



An Investigation of Time Series Models for Forecasting Mixed Migration Flows: Focusing in Germany

Vasiliki Mebelli¹ · Maria Drakaki¹ · Panagiotis Tzionas²

Received: 14 February 2022 / Accepted: 1 March 2023 / Published online: 10 April 2023
© The Author(s) 2023

Abstract

Refugee and migrant (mixed migration) flows in the Mediterranean have been in the spotlight of both policy and research, especially since 2015. Mixed migration is a volatile international phenomenon with considerable and debatable impacts on society and economy. This paper investigates the performance of time series forecasting methods based on EUROSTAT datasets focusing on asylum seekers. Germany has been selected to reflect on the ability of the models to predict the future behavior of an extremely volatile migrant mobility. Exponential smoothing and autoregressive integrated moving average (ARIMA) models have been used for the forecasting of asylum seekers. Monthly records of first-time asylum seekers have been used from January of 2008 up to September of 2020. The results demonstrate clearly that more research is needed on this field, taking into account the complexity of the characteristics of international migration, in order to assist decision-making in migration management.

Keywords Mixed migration · Europe · Time series · Forecasting · Exponential smoothing · ARIMA · Asylum seekers

✉ Maria Drakaki
mdrakaki@ihu.gr

Vasiliki Mebelli
vmebelli@ihu.edu.gr

Panagiotis Tzionas
ptzionas@ihu.gr

¹ Department of Science and Technology, University Center of International Programmes of Studies, International Hellenic University, 14thKm Thessaloniki, 57001 Themi, Nea Moudania, Greece

² Department of Industrial Engineering and Management, International Hellenic University, P.O. Box 141, 57400 Thessaloniki, Greece

1 Introduction

Global migration is a highly complex and dynamic phenomenon. Moreover, European policy makers have focused their attention on irregular migration. Migration policies were reassessed in order to cope with the migration flows arriving in Europe and in particular in the Mediterranean since 2015. The migration-related terminology is expanded in order to depict the characteristics of these flows. Accordingly, “mixed migration” was proposed to describe this category consisting of various types of migrant mobility. According to the United Nations High Commissioner on Refugees (UNHCR), “mixed movements” or “mixed migration flows” is complex and difficult to predict and monitor migratory population movement including refugees, asylum seekers, internally displaced, stateless people, economic migrants and other migrants, often moving irregularly (e.g. without the appropriate documentation for their movement), and for numerous reasons [1].

According to UNHCR, asylum seekers are the people petitioning for international protection, however have yet to be processed. At the end of 2019, 4.2 million people were asylum seekers worldwide [2]. People may seek asylum for a variety of reasons: reasons related to international protection; reasons that may emerge en route, such as to obtain access to a country’s immigration policies, political or economic insecurity; or the decision may be taken upon arrival to the destination country.

Migration constitutes the most unpredictable demographic change. On top of that, asylum-related migration constitutes the most unstable and sensitive type of migration. Therefore, the prediction of asylum-related migration constitutes a difficult task. Asylum seekers are forced to leave their countries for different reasons including war, conflict, violence, or persecution. The choice of a final destination country is greatly influenced by access to information. The “push–pull” factors have been taken into account into theories of forecasting asylum-related migration. Data accuracy issues as well as accurate measurement of uncertainty are main challenges in international migration forecasting [3]. Evaluation and assessment of various forecasting models showed that extrapolation of time series and Bayesian expert-based models were found to be applicable on asylum-related datasets [3].

This paper aims to investigate the performance of time series forecasting based on exponential smoothing and ARIMA methods, in regard to asylum-related migration to Europe and more specifically to Germany. European Statistical Office (EUROSTAT) datasets of first-time asylum seekers have been used, consisting of monthly data from January of 2008 to September of 2020 [4].

In the following, a review of the literature relevant to international migration theories including theories focused on forced migration is presented. Next, the methodology is proposed, followed by the section of corresponding results. Finally, conclusions and topics for future research are presented.

2 Literature Review

Theories related to international and forced migration will be reviewed in this section. A main category of international migration theories is the micro-level approaches for various reasons. Although neoclassical economics showed a potential theoretical framework, its fundamental principles were quite unrealistic [5-7]. According to the neoclassical theory, migrants are a homogenous group motivated by financial benefits when making their decision. More recent studies have focused on the explanatory contribution of a network theory. According to the network theory, individuals are influenced when making decisions to migrate by the fact that family members or other migrants have selected a specific country to migrate [8]. Interestingly, Docquier et al. [9] argued that potential migrants are influenced by the size of the network of previous migrants and the average income in the destination country. Yet the main factor that influences a potential migrant to decide to migrate is the economic growth of the destination country.

