European Journal of Geography

The publication of the European Journal of Geography (EJG) is based on the European Association of Geographers' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual, or empirical way the quality of research, learning, teaching, and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.

Received: 20/01/2023

Revised: 16/03/2023

Accepted: 17/03/2023

Published: 17/03/2023

DOI: 10.48088/ejg.m.lit.14.1.21.34

Research Article

Comparing student mobility pattern models

Marie-Louise Litmeyer[1,*], Philipp Gareis [2] and Stefan Hennemann[1]

[1] Department of Geography, Justus Liebig University Giessen, Gießen, Germany

[2] Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), Germany

* Correspondence: Marie-Louise.Litmeyer@geogr.uni-giessen.de

Abstract: Classically, gravity models have been used to estimate mobility flows. However, in recent years, a number of new models, such as radiation models, have been introduced to estimate human mobility. The focus has generally been on models dealing with commuting movements. There is no systematic application of different versions of the laws of gravity to student mobility. The application of these models to student mobility provides the opportunity to calculate reliable forecasts of student mobility flows at the micro level, make medium- to long-term decisions at the university level, and implement sustainable strategic orientation. Therefore, this article uses different models to estimate interactions to improve the forecast of the regional distribution of students in Germany under data limitations. Using publicly available data on high school graduates and historical data on student flows between German counties, we show that radiative models with parameters are best suited to predict student flows at the level of German counties. Among parameter-free models, the population-weighted odds model yields the best results.

Keywords: Student Mobility, Student mobility data, Gravity Model, Radial Model, Germany

Highlights:

- Forecasting student mobility flows in Germany
- Model comparison between gravity and radiation models
- Implications for the development of universities

1. Introduction

Since the early 2000s, the number of students in Germany has risen sharply (Multrus et al., 2017). The reasons for this include the expansion of higher education programs, the increasing proportion of high school graduates, the introduction of the bachelor and master system, the abolition of compulsory military service and some double year groups of high school graduates (KMK, 2012; Deutscher Bundstag, 2011). As a result, the higher education system is reaching its capacity limit. Previous studies show a significant underestimation of student numbers at the national level (see, among others, Nutz, 1991; KMK, 2000; Gösta & von Stuckrad, 2007; Wis-senschaftliche Dienste des Deutschen Bundestages, 2006). Errors in forecasting also exist at state level, e.g. for Mecklenburg-Western Pomerania (cf. Kramer & Jost, 2005) and Saxony-Anhalt (Lischka & Kreckel, 2006). This estimation of the number of students is important for higher education institutions, especially from a financial point of view, as basic funding and hence jobs in research and teaching are linked to student numbers (HMWK, 2015). Reliable forecasts of mobility flows at the micro level therefore open up the possibility of making medium to long-term decisions at the university level and implementing a sustainable strategic orientation (e.g. establishment of new professorships or specializations; improving the building infra-structure; adaptation to higher/lower student numbers). Nevertheless, the funding and planning of the German higher education system was based on very erroneous predictions in the past. Therefore, the question arises how can we make better predictions for student flows in Germany with existing data?. For this purpose, this study attempts to improve the forecasting of regional allocation of students in Germany utilizing student flow data and numbers of high-school graduates. Through the usage of publicly available data for high-school graduates and the historical data of student flows between German districts, we show that radiation models provide the best forecast of student flows at the German district level. Although many of the allocation mechanisms for student mobility cannot be explained by our models and data, we outperform existing prediction models and offer a novel approach to forecast student mobility with minimal use of (existing) data.

Our paper is structured as follows: Chapter 2 includes a short literature review and introduces Germany`s higher education sector in order to provide a better understanding of the unique characteristics and specifics of the study area. Chapter 3 then presents the common models for migration movements, followed by the results in Chapter 4. Chapter 5 discusses the results and, based on the findings, provides an outlook on the need for further research.

