Fakulteta za logistiko
Mariborska cesta 7
3000 Celje, Slovenija

UNIVERZA V MARIBORU
FAKULTETA ZA LOGISTIKO

Self-Adapted Meta-heuristics for automated generation of machine learning tools for modeling of complex dependencies and forecasting: Applications in the real and public sector considering the COVID-19 pandemic shocks

Research Project (Working Report 2020-2021)

Assoc. Prof. Dr. Dejan Dragan, Prof. Dr. Eugene Semenkin

Faculty of logistics UM

Reshetnev Siberian State University of Science and Technology

Celje, 24.05.2021

Univerza v Mariboru
Fakulteta za logistiko
Mariborska cesta 7
3000 Celje, Slovenija

Document: **Self-Adapted Meta-heuristics for automated generation of machine learning tools for modeling of complex dependencies and forecasting: Applications in the real and public sector considering the COVID-19 pandemic shocks**

Authors: Dejan Dragan, Eugene Semenkin

Type of document: Working report for the research project (the year 2020-2021)

Univerza v Mariboru
Fakulteta za logistiko
Mariborska cesta 7
3000 Celje, Slovenija

Content

1. Background of the research project

In many real applications, as much accurate forecasting as possible is of paramount importance. Namely, even slightly inaccurate forecasts can lead to wrong strategic decisions and wrong financial investments, which can have catastrophic financial consequences, such as inappropriate investments in infrastructure, bad choice of expensive management and control systems, and the like.

In year 2020, Reshetnev Siberian State University of Science and Technology (research investigator Prof. Semenkin), and Faculty of logistics, the University of Maribor (research co-investigator Assoc. Prof. Dejan Dragan) have conducted an agreement about joint research project titled: *Self-Adapted Meta-heuristics for automated generation of machine learning tools for modeling of complex dependencies and forecasting: Applications in the real and public sector considering the COVID-19 pandemic shocks.*

The project's main aim is dedicated to the modeling of complex dependencies in the context of forecasting models and the problem of optimal model selection. It turns out that finding the best model from the entire family of model candidates is not an easy task. Conversely, it requires a lot of mathematical knowledge from different fields and modeling skills and experiences. In many cases, the classical approaches (statistical, AI and machine-learning-based, data-mining based, etc.) to the process of optimal model selection do not give satisfactory forecasting results. Then, the hybrid and combined approaches are needed to improve the predictive power of obtained forecasting models. Nevertheless, sometimes even with such advanced approaches, the developed forecasting models do not reach adequate accuracy, and the necessity to use advanced optimization, and additional metaheuristics for optimal model selection is needed.

There is always a need to design such generic and automated modeling mechanisms that would be simultaneously, with slight modifications, helpful in various research areas and for different practical applications. Naturally, such a desire requires even more advanced modeling skills to achieve this objective. This research project mainly focuses on modeling and forecasting the complex dynamics of various throughputs and material flows via supply chains (SCs), road traffic and transportation, and seaports.

Univerza v Mariboru
Fakulteta za logistiko
Mariborska cesta 7
3000 Celje, Slovenija

Besides, modeling the relationships between the dynamics of the spread of the new coronavirus COVID-19 on one side and the dynamics of material flows running via supply chains must also be considered. For this purpose, the epidemic models from the area of mathematical epidemiology are also needed while discovering the complex relationships of the virus spreading on one side and consequential disruptions of the material flows dynamics via supply chains on the other side. Only when the influence of COVID-19 dynamics can be effectively and holistically linked with the dynamics of material flows running via supply chains, adequate predictions can be made and possible scenarios projected into the future. Since in the case of the total outbreak also the supply chains are going apart, it is very important to adapt them quickly to new circumstances so that their material flows are not severely interrupted or even broken.

This ongoing project consists of the following main stages:

1 Studying different forecasting model selection approaches to find the best model (statistical, fuzzy, neural, hybrid, etc.).

2 Modeling of the dynamics of the spread of the virus with epidemic models.

3 Modeling the dynamics of material flows running via supply chains with additional consideration of COVID-19 dynamics.

4 Possible reduction of the model selection problem into the optimization problem by means of self-adapted metaheuristics.

5 Deployment of developed forecasting models into different practical examples (e.g., Road Freight transport flows, Throughput cargo flows in ports, etc.).

6 Research on COV 19 and the medical supply chains (with particular emphasis on possible bottlenecks regarding the supply of masks, respirators, and mechanical ventilators).

7 Research on COV 19 and the social irresponsibility, mental state of the people, etc.

Until now, the following work has been already conducted within the scope of the project:

1 *DRAGAN, Dejan, MULEJ, Matjaž, KESHAVARZSALEH, Abolfazl. Research on pandemics and COVID-19 virus : a systemic review of methodologies focusing on holistic solutions in logistics and supply chain management. V: HRAST, Anita (ur.), MULEJ, Matjaž (ur.), GLAVIČ, Peter (ur.). Personal and social resposibility for sustainable future : IRDO International Conference : the 15th IRDO International Conference Social Responsibility and Current Challenges 2020 : Maribor, Slovenia, 4-5 June 2020.*

2 DRAGAN, Dejan. COVID-19, modeling, logistic systems, and social responsibility : a holistic overview : IWMMA 2020 - the 9th International Workshop on Mathematical Models and their Applications, held at Siberian Institute of Applied System Analysis, Krasnoyarsk, Russia, November 17, 2020. [COBISS.SI-ID 46843139]

3 DRAGAN, Dejan. Software for bibliometric analysis of social responsibility works during the COVID-19 pandemic. Celje: Faculty of Logistics, 2021. 1 CD-ROM. [COBISS.SI-ID 61062147].

4 DRAGAN, Dejan. Software for mathematical epidemiology and mathematical models to analyze the dynamics of the COVID-19 epidemic. Celje: Faculty of Logistics, 2021. 1 CD-ROM. [COBISS.SI-ID 61061891].

5 DRAGAN, Dejan. Software for optimal selection of ARIMA statistical models using fuzzy logic and Monte-Carlo simulations and an example of the choice of electronic toll system for trucks. Celje: Fakulteta za logistiko, 2021. 1 CD-ROM. [COBISS.SI-ID 61062659].

6 DRAGAN Dejan, Simona ŠINKO, Abolfazl KESHAVARZSALEH, Maja ROSI. Road Freight Transport Forecasting: a Fuzzy Monte-Carlo Simulation-Based Model Selection Approach. Technical Gazette. Accepted, in the publishing procedure.

7 DRAGAN Dejan, JANŠA Rado, IVANUŠA Teodora. Mathematical epidemiology and COVID-19. Nova Science Publishers. In the publishing process.

8. Dejan Dragan, Abolfazl Keshavarzsaleh, Matjaž Mulej, Bojan Rosi, Maja Rosi. Systemic Review on Social Irresponsibility Dimensions during the COVID-19 Pandemic. Economie e Management. In the reviewing process.

