluctuations so complex that we are forced to classify them as random and express them by the theory of probability. Then it is important to create a dynamic theory that can predict the behaviour of the system under the influence of these fluctuating forces, and the prediction can be given by probabilistic expressions.

For this purpose, we assume that the operating system may exist in different states. In a number of cases, these states will form an incomprehensible set and in some cases will be final. If the system reacts to the external and internal forces acting on it, its state changes with time. The theory must be able to predict these changes in terms of probability distribution, with the pi (t) probability that the system is in the state i at time t.

In time t = 0, we can claim that the system is in the state j. This is defined by asking pi (0) = 0 and pi (0) = 0, if i = j. Later at time t, due to the change of forces, we can no longer be sure what the system is. The only thing we can do is predict the probability pi (t) that it will be in the state i at time t. The system of probabilities pi (t) is limited in size by the requirement that probabilities cannot be negative and that their sum must be 1.

$$p_i\,(t)\ \geq 0 \qquad\qquad \textstyle\sum_i p_i\,(t) = 1 \qquad (1)$$

which are actually the basic requirements of any probability distribution. Dynamic theory for stochastic systems must define equations that allow the calculation of changes of these probabilities as time processes. The probability of the condition is related to the probability of transition Pi,j (t0, t1), that the system will be in the state j in time t1, when it is in the state i in time t0.

In the case of restrictions:

- if the transition probability from the initial state to the end state is only one period,
- if it is independent of time t,
- the past state,

then such a system is called Markov, and its behaviour is referred to as Markov´s process.

Examples of discrete time are useful as simple examples and correspond to some operations that are virtually occurring. However, in many operating systems, system state changes may occur at any time, not only at the end of discrete periods. The effects of external errors are recorded and decisions are made with respect to the resulting system activities at any time if the impacts do not occur at the end of the predetermined periods. The use of the Markov model for such systems assumes the time division for infinitesimal periods with a length dt, and the matrix elements Mij must be equal to the probability of transition from state i to state j in time dt. If the events that cause this change are randomly distributed over time, then these probabilities should be proportional to dt, except the transition i → j.

$$M_{ij} \rightarrow R_{ij}dt, \quad i \neq j \qquad (2)$$

$$M_{ij} \rightarrow \sum_{i \neq k} R_{ik}\, dt, \quad i = j \qquad (3)$$

where the average change rate of the system from state i to state j.

If the Markov system is to be, R must be independent of t.

Previous considerations have suggested that the dynamic behaviour of the stochastic operator system is determined by the interaction between the incoming data streams to be served and the operation of the data stream. Arrivals and operations are commonly subjected to random changes and the effect between these two stochastic flows creates changing queues and delays. It is therefore necessary to determine the statistical values of these two processes.

Nepoissonian arrivals can be simulated in a similar way if we imagine them as an infinite number of data flows that must pass through the „Poisson door“ system before they are released to get into the operator's facility. Poissonian arrivals can be obtained using a „simple door“ with a medium speed λ that would open and let a simple flow of data pass in a random time.

$$A_n(t) = U_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{4}$$

Erlang arrivals can be simulated by a sequence of k doors, each of which has a speed k λ and which should release the data stream before this flow of data was free for „arrival" and before another data stream could enter the sequence. Hyperexponential arrivals could be simulated using two alternating randomly selected doors so that the door allows the closest exit flow, while the other door would be closed to use the second flow until the nearest flow of data passed through the door.

The reason for simulating the general distribution of arrivals or operators by combining exponential elements is to use the equation

$$p_j(t + dt) = \left[1 - \sum_{i \neq j} R_{ji} dt\right] p_j(t) + \sum_{i \neq j} p_i(t) R_{ij} dt \tag{5}$$

to reduce the system to an equivalent Markov system. However, this equation can be used only if all the transitions from the state to the state defined by the matrix R are randomly distributed over time with respect to the probability set as in equation (4).

It is clear that the dynamic characteristics of the stochastic operator system are determined by interactions between the incoming data stream statistic and the operator equipment statistics adapted to the rules that describe the behaviour of the incoming data streams before they are comprised i.e. the order in the queue. In fact, the calculation of distribution functions A0 (t) and S0 (t) or densities a(t) and s(t), plus order specification in the queue plus the indication of incoming data streams are in an unlimited number or not, and whether the operator can handle several flows in parallel (multiple operator channels) or not, all must be given or anticipated before we can develop the dynamic characteristics of the system.