2. Student Mobility and Germany`s higher education system

Student mobility forecasts play an important role in both national and international mobility flows (Breznik/Skrbinjek 2020). At an international level, universities aim to attract foreign stu-dents to gain access to research funding, enhance research excellence, and diversify their skill portfolio (Choudaha 2017). Regional forecasts of student enrolments were first introduced by Rees (1981) in the UK to identify potential recruitment markets for universities and policy makers. Student mobility becomes particularly relevant when the number of young adults residing in university towns and cities is more than four times that of locations without a higher education institution with all the consequences related to it (e. g. concentration of high skilled labor force; growing disparities between regional labor markets, etc.) (Champion 2022). By accurately identifying regional student numbers, better strategies can be developed to address skill shortages on a regional level and to enhance the marketing and capacity expansion of higher education institutions both within and outside of these regions. As a location for studying, Germany is a study area with a large number of students and internationally competitive programs in the education sector (Weisser, 2019). The universities and universities of applied sciences are spread evenly throughout Germany. Unlike the USA, this means that there are no "educational deserts" in Germany (cf. Hillman, 2016). In addition, German higher education institutions do not charge tuition fees, and hence there are no consumption effects. All western German states (except Bremen, Rhineland-Palatinate and Schleswig-Holstein) that introduced tuition fees between 2006 and 2007 abolished them by 2014 due to changes in government and politics (Kauder & Potraf-ke, 2013).

Despite these advantages in the study area, the analysis must take into account the fact that the German higher education system has changed rapidly over the last 20 years due to the alignment of study structures with bachelor and master degrees. This also has an impact on the mobility of students (Gareis/Brökel 2022). In addition to changes in the types of degrees, there have been other internal and external changes at the universities. The switch in various federal states from the nine-year to the eight-year Abitur (KMK, 2012) and the suspension of compulsory military service in 2011 (Deutscher Bundestag, 2011) led to a significant increase in student numbers.

As figure 1 clearly indicates, in the time span from 1996 to 2016, the number of first-semester students increased from 267,261 to 509,760, which resembles an average annual growth rate of 3.28 percent (Statistisches Bundesamt, 2008 - 2017). This means an average annual growth rate of 3.28 percent. The largest number of first-year students in a single year was 18,748, registered in 2011 (Statistisches Bundesamt, 2008 - 2017). On the one hand, the double cohorts and the abolition of compulsory military service (KMK, 2012) are cited as the reason for this. On the other hand, the proportion of high school graduates starting university education studies as well as the number of school graduates having acquired a higher education entrance qualification has grown continuously. The average annual growth rate of high school graduates is 1.62 percent. In total, in the time span from 1996 to 2016, the number of high school graduates increased from 215,245 to 296,736, whereby Germany recorded the highest number in 2013, where 318,900 students received a high school degree.

Overall, the number of high school graduates increased from in 1996 to in 2016, and Germany recorded the highest number of in 2013 (INKAR, 2020).