Both physical and mental security are taken for granted in the above theories. Yet forced and international migration are driven by reasons such as violence, persecution, or extreme poverty. Accordingly, academics point out that existing international migration theories are not appropriate theories to explain forced migration due to the fact that the former is voluntary movement and the latter is not [10].

The macro-level approaches constitute another main category of migration theories. Forced migration has been presented as a result of globalization and expressed through a North–South inequality in [10], using a social framework for analysis. According to the authors, North countries have put many restrictions on refugee-to-be criteria, excluding “environmental refugees” as well as families of internationally displaced persons from their countries of origin due to development projects. In Europe, the country laws for potential migrants aiming to attract explicitly highly skilled personnel have been seen as desirable compared to other groups of labor force. In [11], an experiment was conducted and submitted by voters of 15 European countries aiming to categorize the traits that an asylum seeker should possess to be approved as refugee by each individual voter. The results showed that age, host-country language fluency, occupation rank, and religion affected the decision of the voters.

Some literature studies have followed the approach of “push–pull” forces for the development of the various migration theories towards receiving and destination countries. This approach assisted scientists and policy makers to acquire a more representative picture of migration with the goal of making predictions. The first theory that tried to bridge the “pull–push” forces was introduced by Lee [12]. Accordingly, he presented the factors that shape mobility: factors associated with the country of origin, factors associated with the destination country, the obstacles en route, and factors associated with the characteristics of each individual. However, “push–pull” theories are too simplistic and treat individuals as a homogenous group that reacts in the same way to the same stimuli. Furthermore, another simplification of the “push–pull” theories relates to the concepts of peace and violence, treated as polar opposites, with values equal to 0 and 1, respectively.

Additionally, research on human migration network modelling is growing, aiming to quantify the migration flows as well as the impact of various policy regulations [13, 14]. In [13], multiclass and multipath refugee migration networks were developed with routes that captured congestion. In [14], a multiclass human migration network was formulated and analyzed, considering policy interventions to moderate the flow of migrants as well as enhance societal welfare.

The complexity of forced migration being an extremely irregular phenomenon can be attributed as the reason that inaccurate forecasts using probabilistic or deterministic models were produced by numerous studies. Forecast and uncertainty go hand-in-hand [15]. Disney et al. [3] categorized the most widely used probabilistic forecast models for migration into time series, expert-based, Bayesian, and econometric models with covariate information and historical error models. Furthermore, after testing all the above methods on different migration data, the conclusion was that only two of them are applicable on asylum-related datasets, namely extrapolation of time series and Bayesian expert-based models.

Time series is a sequence of numerical data points in successive order. The main purpose of using time series is to make predictions. Time series modeling relies on past data, and because data can be imperfect, it introduces bias and forecast errors [16]. It is important for the model to identify a possible trend and seasonality, yet shocks such as a financial crisis cannot be added which may lead to a false prediction. Time series extrapolations may be applicable for asylum-related migration and may be applied to other forms of migration prediction. Asylum-related migration can be considered more often as a non-stationary process, therefore requiring special handling with respect to forecast time horizon [17]. In [17], the general migration forecast horizon was proposed to be up to 100 years; however, in cases such as asylum-related migration, it was urged that the time horizon should be up to 1 year and could extend to 5 or 10 years [18].

The main characteristic of time series analysis is that observations are correlated and if their pattern is studied then inferences can be drawn about underlying phenomena. Acknowledging the dynamic and static components of a system as well as their interdependence assists in comprehending the system's mechanism and, eventually, estimates its next phase with a level of uncertainty. This uncertainty is observed in stochastic time series where the future values have a probability distribution conditioned by a knowledge of past values. Time series models can include trend, seasonality, and randomness components. A system can exhibit one or more components [16].