There are many examples of how to arrive, which are between free and bound cases. E.g. we can limit the length of the queue, so incoming data streams (if the queue length is less than a certain N value) enter the queue and remain there until they are computed, but the incoming flows if the queue length is N does not exist. In other cases, flows may show impatience by leaving the queue (fall off) if the delay is too high. Individual flows may vary in their behaviour, so the effect can be expressed in terms of probability.

The simple system has poisson-bound data stream arrivals that are left in the queue until the operator is executed by individual operator devices with an exponential time slot layout. By such a system, we find that by reducing the average delay time or the average length of the queue, the ratio ρ between the average arrival speed and the average speed of the operator is reduced and thus the average value of the time 1- ρ during which the operator´s equipment is ineffective. Due to variability in arrival and service times, the operator must be inactive in time to eliminate random arrivals or delays in the operator without showing the wait time in the queue. It is therefore advisable to consider whether there are other options, apart from influencing, to make the desired change.

E.g. we can test what effect the change in the operator operation's statistical expression would have by treating the operator operations more regularly (by reducing scattering) but without changing its average speed, thereby reducing the average length of the queue. This effect can be expressed by simulating non-exponential operation in terms of the combination of exponential phases. E.g. the Erlang operator of the operator time to k, i.e. Ek-1(kμt) corresponds to the phases of the operator phase's k, each of which is at a speed μ that must be terminated gradually before the flow is released and another flow can be served. In this case, the state of the system can be denoted by n, which is the number of flows in the system (in the queue or in the operator) and the variable with the flow phase in the operator, the flow being processed in the phase and proceeding in stages to Phase 1, (in state 0, the operator is ineffective and irrelevant).

In designing systems may be used in essentially two approaches, i.e. direct measurement and experimentation with the real system or modelling. Real system measurement, however, remains an integral part of system models because it provides important inputs to the model and also allows for model validation. Based on the model description approach, we divide the models into analytical or simulation.

The basis for the formulation of the analytical model is usually the representation of the original as a Queuing system (QS). QS consists of:

- devices that provide the operator with data flow,
- data stream queues that are waiting for the operator.

In analytical models, we represent the load in the form of data flows that are required from system resources. From the QS point of view, it is intervals, respectively. Probability distribution of intervals between the arrivals of data streams requesting the service at a given service centre and the operating time; Probability of distribution of the service life of the service centre.

According to the method of analytical solution QS, analytical models can be divided into three groups:

- deterministic models - Middle value models, Extreme models,
- probability models – models with one service centre, service networks,
- models based on the assumptions of the operational analysis.
- The principles of modelling can be divided into two basic parts:
- defining the problem of interest and gathering important data,
- create a mathematical model.

3. Application example of probability distribution of statistical random variables

The example focuses on the analysis of train arrivals and their processing at marshalling yard A in terms of statistical random variables. The refereed observation period of entry requests to the station represented working days, a total of 200 trains.

Within the example, statistical hypothesis testing was performed on:

- input demand stream (train arrivals),
- requirement handling (train preparation for marshalling).

The above testing will be performed using Pearson's chi-squared test. The Pearson chi-squared test is usually used to assess the similarity/difference of the observed distribution (observed frequencies) from the theoretical distribution (expected frequencies). The chi-square test of independence is suitable for comparing two categorical features (Štefancová, V., et. al., 2022). The hypotheses of the test can be formulated as follows: $H0$: there are no statistically significant differences between the compared distributions, $H1$: there are statistically significant differences between the compared distributions. There is always a comparison χ2TAB and χ2VYP. The test statistic of the Pearson chi-squared test calculated (χ2VYP) takes the form

$$\chi^2 = \frac{[n_i - f(x_i)]^2}{f(x_i)} \quad (6)$$

Theoretically, the Pearson χ^2 test statistic asymptotically has χ^2 - distribution with (k - r - 1) degrees of freedom, where p is the number of parameters of the theoretical probability distribution of chi-square tabular (χ2TAB)

$$\chi^2_\alpha\,(k - r - 1) \quad (7)$$

Testing the statistical hypothesis that the random variable of trains arriving within one hour has a Poisson probability distribution at a significance level of $\alpha = 0.05$.