2. Some details about the research project

While solving different problems in real practice, let be in the private or public sector, one is often dealing with highly complex, nonlinear, stochastic, and dynamical processes. Since there is a significant need to mimic those processes for different purposes in many scientific disciplines, the construction of appropriate mathematical models is essential. The spectrum covering mathematical models is broad, from various optimization-based models, statistical and stochastic models, all over to different simulation models.

The model selection procedure to find the best model is one of the most essential steps during the modeling process (of a particular type of model). Models are, in principle, determined by their structure and parameters by using the identification (and estimation) procedure. Often, the structure and parameters can only be partially derived based on prior knowledge (first principles) and recorded historical data (grey modeling principle). Several criteria characterize the quality of the model's fit to the real data (with corresponding model's error), let be the information-based (e.g., AIC, BIC), residual-based (e.g., MASE, RMSE, MAAPE, sMAPE, UMBRAE, and other criteria), or statistical ones (e.g., number of statistically significant regressors due to parsimony principle (Occam's razor), or inveribility, stationarity and stability (for time series (TS) models)). Another important measure often used is the so-called ROC curve, Area under the curve (AUC), and confusion matrix. When searching for the best model among many model candidates, the optimal compromise must be achieved between various criteria that are often contradictory. At this point, the problem of searching for the best model can be converted (reduced) into the optimization problem. From here on, space opens for the use of different optimization approaches, from the classical ones to the modern techniques based on heuristics and metaheuristics.

There exist many modeling approaches, from the classical statistical ones, or those based on systems' dynamics, all over to modern AI approaches from the field of Machine Learning and Data mining (e.g., the neural models, fuzzy models, neuro-fuzzy models, support vector regression approach, and many others). It is often also recommended to mix different techniques, for example, to use one classical method (e.g., the statistical model to cover the linear part of the model), and the other that belongs to modern techniques (e.g., the neuro-fuzzy model to cover the nonlinear phenomena). Such kind of hybrid model usually provides better performance and ensures better model's fit to the real data. Within this scope, there is also a strong desire to accurately model and reveal complex relations between model's exogenous and endogenous variables that are comprised in the model's representation.

When the model is once derived and validated, it can be put into the real practice. In the case of predictive models, their main task is to forecast certain physical quantity, material flow, energy flow, or some other physical variable. However, one should be awared that during the online applications, the circumstances in the observed environment can quickly change since they are, in principle, also stochastic, dynamic, and exposed to the future uncertainty. Also, the observed process's parameters and other characteristics can change so that the identified model's structure and parameters are not valid anymore. In this case, the model should have been self-adapted and re-calibrated with respect to the process changes. Sometimes, there is even a need to deploy the regime-switching models to have several different hybrid models for different process situations. In such situations, to avoid human interventions and non-stop need of experts in behind, it might have been better to automate such regime changes. Thus, to develop a certain mechanism that would automatically choose the most appropriate model type for each particular process situation. The change of the model also requires the repetition of the optimization procedure (recall that it was mentioned that the model selection process can be reduced to the optimization problem).

When discussing optimization procedures, there is often a strong need to achieve global optimization and to avoid getting stuck in the local extrema. For this purpose, different nature-inspired heuristics are more and more important due to their efficiency, relatively fast computations, and other good properties. The basic metaheuristics include the genetic algorithms, simulated annealing, ant colony optimization, particle swarm intelligence, differential evolution, and so on. Particularly important are modern stochastic metaheuristics (or combination of several metaheuristics), that use different perturbation maneuvers to search across the search space quickly and as soon as possible find the feasible solution and global optimum.

The only problem with such metaheuristics is that they also demand a certain setup of their parameters (usually by the human expert), for which always exists a certain danger that the choice of the bad parameters might lead us to unsatisfactory solutions. Thus, the further idea is to automate the choice of most appropriate metaheuristics somehow as well and automatically optimally set new parameters for them without human intervention. Naturally, all automatic settings, let be for models or metaheuristics that drive the corresponding optimization, should be conditioned by information carried in the (new) data explaining the changes in the environment of the observed process. Finally, in this context, genetic or evolutionary programming also has quite a big potential. Namely, as a technique of evolving programs, it fits for a particular task by applying operations analogous to natural genetic processes to the population of programs and trying to find an optimal program among the space of all programs.

At a certain point in the research, it would also be perhaps worth including the control theory and the design of different controllers (e.g., fuzzy controllers) to automatically tune the parameters of different combined bio-inspired metaheuristics for solving different optimization problems. Those controllers would aim to conduct the automatic tuning of certain parameters of bio-inspired algorithms when there appears a need to deploy such intervention (control/regulation).

When traditional machine learning (ML) tools are applied to real-world problems, they can often be resource-intensive, requiring significant domain knowledge and time to produce and compare dozens of different model types. In this case, the so-called automated ML tools would likely be a better solution since they accelerate the time needed to get production-ready models in a very effective way. For this purpose, the parallel pipelines are created that try different algorithms and parameters. In each iteration, a new model is produced with a certain training score, while the models with the higher scores are more likely to have a better fit to the real data.

3. Applications and real examples

In this project, the described methodological research will be tested for a plethora of real examples in the real and public sector, considering the COVID-19 pandemic shocks. The main emphasis will be on the following areas:

1. **Modeling of the relationships between the dynamics of the spread of the virus on one side, and the dynamics of material flows running via supply chains**. The goal is to make good predictions regarding the possible scenarios projected into the future. Thus, more precisely, the main emphasis of the research directions here is to reveal the relationship between the modeling of the spread of COVID-19 virus dynamics on one side, and the development of the mathematical (statistical, simulation-based, forecasting, etc.) models of supply chains' material flow on the other side. The achieved results will help us to understand better how the virus will affect the global supply chains as a whole, as well as the individual SC members (e.g., Cargo Throughputs in the Seaports, Airports, etc.). Many industries have deployed product variety and outsourced manufacturing as their competitive advantage strategies to gain market share and cost advantage. Although such strategies seem to be effective during a normal environment, they can make industries' supply chain vulnerable to unpredictable disruptions when uncertainties arose due to epidemic spread. Such unpredictable disruptions have both immediate and sever hit on supply chain structures, networks, and components as they might create discontinuity of interactions between supply chain components (i.e., manufacturers, suppliers, distributors) and logistics operations (procurement, transportation, shipping, freight) temporarily. These, in turn, will lead to performance degradation and ripple effects, including fluctuations in demand and lead time, decreased productivity and service quality level due to delivery delays, price, and currency fluctuations and raw material shortage. Also, presumably, the antiglobalization will erupt. At the same time, the paths of the SCs will be shorter, and the market will likely reduce more to the regional (e.g., the EU) or even the national level. Thus, to summarize, for the purpose of research, different mathematical and statistical models should be developed here. Also, the multivariate statistical analysis (MSA) and the development of forecasting models should be included besides the modeling of pandemic spread. One of the main goals is to find out how the dynamics of Cov-19 will impact on the future material flow via the

global supply chains, e.g., throughputs in seaports, airports, etc. Regarding the methodology, also the Monte Carlo scenario-playing framework might have been included, as well as different kinds of simulations (continuous, discrete event, agent-based).