3 Methodology

Accurate forecasting depends on appropriate data. National, European, and international agencies track almost every quantitative part of everyday life for data-driven decision-making. Internal European statistics are provided by EUROSTAT, European Asylum Support Office (EASO), and National Agencies, as well as external ones such as UNHCR and International Organization for Migration (IOM). Yet many challenges are posed throughout the process. EASO published a methodology

for evaluating potential data sources [19]. The evaluation was based on six criteria, namely, frequency, definition, coverage, accuracy, timeliness, and quality assurance. A database that would meet those criteria perfectly would publish statistics reflecting the definitions agreed on asylum-related migration with no errors. Nevertheless, such a perfect database does not exist.

It is true that modern database systems offer new and extensive functionalities in terms of data visibility, validation, and data insights. Additionally, this is further enhanced by managed database services that are provided by big cloud providers. Though these functionalities and services enable users to gain access to hidden knowledge in their data, it still does not combat several problems that are identified in the existing data.

Our comment regarding the “perfect database” reflects to the definition of EASO. More specifically, we could argue that by utilizing the nuances on database management systems, one could extract and produce the evaluation metrics that the EASO proposes as criteria for the database evaluation; by then analyzing those criteria, it is clear that the existing databases rarely reach high values that impact the quality and quantity of the data. This data deficiency leads to our statement that “a perfect database” does not exist.

In conclusion, modern database systems as well as managed database services can enhance the ability to analyze and interpret data but cannot generate the data gap that exists in terms of quality data in the existing databases so far.

In this paper, EUROSTAT has been chosen as the preferred data source [4]. EUROSTAT publishes monthly data on asylum-related migration and therefore provides an increased sample size. The final sample set consisted of 153 observations, from January of 2008 to September of 2020 [20]. Germany has received the majority of asylum claims since 2015 and therefore was selected in this study.

The original data consisted of monthly asylum and first-time asylum applications by citizenship, age, and sex from January 2008 till September 2020 (153 observations). The sex profile consists of two genders (male and female), and the country of citizenship of the applicant was not restricted in any way. The topic of interest was new records of asylum-related applications. The overall number of applications depends on new and pending asylum cases. Only first-time asylum applications were taken in consideration. The measure of prediction accuracy was selected to the mean absolute percentage error (MAPE) due to its popularity and simplicity [15]. An upper limit has also been set to 45% MAPE [20].

3.1 Exponential Smoothing Methods

The main characteristic of exponential smoothing methods is that recent values hold more predictive power and importance than past ones. They assign exponentially decreasing weights as the observations get older. They combine error/residuals, trend, and seasonal components in a smoothing calculation. Those components can be combined either additively or multiplicatively. The model represents a weighted average of the observations. In R programming language, the model can be used with the forecast package. The general form of every possible exponential

smoothing model is of the form $ETS(X, Y, Z)$ with E corresponding to the error, T to trend, and S to seasonality. Hence, $ETS(X, Y, Z)$ correspond to the simple exponential smoothing method with multiplicative errors. This would be the simplest case where no trend or seasonality seems to be present.

Exponential smoothing methods combine the three components in a smoothing calculation. There are three smoothing parameters, namely, α , β , and γ , and their existence depends on how many of the time series components are present in the specific time series. In the simplest case with no trend and seasonal component, there is only one smoothing parameter (α); however, when all the components are present, all smoothing parameters are included. The smoothing parameter in each case is between 0 and 1, and the closer the value of the parameter is to 1, the heavily the time series relies on recent data.

3.2 ARIMA

ARIMA models are considered the most general class of models for forecasting non-stationary time series. It is a class of models that support that previous observations can be used alone to predict the future values.

In R programming language, the package `forecast` will be once more the tool to test whether an ARIMA model is a good fit for the time series. Generally, the form is $ARIMA(p, d, q)$ where p corresponds to the order of the auto-regressive part, d to the degree of first differencing, and q to the order of the moving average part.

4 Results

The number of annual first-time asylum applicants since 2008 in Germany is shown in Fig. 1. In the beginning of 2014, the number of asylum claims is shown to increase with its peak being around 2016. The main reason lies behind the closing of the Balkan and Mediterranean routes that asylum seekers used to reach other European countries besides the country of first arrival.

The actual number of asylum claims per month, the predicted numbers obtained from the simple exponential smoothing method, and the error between the two is presented in Table 1. Figure 2 shows the time series graph representing monthly first asylum claims in Germany and the predicted values using simple exponential smoothing.