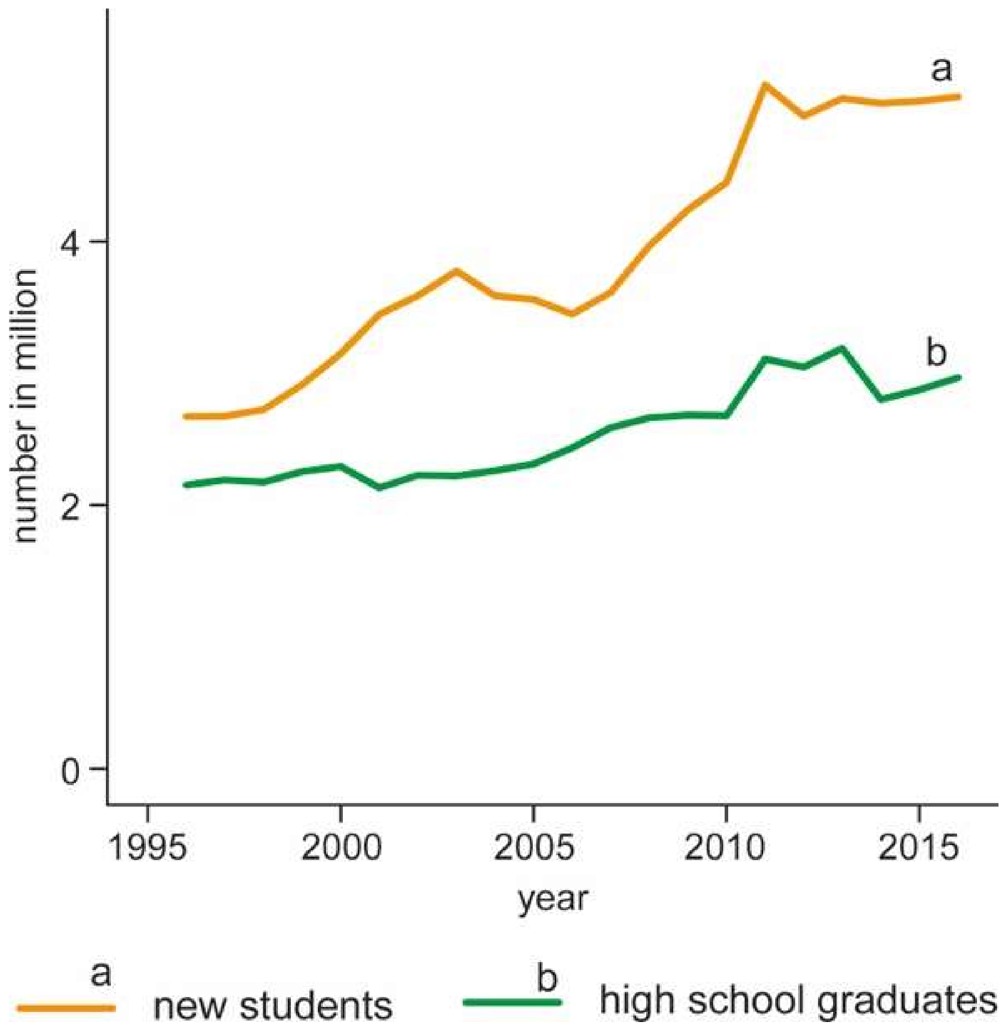

Figure 1. Comparison of high school graduates and new students. Description: The development of the number of school leavers and students shows clear increases over time. Source: Statistisches Bundesamt, 2008 – 2017