The Poisson distribution is given by the probability distribution:

$$p(x_i) = \frac{\lambda^k . e^{-\lambda}}{k!} \quad (8)$$

where

λ – estimated parameter

$$\lambda = \frac{\sum(n_i . x_i)}{\sum n_i} \quad (9)$$

The functional value f(xi) for each probability xi is calculated according to the following relation:

$$f(x_i) = p(x_i). \sum n_i \quad (10)$$

The observed statistical data and the results of the above computational steps are shown in the following table, where xi represents the hours when 0-6 trains entered the system and ni the number of hours when 0-6 trains entered the system.

Table 1 Calculation table for Poisson probability distribution

The observed statistical data and the results of the Poisson probability distribution					
x_i	n_i	$x_i * n_i$	$p(x_i)$	$f(x_i)$	χ^2
0	17	0	0,19	23,17	1,64
1	49	49	0,32	38,30	2,99
2	32	64	0,26	31,65	0,00
3	11	33	0,14	17,44	2,38
4	7	28	0,06	7,21	0,01
5	4	20	0,02	2,38	1,10
6	1	6	0,01	0,66	0,18
Total	121	200	1,00		8,30

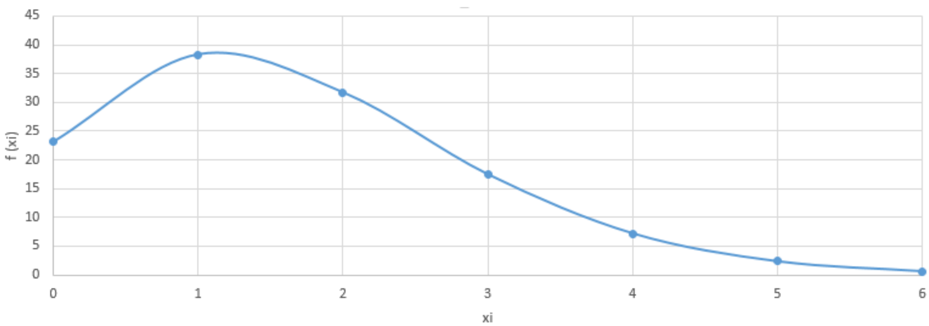
Source: Author

Estimated parameter λ has, when added to formula (9), the value 200/121 = 1.65.

According to the critical value table χ2α at α = 0.05 considering χ20.05 (7-1-1)= χ20.05 (5)=11.1.

It is clear from the result of the calculation that χ2TAB > χ2VYP.

Graph 1 Graphical progression of values for the Poisson distribution calculation table



Source: Author

We can conclude that we accept the hypothesis that the number of trains arriving within one hour has a Poisson probability distribution at the 0.05 significance level (train arrivals at the marshalling yard meet the criteria of a Poisson distribution on α).

The second statistical hypothesis to be tested, is to verify that the random variable of the train service in the entrance track group of yard A has an exponential probability distribution at the significance level α = 0.05.

The exponential distribution is given by the probability distribution:

$$p(x_i) = \int_d^h b.\,e^{-b.x}.\,dx = e^{-bd} - e^{-bh} \quad (11)$$

where

d - lower limit of the specified interval,

h - upper limit of the specified interval,

b - estimated parameter, which is calculated as:

$$b = \frac{1}{\bar{x}} = \frac{1}{\sum(x_i.n_i)} \quad (12)$$

The functional value f(xi) for each probability xi is calculated:

$$f(x_i) = p(x_i) . \sum n_i \quad (13)$$

The observed statistical data and the results of the above computational steps are presented in the following table.