The simple exponential smoothing method held the last observation of the train dataset, applied a smoothing parameter, and kept it unchanged throughout the time horizon of 9 months. This function of the model can be proved efficient in times of relative stability. However, if the last observation in such a model is a product of anomaly such as an economic recession, the prediction will be subject to systematic error. Consequently, its major limitation lies in the fact that the selected time horizon is not taken into account. Especially, in a phenomenon linked with frequent fluctuations such as forced migration, the method can be replaced with more complex models. In the

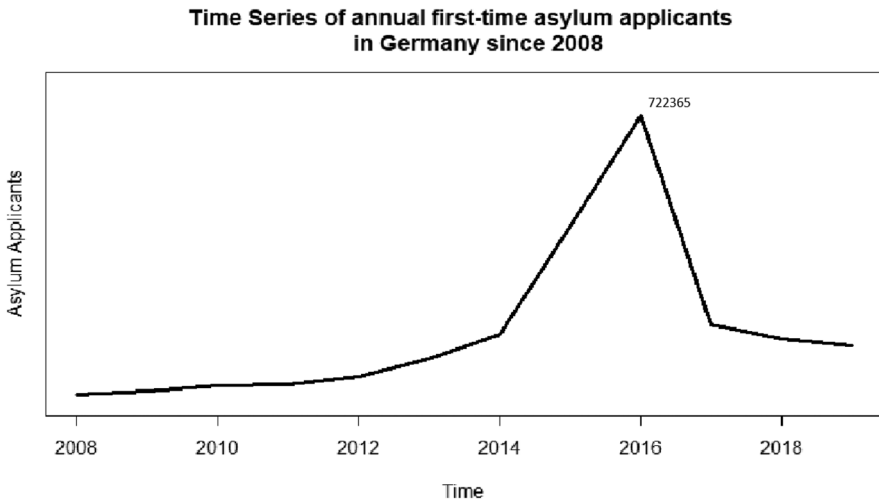


Fig. 1 Annual first-time asylum applicants in Germany based on EUROSTAT

Table 1 Results of simple exponential smoothing time series method in Germany

| | Actual numbers | Predicted numbers | Error |
|-----------|----------------|-------------------|-------|
| January | 13,385 | 8360.28 | 0.37 |
| February | 11,140 | 8360.28 | 0.25 |
| March | 7935 | 8360.28 | 0.05 |
| April | 5645 | 8360.28 | 0.48 |
| May | 3775 | 8360.28 | 1.21 |
| June | 4790 | 8360.28 | 0.74 |
| July | 7590 | 8360.28 | 0.10 |
| August | 7275 | 8360.28 | 0.15 |
| September | 9300 | 8360.28 | 0.10 |
| MAPE | | | 38% |

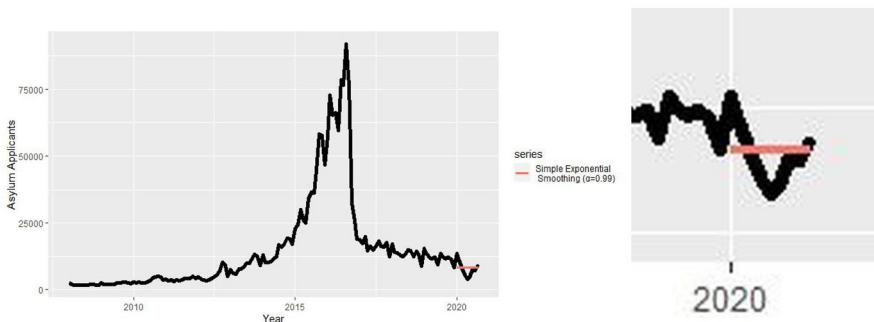


Fig. 2 Time series graph of asylum claims in Germany as well as the predicted numbers using simple exponential smoothing method

case of Germany, the last observation turned out to be a fortunate pick and produced a MAPE below the chosen threshold.

Interestingly, exponential smoothing was identified as applicable method for migration forecasting applications [21]. ARIMA models and exponential smoothing method were identified as univariate time series methods that could be applied on migration forecasting. The exponential smoothing was found preferable due to its simplicity. The arbitrary choice of the smoothing parameter and overreliance on past data were identified as the main disadvantages of the exponential smoothing method.

Exponential smoothing has been widely used for migration forecasting by many statistical agencies of developed countries, mainly due to its simplicity. Major limitations of exponential smoothing include the overreliance on past data and an arbitrary choice of the smoothing parameter. Univariate ARIMA models constitute a standard approach for time series extrapolations. A main advantage of the models is their versatility, the existence of a wide range of applications and software. Additionally, they give an estimate of forecast uncertainty with confidence intervals [18, 21].