2. **Forecasting of different throughputs and material flows in traffic, transportation, and seaports.** The baseline here would be the research work and development of forecasting models already done. In the case of numerous influential exogenous variables (due to different geographic clusters and economic variables), the dynamic factor analysis and DF model (DFM), as a result, is recommended to lower the dimension of the input variables' space. The prefilter based on Monte Carlo simulation and principal component regression might have been also used to identify truly influential input variables. As additional input variables, the ones related to the COVID pandemics longitudinal dynamic model might have been additionally considered. During the model selection, the fuzzy logic might have been used to identify the best model as a compromise of different information-based, residual-based, and statistical criteria. Accordingly, the fuzzy-ARIMAX or fuzzy-DFM-ARIMAX model of the predicted throughput might have been conducted. In order to capture the nonlinear dynamics and phenomena, this model might have also been combined with an additional neuro-fuzzy (e.g., ANFIS) or neural (e.g., NARX) model.

3. **Descriptive statistics, comparison, multivariate statistical analysis (MSA), and statistical modeling with respect to some of the typical countries**: Italy, Spain, The US, the UK, France, Belgium, Sweden, Norway (Western Countries), Russia (the biggest country in the World), India (Central Asia), China, Malaysia, Singapore, Japan, Taiwan, S. Korea, Hong Kong (Far-East Asia), and Brazil, Mexico (Central, South America). Here, we should explore the data of the COV 19 total cases, deaths, tested people, recovered people (which left the hospitals), etc., on the national, regional, and municipal-county levels on one side. On the other hand, we should observe for all five aforementioned geographic clusters the following: A) (Predicted) GDP and GDP per capita; B) Level of purity; C) number of employed/unemployed; D) age clusters (e.g., 0-10, 10-20, 20-30,…,80-90), E) The number of available m2 per person; F) m2 of the surface of the municipality, county,…, state; G) An average (GROSS, NET) income per household; H) Number of people in county…; I) Number of rural/urban people in the county, and so on.

   Objectives, hypotheses, and dilemmas that must be investigated:

A). The people in regions with a higher density of population have more cases, deaths, etc.

B). The people in dense urban areas suffer more severe casualties (SMSC).

C). The people with a lower Number of available m2 per person SMSC.

D). The people with a lower GDP per capita SMSC.

E). The poor people with lower income and worse living conditions SMSC (example: in the UK, poor people suffered twice more casualties than rich people).

F). Why has Sweden sacrificed so many innocent victims for nothing? Namely, the prediction of the forthcoming GDP is similar, as in the case of Norway (-6% at least), but the latter has only 40 deaths/mil., while Sweden almost ten times more.

G.) Why is Belgium so hardly punished with the (biggest!) number of deaths per million, if they have appropriately conducted their lockdown timely?

H.) It would be interesting to conduct the MSA for many countries and observe the time-series ratio of the deaths(t)/recovered(t). Here, some fascinating patterns can be seen that should be investigated using data mining techniques. Also, some serious measurement bias can be detected, particularly during the weekends.

I.) Is it true that the Far-East Asian countries (see above) have sacrificed the economy to save the people, while the western countries have done just the opposite (sacrificing of the people to save the economy)?

J.) The time-periods from the first positive COV identification until the first serious reaction (lockdown) of the authorities were different for various countries. It would be perhaps interesting to make an MSA and compare the situation for different countries with varying rates of mortality. This way, the motives for the too-late response of some countries might have been revealed.

4. Research on COV 19 and the medical supply chains (with particular emphasis on possible bottlenecks regarding the supply of masks, respirators, and mechanical ventilators).

5. Research on COV 19 and the mental state of the people.

6. The statistical structural equation modeling (SEM) and/or logistic regression (LOGIT) modeling (for classification) related to socio-economic factors on one side, and COV 19 factors on the other side. Such models would interrelate the economic and social variables. In the case of the SEM models, the dependent (endogenous) variables might have been the number of cases and mortality rates. Constructs might have been social, economic, and COV-19 based.

7. Analysis of the dynamics of COV 19 and its modeling (+ simulation + gis if necessary) in conjunction with the modeling of stochastic processes of cargo throughputs and other mechanisms in logistics and transport.

8. Analysis of the dynamics of COV 19 and its modeling (+ simulation + gis if necessary) in conjunction with the modeling of stochastic processes of the movement of electric loads, oil prices, and other strategic commodities.

9. Analysis of COV 19 dynamics and its modeling (+ simulation + gis if necessary) in conjunction with the modeling of stochastic air pollution processes (environmental and sustainable issues).

4. Conceptual framework of the research

Figure 1 shows the conceptual framework of the key elements of the entire research project. The interested reader can investigate more in-depth details of given modeling mechanisms in our previous works ((Dragan, Keshavarzsaleh, et al., 2020; Dragan, Kramberger, et al., 2020; Dragan, Mulej, et al., 2020; Dragan et al., 2019; Hammad et al., 2020; Intihar et al., 2017)). Epidemic outbreaks undoubtedly lead to global chaos; thus, the catastrophic scenarios should be prevented at any price. Since in the case of the total outbreak also the supply chains are going apart, it is very important to adapt them quickly to new circumstances so that their material flows are not severely interrupted or even broken. The emergence of COVID-19 revealed a higher degree of disruptions at multiple levels in comparison to outbreaks that had happened previously. The reason is in today's interconnected world, where many socio-spatial, ecological, and economic factors play a role as catalysts for speeding epidemic outbreaks. Globalization, population mobility, population density, climate change, and urbanization are a few examples of such catalysts that accelerate the spread of communicable diseases.

The SC and logistic operations (LOs) are of utmost importance to controlling any pandemics, be it before, during, or after outbreaks. Therefore, the establishment of logistic operations' decision-support systems (DSS), evaluation of the current situation, forecasting the spread and dynamics of the outbreak, building an emergency SC mechanism, and deploying monitoring technologies are of central importance to controlling pandemics in today's interconnected world. The LOs, including transportation of humans and commodities, along with the deployment of temporary medical facilities, are recognized as the most important indicators of managing a disastrous crisis by WHO. In order to efficiently manage the epidemic and analyze and forecast its dimensions and dynamics, mathematical epidemic models play a central role. There is also a strong challenge to effectively and holistically link the influence of COVID-19 dynamics with the dynamics of material flows running via SCs to make the correct prediction of possible scenarios projected into the future.