https://doi.org/10.48088/ejg.m.lit.14.1.21.34

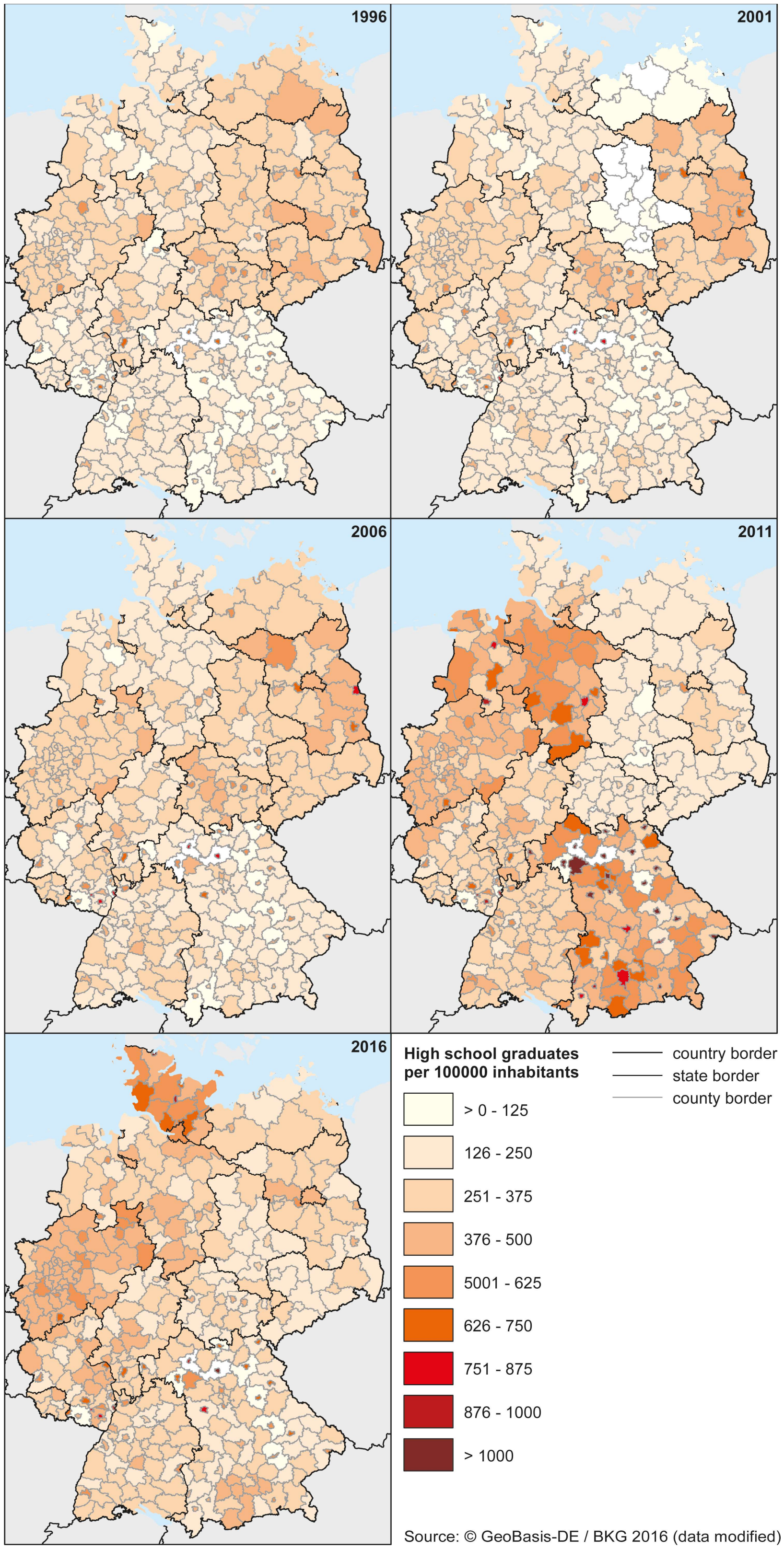

Figure 2. Development of high school graduates per 100,000 inhabitants. Description: The high school graduates per 100,000 inhabitants for the years are presented. It is important for the interpretation that in 2011 in Bavaria and Lower Saxony and in 2016 in Schleswig-Holstein two cohorts took their A-Level exams.

The development of high school graduates per 100,000 inhabitants shows that the numbers in the new federal states were initially somewhat higher than in Germany as a whole (Abb. 2). The numbers for Saxony-Anhalt were particularly low in 2001, and in the following years, especially 2011 (Lower Saxony, Bavaria) and 2016 (Schleswig-Holstein), the effects of the double cohorts are visible.

The changes and high growth rates in the number of students and high school graduates with higher education entrance qualifications are evidence of incalculable factors that could not be included in the forecasts for previous years or do not provide sufficient explanatory content for the allocation of students. For this reason, the preparation of more accurate forecasts at county and independent city level is the aim of this paper.

3. Data and Methodology

The 401 NUTS-3 regions and independent cities in Germany serve as the study area for the forecasting instruments presented. The data is available for the period between 1996 and 2016, with the number of high school graduates coming from the database "Indicators and Maps of Spatial and Urban Development" of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (INKAR, 2020). The query of the number of students, the number of first-year students and the respective internal migration flows of future students from their home district to the university location is carried out in the research database of the federal statistical offices of Germany (FDZ, 1992 - 2017). For data protection reasons, only migration flows with more than three students are considered. In addition, the Euclidean distances between all district pairs are calculated using the Haversine formula based on the geometric centers. The data source is adjusted for district reforms and is based on the starting point of 2017.

For over 100 years, gravitational models, based on Newton's law, have predicted mobility flows. The interaction between two locations is proportional to their size (e.g. number of inhabitants, number of students, etc.) and the attraction decreases with increasing distance (c. s. Carey, 1858; Zipf, 1946; Wilson, 1970; Erlander & Stewart, 1990; Maris et al. 2019). The application of universal gravitation occurs in many areas, such as transportation (Jung et al., 2008; Kaluza et al., 2010; Odlyzko, 2015; Hong & Jung, 2016; Bartzokas-Tsiompras & Photis, 2019; Azad et al., 2021), population flows and commuting (Griffith, 2009; Murat, 2010; Lenormand et al., 2012; Liang et al., 2013; Masucci et al., 2013; Thomas & Tutert, 2013; Liu et al., 2014; Lenormand et al., 2016; Kluge & Schewe, 2021), freight transport (Kaluza et al., 2010), international trade (Bergstrand, 1985; Fagiolo, 2010), telecommunications (Krings et al., 2009), scientific collaboration (Pan et al., 2012) and the consideration of human mobility in the context of infectious disease spreads (Viboud et al., 2006; Balcan et al., 2009; Tizzoni et al., 2014; Sallah, 2017 Marshall et al., 2018; Tuite et al., 2018). Estimates of student migration flows also use gravity models (Sá et al., 2004; Alm & Winters, 2009; Cooke & Boyle, 2011; Faggian & Franklin, 2014).