In migration forecasting studies, the random walk with drift model was found to be a good choice due to its simplicity as well as model selection criteria [3, 18, 21]. Furthermore, the non-stationarity of the model reflects the volatile feature of forced or irregular migration. The random walk model, being a special case of ARIMA(0, 1, 0), provided better estimates of uncertainty compared to other ARIMA models for asylum-related migration forecasting [18, 21].

The main limitation of using both exponential smoothing and ARIMA models for migration forecasting is their dependence on data only. Data accuracy is a significant challenge for international migration forecasting [3]. Moreover, in case of shock events, ARIMA models produce unreliable predictions.

In this study, the ARIMA(0, 1, 0) model was used. This model corresponds to a random walk model. Various ARIMA models have been explored [20]. ARIMA(0, 1, 0) showed to be the best performing model.

Table 2 shows the actual number of asylum claims per month, the predicted numbers from ARIMA(0, 1, 0), and the error between the two. Figure 3 shows the time series graph representing monthly first asylum claims in Germany and the predicted values using ARIMA(0, 1, 0).

Table 2 Results of ARIMA(0, 1, 0) in Germany

| | Actual numbers | Predicted numbers | Error |
|-----------|----------------|-------------------|------------|
| January | 13,385 | 8360 | 0.37 |
| February | 11,140 | 8360 | 0.25 |
| March | 7935 | 8360 | 0.05 |
| April | 5645 | 8360 | 0.48 |
| May | 3775 | 8360 | 1.21 |
| June | 4790 | 8360 | 0.74 |
| July | 7590 | 8360 | 0.10 |
| August | 7275 | 8360 | 0.14 |
| September | 9300 | 8360 | 0.10 |
| MAPE | 38% | | 38% |

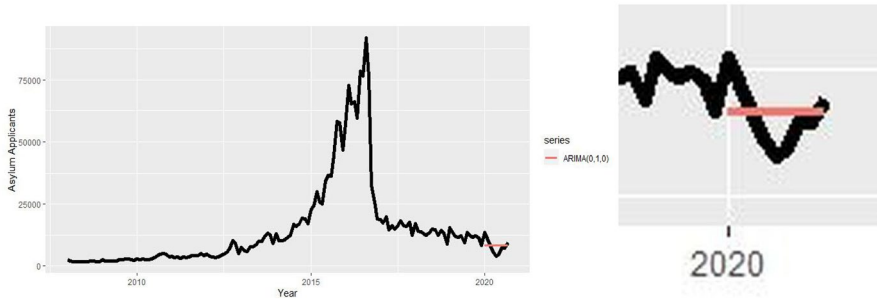


Fig. 3 Time series graph of asylum claims in Germany as well as the predicted numbers using ARIMA(0, 1, 0)

Future research will explore the inclusion of other parameters, besides data, in asylum-related migration forecasting methods, followed by sensitivity analysis, in order to better depict and take into account the complexity and volatility of this type of migration [18, 21]. Interestingly, the exclusive reliance on data has been the main criticism of ARIMA models in migration forecasting, since reliance on data only does not take into account shock events that may possibly occur [18, 21].

5 Conclusions

Forced migration flows constitute the most volatile and sensitive type of migration. Refugees are a vulnerable group globally and asylum applicants need to prove they have suffered or are at risk of suffering persecution. Refugees rely on host governments, international organizations, and non-profit organizations to assist them with access to schooling, medical care, and work among others in the country of destination. When humanitarian aid is not effectively managed, the distress of affected population increases [22, 23].

Existing tools to monitor forced migration flows appear to be weak. Specifically, in this study, the time series models produced results under a specific error threshold; however, the uncertainty in the forecasted values is large. When many coexisting factors interact, they can influence so noticeably the results. Some European countries rely exclusively on statistical modeling. Although models can be efficiently used as decision support systems, they are also widely used as early warning and alert systems. Some of the models are also used to fine tune those systems. This way, the preparedness of policy makers is improved. Relevant systems include the EU Integrated Political Crisis Response (IPCR), the IOM's Displacement Tracking Matrix System (DTM), and the EASO's Early Warning and Preparedness System. In the same vein, some surveys track people's intentions to migrate. The reality is that even with those supportive tools, the forecasts produce big errors because the system's dynamic behavior cannot be tracked.