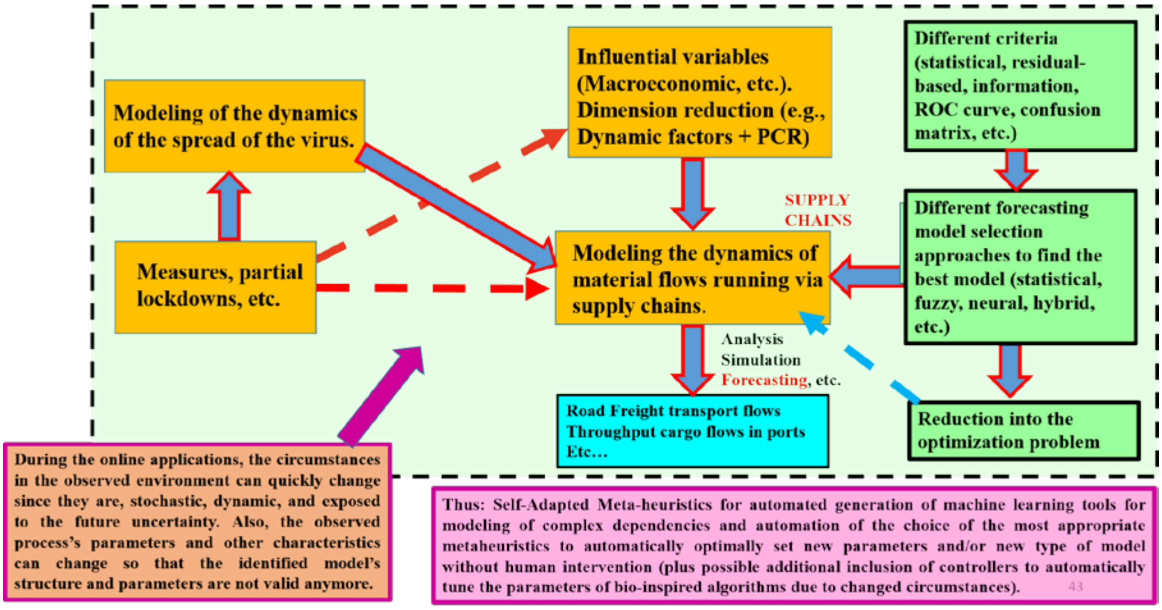

Figure 1: The conceptual framework of the key elements of entire research project.

It is desired to develop such model selection mechanism, which helps to combine the best properties of different criteria for selecting the optimal model, statistical, information-based, residual-based, and others. This way, the shortcomings of other approaches, which usually unilaterally consider only one category of criteria, can be omitted. Moreover, the possibility of genuinely choosing the optimal forecasting model becomes also increased. Thus, the resulting predictive power might lead to noticeably accurate practical results and, hence, the right decisions without too much risk of financial losses.

As shown in Figure 1, two types of modeling are trying to be combined, modeling the dynamics of the spread of the virus, and modeling the dynamics of material flows running via supply chains. In this context, all influential external factors must also be considered, from epidemic measures, partial lockdowns all the way to the other influential variables (such as socio-economic). When both types of mathematical models, epidemic and those of material flows, are successfully interrelated, different analyses, simulations, and forecasting can be applied. Moreover, the developed generic models can be deployed for different kinds of applications, such as forecasting of road freight transport flows, throughput cargo flows in ports, etc. Such developed types of models should also have an ability to predict and encompass possible disruptions in the SCs due to potential new lockdowns and other possible disturbances in uncertain environments.

During the online applications, the circumstances in the observed fuzzy environment can quickly change since they are stochastic, dynamic, and exposed to future uncertainty. Also, the observed process's parameters and other characteristics can change so that the identified model's structure and parameters are not valid anymore. When this happens, the developed models should be re-calibrated. These facts lead us to use an additional help of self-adapted meta-heuristics. They can provide automated generation of machine learning tools for modeling of complex dependencies and automation of the choice of the most appropriate metaheuristics to automatically optimally set new parameters and/or new type of model without human intervention. In some cases, possible additional inclusion of controllers to automatically tune the parameters of bio-inspired algorithms due to changed circumstances might have also been needed.

5. Research Case: Monte Carlo fuzzy logic based model selection algorithm to find the best ARIMA model to predict future transport in Slovenia to year 2030

The research case was a part of this research project. The main aim here was to develop a novel two-stage fuzzy-based model selection procedure, which is combined with the Monte-Carlo simulations (for details see (Dejan DRAGAN, 2021)). The proposed mechanism represents an alternative to the commonly used model selection approaches, where the best model is chosen based on observing, e.g., the Akaike's AIC information criterion or Bayesian BIC criterion (Akaike, 1973, 1974). In some cases, these criteria do not identify the most parsimonious model, thus they have certain deficiencies that our approach strives to overcome.

The Conceptual Framework with a novel two-stage fuzzy-based model selection procedure is shown in figure 2. In the first stage, for different combinations of orders p and *q*, the whole family of plausible ARIMA model candidates is generated with different levels of fit to the given freight transport historical data. For each of these candidates, various information, statistical, and residual-based criteria $X(i,j), j=1,\ldots,J$ are also calculated.

Afterward, in the second stage, the fuzzy model (FM) of type Mamdani (Mamdani Fuzzy inference system (MFIS)) is employed and implanted into a wider Monte Carlo Simulations

(MCS) framework. While designing the FM, $X(i,j), j=1,...,J-1$ are injected as inputs into the fuzzy model, while one criteria $Y(i)=X(i,J)$ represents its output. The primary assumption is that the best ARIMA model is "hidden" behind that index i^*, for which the inputs $X(i^*,j), j=1,...,J-1$ induce the closest fit of the FM's output $\hat{Y}(i^*)$ to the real one, i.e., the $Y(i^*)$. In order to verify this assumption, the $k=1,...,M$ MCS iterations are executed that produce M different random sets of weights $w_k(s), k=1,...,M, s=1,...,S$ loaded on S fuzzy rules $FR(s)$. This way, a sequence of M fuzzy models and their belonging errors $e_{FM}(i,k)=\hat{Y}(i,k)-Y(i)$ is generated for each *i-th* ARIMA model candidate. From these errors, a further analysis clearly isolates the "optimal" index i^* that belongs to the best ARIMA model. The designed FM-MCS model selection mechanism was first tested for the simulation example and then used in the real case of freight transport data. Prediction results show that the developed forecasting model can provide a well-fit to the real data.