Table 2 Calculation table for Exponencial probability distribution

The observed statistical data and the results of the Exponencial probability distribution							
i	x_i	x_{istr}	n_i	$n_i * x_{istr}$	p (x_i)	f (x_i)	χ^2
1	<0;10)	5	0	0	0,13	26,51	26,51
2	<10;20)	15	3	45	0,11	22,71	17,11
3	<20;30)	25	34	850	0,10	19,46	10,86
4	<30;40)	35	20	700	0,08	16,67	0,66
5	<40;50)	45	25	1 125	0,07	14,29	8,03
6	<50;60)	55	20	1 100	0,06	12,24	4,92
7	<60;70)	65	22	1 430	0,05	10,49	12,64
8	<70;80)	75	18	1 350	0,04	8,99	9,04
9	<80;90)	85	15	1 275	0,04	7,70	6,92
10	<90;100)	95	9	855	0,03	6,60	0,88
11	<100;110)	105	13	1 365	0,03	5,65	9,55
12	<110;120)	115	4	460	0,02	4,84	0,15
13	<120;130)	125	7	875	0,02	4,15	1,96
14	<130;140)	135	2	270	0,02	3,55	0,68
15	<140;150)	145	3	435	0,02	3,05	0,00
16	<150;160)	155	3	465	0,01	2,61	0,06
17	<160;170)	165	1	165	0,01	2,24	0,68
18	<170;180)	175	1	175	0,01	1,92	0,44
Total			200	12 940	0,93		111,09

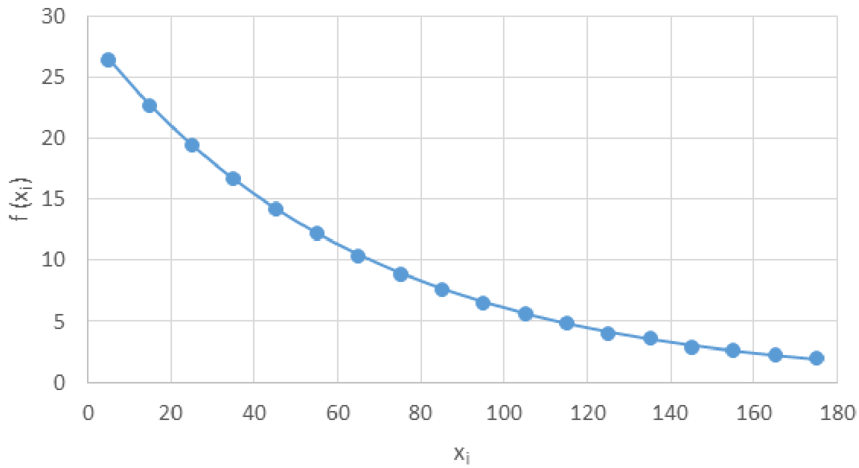
Source: Author

Estimated parameter b has when added to the formula (12) value 200/12,940 = 0.015.

According to the critical value table χ2α at α = 0.05 considering χ20.05 (18-1-1)= χ20.05 (16)=26.3.

It is clear from the result of the calculation that χ2TAB < χ2VYP.

Graph 2 Graphical progression of values for the Exponential distribution calculation table



Source: Author

We can conclude that we reject the hypothesis that the train service time in the entrance track group has an exponential probability distribution at the 0.05 significance level (train arrivals at the marshalling yard do not meet the critical values of the exponential distribution at α).

Conclusion

The complex phenomenon that a transportation system undoubtedly is in the context of traffic model building requires the use of probability theory in many cases to estimate the dynamic evolution of the behaviour of the elements of the system variables over time and to estimate the states of the system over time. For the process of traffic modelling, knowing the probability of occurrence of states of some parameters is crucial in terms of knowing their influence on other elements of the system and the results of the model as a whole. Outputs that can be considered relevant, describing the real system as accurately as possible, are key for the evaluation of the model and at the same time a basic prerequisite for adequate decision making on the chosen solution of the problem in the context of financial, socio-economic and environmental aspects of transport projects.

References

MAY, Adolf D. Traffic Flow Fundamentals. Prentice Hall. 1990. Traffic Flow Fundamentals. ISBN 978 0139260728

BORSHCHEV A. The Big Book Of AnyLogic, AnyLogic. Online, http://www.anylogic.com/the-big-book-of-anylogic

BUKVIĆ L., ŠKRINJAR J., ABRAMOVIČ B., ZITRICKÝ V. Route Selection Decision-Making in an Intermodal Transport Network Using Game Theory. In: Sustainability 2021, vol. 13 (issue 8) ISSN: 2071-1050 online: https://www.mdpi.com/2071-1050/13/8/4443

BYCHKOV I., KAZAKOV A., LEMPERT A., ZHARKOV M., Modeling of Railway Stations Based on Queuing Networks, IDSTU SB RAS, Irkutsk, Russia, 2021, online: https://www.mdpi.com/2076-3417/11/5/2425