A large part of the literature contains case studies where univariate time series models were used. Similarly, in this paper, the models relied on data only for asylum-related migration forecasting. However, forced migration can be the result of civil wars, poverty, human right violations, and others. If all these factors are taken into account, a more holistic understanding of the complex and volatile phenomenon will be acquired and lead to treat the causes of it and not its symptoms.

Future research will investigate different time series models including multivariate ones in different European countries. For instance, a comparison between results obtained in this paper and results obtained by a newer time series method using generalized network autoregressive (GNAR) processes [24] could provide further insights about the efficiency of time series modeling in forecasting asylum-related migration. Moreover, other datasets besides the ones obtained from EUROSTAT will be explored.

Funding Open access funding provided by HEAL-Link Greece.

Data Availability The datasets will be made available on reasonable request.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. UNHCR: available at: <https://www.unhcr.org>. (Accessed 24 Sep 2020)
2. Global Trends, Forced Displacement in 2019, UNHCR, Copenhagen, Denmark, 2018
3. Disney G, Wiśniowski A, Forster JJ, Smith PW, Bijak J (2015) Evaluation of existing migration forecasting methods and models. Commissioned research. ESRC Centre for Population Change, University of Southampton, Report for the Migration Advisory Committee
4. EUROSTAT: available at: <https://ec.europa.eu/eurostat>. (Accessed 28 Sep 2020)
5. Lewis WA (1954) Economic development with unlimited supplies of labour. *Manch Sch* 22:139–191
6. Harris JR, Todaro MP (1970) Migration, unemployment and development: a two-sector analysis. *Am Econ Rev* 60:126–142
7. Ahamer G, Korotayev A, Zinkina J (2014) On the structure of the present-day convergence. *Campus-Wide Information Systems* 31:139–152
8. Spinks H (2013) Destination anywhere?: Factors affecting asylum seekers' choice of destination country (pp. 2012–13). Australia: Parliamentary Library
9. Docquier F, Peri G, Ruysen I (2014) The cross-country determinants of potential and actual migration. *Intern Migrat Rev* 48(1_suppl):37–99
10. Castles S, de Haas H, Miller MJ (2014) The age of migration: international population movements in the modern world, 5th edn. The Guilford Press, New York, NY/London

11. Bansak K, Hainmueller J, Hangartner D (2016) How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science* 354:217–222
12. Lee ES (1966) A theory of migration. *Demography* 3:47–57
13. Nagurney A, Daniele P, Nagurney LS (2020) Refugee migration networks and regulations: a multiclass, multipath variational inequality framework. *J Global Optim* 78:627–649
14. Nagurney A, Daniele P, Cappello G (2021) Human migration networks and policy interventions: bringing population distributions in line with system-optimization. *Int Trans Oper Res* 28(1):5–26
15. Keilman N, Pham DQ, Hetland A (2002) Why population forecasts should be probabilistic-illustrated by the case of Norway. *Demogr Res* 6:409–454
16. Brockwell PJ, Davis RA (2016) *Introduction to time series and forecasting*, Third edition, Springer
17. International Organization for Migration (2020) *The future of migration to Europe: a systematic review of the literature on migration scenarios and forecasts*
18. Bijak J, Disney G, Findlay AM, Forster JF, Smith PWF, Wiśniowski A (2019) Assessing time series models for forecasting international migration: lessons from the United Kingdom. *J Forecast* 38(5):470–487
19. European Union Agency for Asylum (2017) *Quantitative assessment of asylum-related migration. A survey of methodology*
20. Mebelli V (2021) *Assessing time series forecast models for mixed migration flows in Europe*. M. Sc. Dissertation Thesis. Intern Helle Univ
21. Bijak J (2012) *Migration assumptions in the UK national population projections: methodology review*. Report for the Office for National Statistics. Southampton: Centre for Population Change
22. Drakaki M, Tzionas P (2021) Investigating the impact of site management on distress in refugee sites using fuzzy cognitive maps. *Int J Disaster Risk Reduct* 60:102282
23. Drakaki M, Gören HG, Tzionas P (2018) An intelligent multi-agent based decision support system for refugee settlement siting. *Int J Disaster Risk Reduct* 31:576–588
24. Urrutia P, Wren D, Vogiatzis C, Yoshida R (2022) SARS-CoV-2 dissemination using a network of the US counties. *Operations Research Forum* 3:29

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.