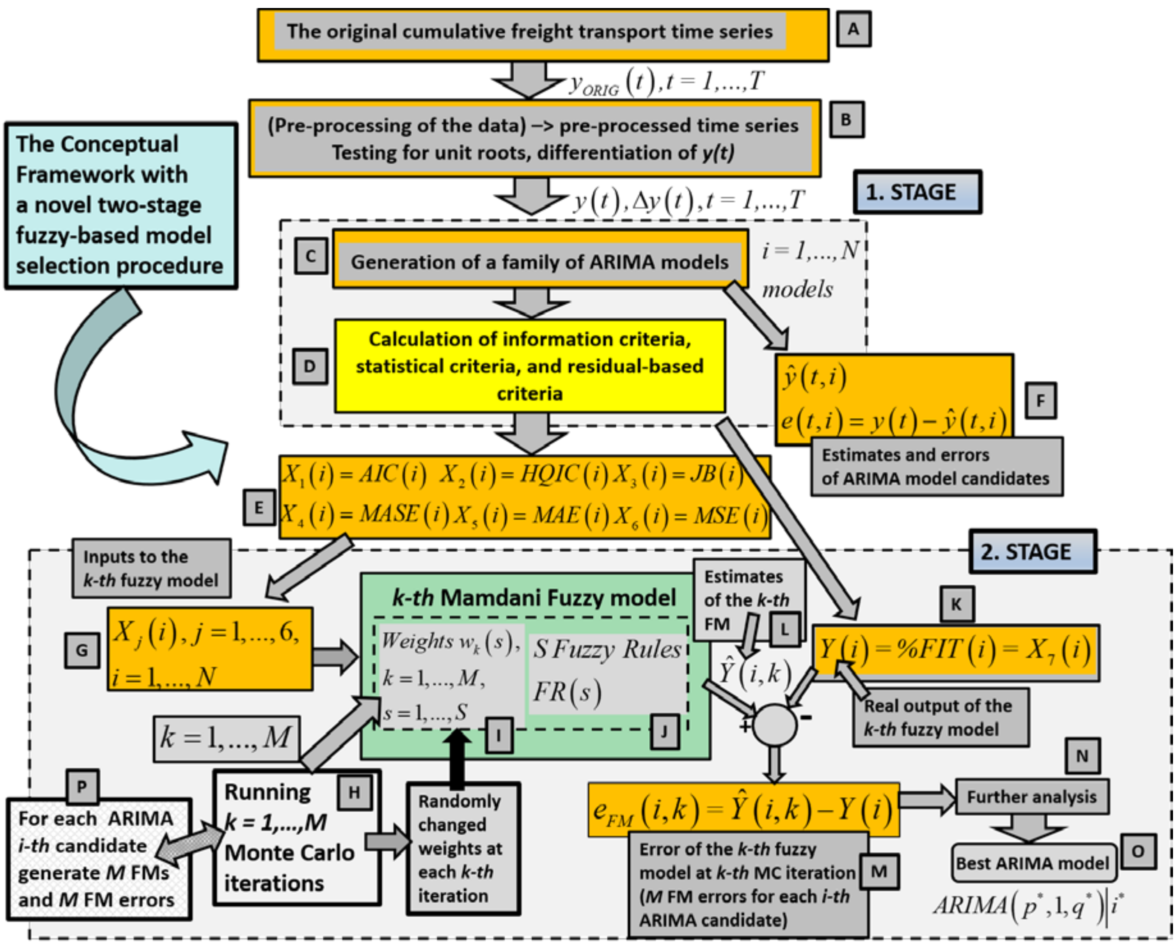

Figure 2: The Conceptual Framework with a novel two-stage fuzzy-based model selection procedure

In order to verify the capabilities of the developed mechanism, it was first tested for the case of simulation example. For this purpose, the simulated $ARIMA(2,1,1)$ process was taken into consideration. In addition, the family of 36 $ARIMA(p,1,q), p=1,...,6, q=1,...,6$ different model candidates was generated. The output of the simulated process was intentionally corrupted with a nonlinear threshold auto-regressive (TAR) noise of the significant magnitude. Despite the substantial noise, the designed FM-MCS mechanism has undoubtedly managed to correctly estimate the exact model's orders $p^*=2, q^*=1$, while the parameter estimates were reasonably close to the real values of the simulated parameters. On the other side, if only the observations of the AIC measure were taken into account, the model selection mechanism has failed to estimate the exact orders and parameters of the simulated process resulting in incorrect identification results.

The next step was to apply the developed mechanism for forecasting of real freight transport data. Figure 3 shows the 100 quarterly historical time series data for Slovenian freight transport, measured over the period from 1990 up to 2014. For the modeling purposes, the data were appropriately pre-filtered and pre-processed (see (Dragan et al., 2019) for details). From Figure 3 it can be noticed that freight transport significantly increased after the year 2004 when Slovenia entered into the EU. During the time of economic crisis, an oscillatory behavior can be noticed with a significant drop that has resulted in a local minimum reached in the third quarter of the year 2009. Later, the time series was stabilized with a slight drift that was present until the year 2014. The data in the forthcoming years are not shown in figure 3.

Univerza v Mariboru
Fakulteta za logistiko
Mariborska cesta 7
3000 Celje, Slovenija

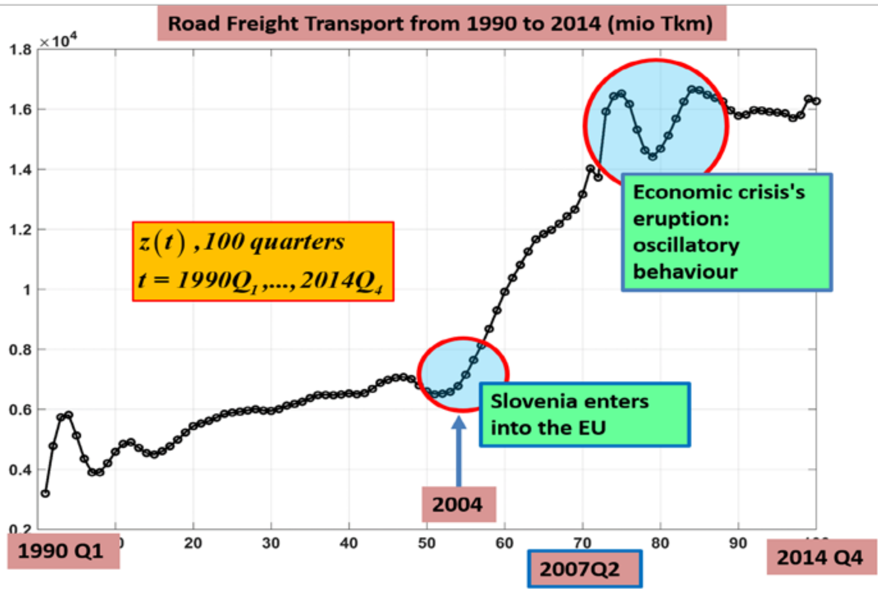

Figure 3. The quarterly data for road freight transport in Slovenia from 1990 to 2014.

The modelling process and all other calculations were carried out in the MATLAB technical computing environment, with the additional use of the three MATLAB's toolboxes, i.e., the Fuzzy Logic Toolbox, the Econometrics Toolbox, and the Statistics and Machine Learning toolbox.

The MCS procedure in figure 2 ($k = 1,...,M = 1000$ iterations) has created an entire family of M different MFIS models with corresponding $e_{FM}(i,k) = \hat{Y}(i,k) - Y(i)$ FMs errors. The next step was to carry out further analysis to obtain the best $ARIMA(p^*,1,q^*)|i^*$ model and corresponding optimal index i^* by means of designed model selection mechanism. For this index, the number of occurrences $N_{\max}(i'^*,k)$ of the minimal FMs errors $min[e_{FM}(i',k)]$ should have been maximal if compared with the other $ARIMA(p,1,q)|i$ models.

The modelling mechanism has calculated the histogram of occurrences of the minimal FMs errors $N\{min[e_{FM}(i',k)]\} = N(i',k)$ shown in figure 4, as well as the corresponding point diagram $min[e_{FM}(i',k)]$ shown in figure 5. A careful observation of Figures 4 and 5 convinces us that for the far biggest number of MCS iterations (624 times) have happened that the maximal number of occurrences of the minimal FMs errors $N\{min[e_{FM}(i',k)]\} = N(i',k)$ had occurred for the index $i'^* = i^* = 12$, which means $N_{\max}(i^*,k) = N_{\max}(12,k) = 624$. On the other side, the frequency of occurrences of the minimal FMs errors for the other indices i' was significantly lower, e.g., $N(2,k)=76, N(4,k)=104, N(9,k)=59, N(10,k)=51$, etc.