DRÁBEK M., JANOŠ V., Construction of Differentiated Periodic Freight Train Paths in Dense Mixed Traffic, In: Sustainability 2021, vol. 13 (issue 15) ISSN: 2071-1050 online: https://www.mdpi.com/2071-1050/13/8/4443

KŘIVDA, V. Problematika homogenizace dopravního proudu v silniční dopravě, In: Perner's contacts, 2008, Sv. III, 3, str. 38 – 43, eISSN 1801-674X,

KŘIVDA, V., ŠKVAIN, V. Učební texty, Městské komunikace a křižovatky, Online, 2011, VŠB – Technická univerzita Ostrava,

LAW, A., M. Simulation modelling and analysis, Boston [u.a.], McGraw-Hill, 2007, ISBN 978-007-125519-6

LIEBERMAN E., RATHI A. K. Traffic Simulation, 1992, online, https://docplayer.net/32619516-Traffic-simulation-by-edward-lieberman-18-ajay-k-rathi-president-kld-associates-inc-300-broadway-huntington-station-ny-11746.html

MACAL, Ch., M., NORTH, M., J. Tutorial on Agent-Based Modeling and Simulation, Online, 2005, http://www.informs-sim.org/wsc05papers/002.pdf.

ŠALAMON, T. Development of Agent-based Models for Economic Simulation. Praha, VŠE-FIS, 2010, Disertační práce

SHANNON, R., E. Introduction to the art and science of simulation, Online, 1998, http://www.informs-sim.org/wsc98papers/001.PDF

SIEBERS, P-O. G54 SIM - Simulation for Computer Scientists, 2011/2012 (Autumn), School of Computer Science, The University of Nottingham, 2011

ŠIROKÝ J., NACHTIGALL P., TISCHER E., GAŠPARÍK J., Simulation of Railway Lines with a Simplified Interlocking System, Sustainability 2021, vol. 13 (issue 3) ISSN: 2071-1050 online: https://www.mdpi.com/2071-1050/13/3/1394

ŠTEFANCOVÁ V., KALAŠOVÁ A., ČULÍK K., MAZANEC J., VOJTEK M., MAŠEK J., Research on the Impact of COVID-19 on Micromobility Using Statistical Methods, Applied sciences, 2022, vol. 12 (issue 16), ISSN: 2076-3417, online, https://www.mdpi.com/2076-3417/12/16/8128

STERMAN, J., D. A Skeptic's Guide to Computer Models, Online, 1991, http://web.mit.edu/jsterman/www/Skeptic's_Guide.pdf

VÁVRA R., JANOŠ V., MICHL Z., Arrangement of the Internal Environment Structure of the SW Database for the Technological Timetable Construction, Transportation Research Procedia, 2021, vol. 53, (p. 146 – 153) ISSN: 2352-1465 online: https://reader.elsevier.com/reader/sd/pii/S2352146521001794?token=27AECB6F063EF7A93817748660E9C6AD6FD5FFF6BCBB592E540E60DD56F327D7F03276F5933B333814B868B49249AAB4&originRegion=eu-west-1&originCreation=20230130142644

VOLNER, R., JANIGA, R.: Projektovanie špeciálnych informačných systémov, (Designing special information systems), Verbum KU Ružomberok 2015, ISBN 978-80-561-0206-0, 209 str., [in Slovak]

VOLNER, R., VOLNER, Ľ.: Basic security model of the airport, Journal of Multidisciplinary, Engineering Science Studies, Vol. 3, Issue 9, September 2017, pp. 2122-2127, ISSN 2458-925X, http://jmess.org/index.php/vol-3-issue-9-september-2017/

VOLNER, R., VOLNER, Ľ.: L410 Simulation – Initial Conditions, Flight Dynamics, Journal of Multidisciplinary Engineering Science Studies, Vol. 3, Issue 8, August 2017, pp. 1968 -1973, ISSN 2458-925X, http://jmess.org/index.php/vol-3-issue-8-august-2017/

VOLNER, R.: Modelovanie a simulácia, (Modeling and simulation), Verbum KU Ružomberok 2014, ISBN 978-80-561-0165-0, 209 str., [in Slovak]

Contact:

Ing. Ľubomír Volner,
University of Žilina
Adress Univerzitná 8215/1
010 26 Žilina
Tel. +421 902 301 224
Email: volner@stud.uniza.sk