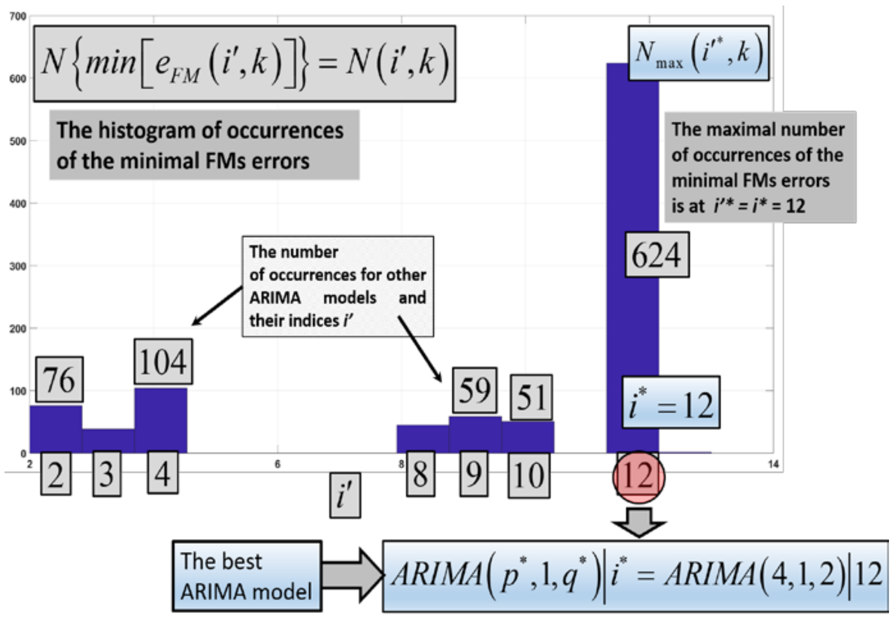

Figure 4: The histogram of occurrences of the minimal FMs errors $N\{min[e_{FM}(i',k)]\}=N(i',k)$; $k=1,\ldots,M=1000$.

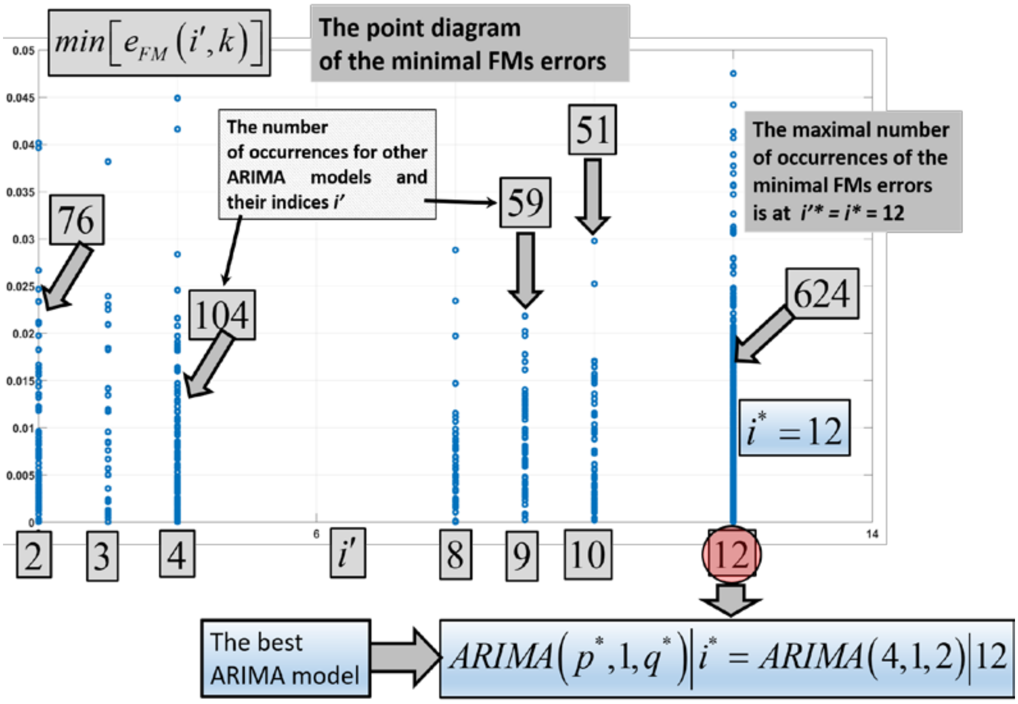

Figure 5: The point diagram $min\left[e_{FM}\left(i',k\right)\right]$ of the minimal FMs errors for different indices i' and k; $k=1,...,M=1000$.

Based on these facts, the following best ARIMA model can be isolated with the detected optimal index $i^*=12$:

$$ARIMA\left(p^*,1,q^*\right)\Big|i^* = ARIMA(4,1,2)\Big|12$$

Accordingly, the following form of the structure of the best model appears:

$$\hat{y}(t)=\frac{\hat{\theta}\left(B,q^*\right)}{(1-B)\cdot\hat{\phi}\left(B,p^*\right)}\cdot\varepsilon(t)=\frac{\hat{\theta}(B,2)}{(1-B)\cdot\hat{\phi}(B,4)}\cdot\varepsilon(t)=\frac{\left(1+\hat{\theta}_1\cdot B+\hat{\theta}_2\cdot B^2\right)}{(1-B)\cdot\left(1-\hat{\phi}_1\cdot B-\hat{\phi}_2\cdot B^2-\hat{\phi}_3\cdot B^3-\hat{\phi}_4\cdot B^4\right)}\cdot\varepsilon(t)$$

for which the estimated values of parameters, their standard errors, t values, and p values are depicted in table 1 ($\hat{\phi}_i \to AR(i);\hat{\theta}_i \to MA(i)$).

Table 1: The estimated parameters and their significance.

Parameter	Estimate	Standard error	p - value	t- value
AR(1)	0,55484	0,16745	0,001324	3,3134

AR(2)	0,52508	0,15787	0,001272	3,3260
AR(3)	-0,47366	0,14034	0,001086	-3,3751
AR(4)	0,25811	0,12265	0,038027	2,1053
MA(1)	0,20872	0,11867	0,081961	1,7589
MA(2)	-0,74033	0,19931	0,000351	-3,7145

As can be seen from table 1, all the parameters are statistically significant at $\alpha = 0,05$ level (t values bigger than 1,96), except the parameter MA(1) (t value is 1,7589) that is significant at $\alpha = 0,07$ level.

Figure 6 shows the predictive performance of the identified best $ARIMA(4,1,2)|12$ model. As can be seen from figure 6, the model provides surprisingly good forecasts $\hat{y}(t,i^*)$ of the real road freight transport time series $y(t)$ (measured in Tones km). There can be noticed some very sophisticated details of the TS dynamics, which were not entirely captured by the model. The reason is that the freight transport data dynamics likely contains a complicated nature with perhaps even a certain kind of nonlinear behaviour included. Despite this, the model generally provides an encouragingly good fit to the real data, particularly considering the key movements of the time series trend.

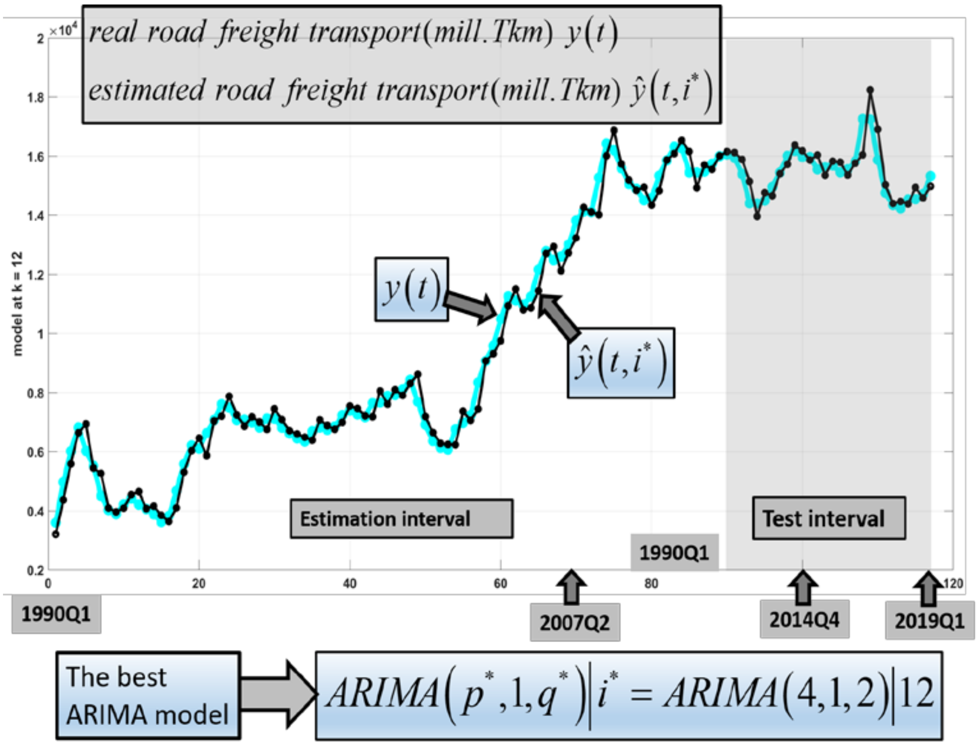

Figure 6: Forecasting results for the road freight transport data in Slovenia (Estimation interval–First 89 samples; Test interval – Last 28 samples); Tkm = Tones km.

The model's predictive performance should also be analytically verified by calculating the achieved measures of the quality of the derived forecasting model. Besides the well-known classical measures (e.g., the MAE, MSE, AIC, BIC, etc.), there have been many new measures applied in the last decade, which had become significantly attractive during the well-known Makridakis competitions (Dragan, Keshavarzsaleh, et al., 2020). Table 2 depicts the meaning of some of these new applied measures (Dragan, Keshavarzsaleh, et al., 2020):

Table 2: The important measures and their requirements.

Measure	**The meaning of measure**	**Requirement**
MAPE	Mean – Absolute – Percentage – Error (%)	low!
sMAPE	Symmetric – Mean – Absolute – Percentage – Error (%)	low!

MAAPE	Mean – ArcTangent – Absolute – Percentage – Error (%)	low!
MASE	Mean – Absolute – Scaled - Error	<1!
GMRAE	Geometric – Mean – Relative – Absolute – Error	<1!
MdRAE	Median – Relative – Absolute – Error	<1!
MBRAE	Mean – Bounded – Relative – Absolute – Error	
UMBRAE	Unscaled – Mean – Bounded – Relative – Absolute – Error	<1!
<1! → better than benchmark method (simple, usually a random walk)		

Measures such as (*sMAPE, MAAPE, MASE, etc.*) can be treated as the most reliable and were also used to check the model's quality. Table 3 shows the calculated measures for the case of the best ARIMA model in order to verify its quality and performance. Some of achieved values in table 3 are compared with the prescribed thresholds recommended in the literature. Computed measures confirm the adequate quality and performance of the derived model.

Table 3: Measures of the quality of the derived forecasting model.

Measure	**Value**
AIC	1753,8
BIC	1773,1
GMRAE*	0,9562 < 1
MdRAE*	0,85155 < 1
MBRAE*	0,47796
UMBRAE*	0,91558 < 1
UMBRAE_better* (%)	8,4422% > 0%
MAPE* (%)	3,5944 < 10%
sMAPE* (%)	3,5975 < 10%
MAAPE* (%)	3,5873 < 10%
MSE*	1,6808e+05
RMSE*	409,98
MAE*	320,53
MASE*	0,81125 < 1

Stand. Deviation*	411,72
SKEWNESS*	0,14743
KURTOSIS*	3,7056
JB VALUE*	2,8508 < 5,09 (critical value)
%FIT*	92,829 (a well-fit)

* These measures address the model' error.

Despite the fact that designed FM-MCS mechanism that gives the best ARIMA model works surprisingly well and provides a well-fit to the real road freight transport data, its prototype is currently developed for short to medium term forecasting only. Thus, similarly as in the previous research (Dragan et al., 2019), it might have been appropriate to combine the FM-MCS algorithm with a Monte Carlo Scenario Playing (MCSP) mechanism for generating future scenarios about the road transport trends. The MCSP mechanism might have represented the basis for calculating the long-term interval-type forecasts, for which the corresponding probability intervals in future time-points could be computed. Besides, there are likely many other practical examples and cases not only in the infrastructure and road transport planning sphere, but also in many other theoretical and practical fields, where the designed FM-MCS mechanism might have offered an opportunity to the scholars and practitioners to facilitate the model selection procedure.

6. The role of epidemic models from the field of mathematical epidemiology

Forecasting models that incorporate several parameters, features, and variables to predict epidemic outbreaks, mortality, and recovery rates, along with social life continuity during pandemics, are of essential importance. A wide range of mathematical modeling approaches has been deployed that can be categorized as time series, compartmental, agent-based, and metapopulational approaches. Here, particularly the compartmental models play an important role. The latter are well-studied and developed and have been successfully used for many real cases of epidemics in the past. Compartmental modeling approaches categorize the population as typical compartments according to different states of the diseases. The deployed examples of such epidemic modeling approaches are various such as for example: Susceptible–Exposed–Infected–Recovered (SEIR); Susceptible–Infectious–Recovered-Died (SIRD); the SIR modeling, etc.

These approaches follow the simple mechanism of defining multiple compartments to classify subpopulations, using the Branching or Combination process. An important classification of mathematical models can also encompass the deterministic-based and stochastic-based models (see figure 7) (Dadlani et al., 2020). Stochastic models are important for situations when the deterministic models cannot successfully enough describe the dynamic behavior of the real epidemics. They can be basically divided into the discrete-time Markov chains (DTMC), continuous-time Markov chains (CTMC), and stochastic differential equations based models.

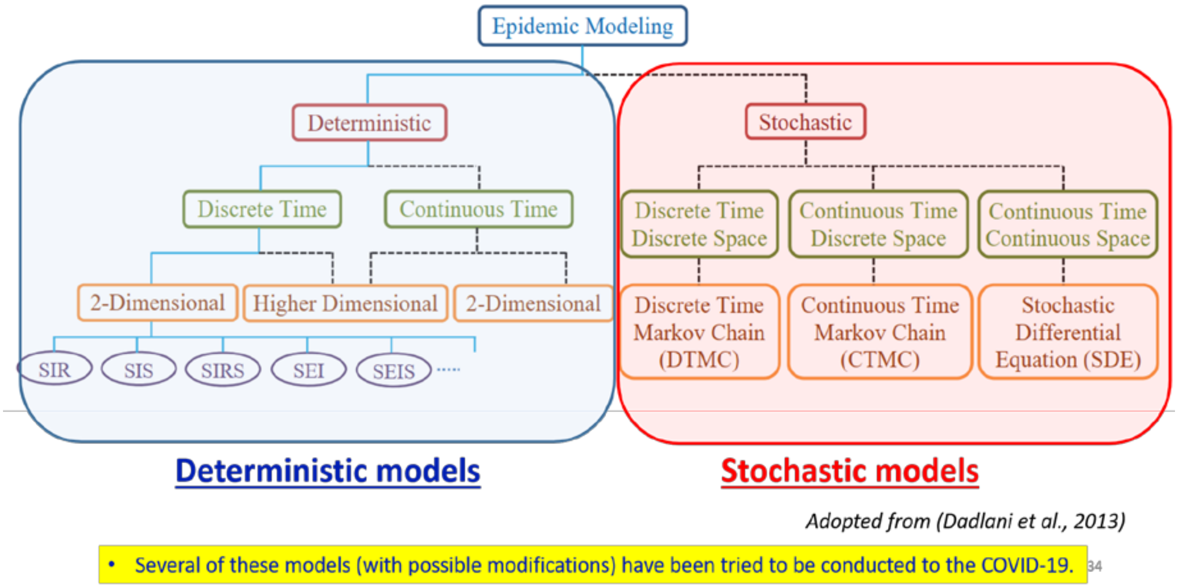

Figure 7: Classification of epidemic models.

The mathematical models, let be deterministic or stochastic, are not important only for the prediction of epidemic spread, but they can be useful for other purposes as well. For instance, by means of the different analyses deployed based on these models, the current progress of epidemic can be more in-depth analyzed, identification of possible failures in the fighting strategy against the epidemic can be detected, or the future strategies of vaccination and/or (partial) lockdowns can be constructed. Moreover, by using analysis based on these models, different scenarios can be played, and stability points of the nonlinear dynamic behavior can be isolated and studied.

All in all, the primary role of the models is to help decision-makers to identify the fragile optimal compromise between the strength of the lockdown on one side (to save as many lives as possible), and simultaneously keep the economy open to some level (to prevent too devastating economic damage). The aforementioned mathematical models can be assigned to the scientific discipline of mathematical epidemiology. This discipline, with its models, plays an essential role in the time of the current COVID-19 pandemic as well. A diverse range of mathematical models is being developed to forecast or to capture the dynamics of COVID 19 pandemics globally. Such efforts have substantially contributed to practical endeavors of curbing the current ongoing pandemic. As crucial as these endeavors are, one should deduce that the application of mathematical models in pandemics modeling is to reduce the burdens imposed by COVID 19 rather than generating precise models relevant to expanding the body of knowledge.

The reason is, as the dynamics of COVID 19 are going to become more and more complex, mathematical modelers should pay more attention to the transferability of their approaches; thereby, consumers of such models (e.g., policymakers) will benefit accordingly. The last major pandemic (Spanish flu) occurred already 100 years ago, and the more recent epidemics in this century (e.g., SARS, MERS, etc.) have been timely stopped. Maybe these facts were the reason that humankind, together with its experts, have forgotten of how devastating pandemic can be and how can mathematics plays a crucial role in lowering down all consequences and casualties.

Thus, it is not surprising that the present COVID-19 situation shows that the different experts have still underestimated the important role and usefulness of mathematical epidemiology. Is can be clearly seen that not only the key decision-makers (politicians) do not understand the meaning of projections of mathematical models, but even the experts from various areas (physicians, virologists, epidemiologists etc.) do not possess enough knowledge to truly understand the practical implications that can be deduced from mathematical epidemic models.

Celje, Krasnoyarsk, May 2021
Written by:
Assoc. Prof. Dr. Dejan Dragan
Prof. Dr. Eugene Semenkin
(Research Co-Investigators)

7. Literature

Akaike, H. (1973). Information theory and an extensión of the maximum likelihood principle. *International symposium on information theory*(1973), 267-281.

Akaike, H. (1974). 'A new look at the statistical model identification'. *IEEE Transactionson Automatic Control*, *19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Dadlani, A., Afolabi, R. O., Jung, H., Sohraby, K., & Kim, K. (2020). Deterministic Models in Epidemiology: From Modeling to Implementation. *arXiv:2004.04675 [physics, q-bio]*.

Dejan DRAGAN, S. Š., Abolfazl KESHAVARZSALEH, Maja ROSI. (2021). Road Freight Transport Forecasting: a Fuzzy Monte-Carlo Simulation-Based Model Selection Approach. *Technical Gazette, In press*.

Dragan, D., Keshavarzsaleh, A., Intihar, M., Popović, V., & Kramberger, T. (2020). Throughput forecasting of different types of cargo in the adriatic seaport Koper. *Maritime Policy & Management*, *48*(1), 19-45. doi:10.1080/03088839.2020.1748242

Dragan, D., Kramberger, T., & Popović, V. (2020). OPTIMIZATION METHODS AND HEURISTICS AND THEIR ROLE IN SUPPLY CHAINS AND LOGISTICS. In M. Vidović (Ed.), *Quantitative methods in logistics* (pp. 139-160). Belgrade: Faculty of Transport and Traffic Engineering.

Dragan, D., Mulej, M., & Keshavarzsaleh, A. (2020). *Research on Pandemics and COVID-19 virus: A systemic review of methodologies focusing on holistic solutions in logistics and supply chain management*. Paper presented at the IRDO 2020.

Dragan, D., Popović, V., Keshavarz, A., Jereb, B., & Kramberger, T. (2019). *FORECASTING FUTURE TRENDS IN FREIGHT TRANSPORT IN SLOVENIA UNTIL THE YEAR 2030*. Paper presented at the LOGIC.

Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logistics & Sustainable Transport, 11*(1), 51-76. doi:10.2478/jlst-2020-0004

Intihar, M., Kramberger, T., & Dragan, D. (2017). Container Throughput Forecasting Using Dynamic Factor Analysis and ARIMAX Model. *PROMET - Traffic&Transportation*, *29*(5), 529-529. doi:10.7307/ptt.v29i5.2334