The gravity models used in studies so far have some weaknesses. In practice, the use of different functions and up to nine parameters are necessary to match the empirical data. (Viboud et al., 2006; Balcan et al., 2009; Kaluza et al., 2010; Krings et al., 2009; Simini et al., 2012). In addition, empirical data must be available for comparison or estimation. Only in regions where systematic comparative data already exists are predictions possible. Moreover, the gravity model exhibits a systematic prediction discrepancy due to the lack of consideration of locations between the target and source areas. This means that similarly-sized fluxes are predicted for interactions between two sites with similar origin and destination measurements and comparable distances, without considering the space in between. The forecasted migration movement rises indefinitely when the location dimensions of the target region are increased. Moreover, the classical gravity model cannot account for variability in interactions between two sites (Simini et al., 2012). For this reason, Simini et al. (2012) develop the classical radiation model as an alternative approach to estimate mobility flows. The advantages of this approach are sparse use of data and parameter freedom. We introduce different models to highlight the negative and positive aspects as well as the applicability to our topic.

Despite a multitude of other motives (tuition fees, cost of living, the university's prestige, influence of friends and relatives) influencing the choice of study location (Perna, 2006), distance, transition rates and previous student mobility patterns are the only variables used for our modeling. This guarantees a minimum use of data for reproducibility and leads to the following definition: m_i denotes the number of high school graduates of a year t at the home location i. For the future university location j, the indicator n_j is chosen at time t-1. It represents the number of students in the first semester. In addition, s_ij describes the number of students in the first semester at time t-1 between the home district and the future university location. The calculation includes all other potential university locations with this variable. For the further calculation of the average student migration of first-semester students T_ij from location i to j, the average migration rate of first-semester students at time t-1 is calculated for the entire federal territory. In addition, the sum of students T_i who leave their home country i to study is proportional to m_i (ϑ/M). ϑ stands for the total number of all mobile students (without internal commuters), r for the distance between home and university region and M for the number of all first-year students in Germany.

According to Simini et al. (2012), we define the classical gravity model as follows, where α,β,γ are freely selectable parameters:

$$T_{ij}^{grav} = \frac{m_i^{\alpha} n_j^{\beta}}{r^{\gamma}} \quad (1)$$

and the definition of the radial model is:

$$T_{ij}^{radial} = T_i \frac{m_i n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})} = \frac{\vartheta}{M} \frac{m_i^2 n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})} \quad (2)$$

In previous studies for commuters, the main advantage of the classic radial model is its distance independence (Masucci et al., 2013). Furthermore, the radial model is free of parameters and hence no calibration is required (Lenormand et al., 2016). Due to the polycentricity of the higher education system and the 1970s regionalization through the expansion of the number of university locations (Framhein, 1983), it can be assumed that the prerequisite (cf. Masucci et al., 2013) of the uniform distribution of the object of study is fulfilled regarding student migration behavior in Germany. The original version of the model by Masucci et al. (2013) is intended for an infinite system by deriving it with the help of the thermodynamic limiting case (thermodynamic limiting assumption). For a finite system, a normalization factor is added to the model. The data flows are systematically underestimated by the factor 1/(1-m_i/M). Consequently, we use a slightly modified variant of the radial model for finite systems (Masucci et al., 2013):

$$T_{ij}^{finit} = T_i \frac{m_i n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})} = \frac{1}{1-\frac{m_i}{M}} \frac{\vartheta}{M} \frac{m_i^2 n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})} \qquad (3)$$

To improve the applicability of the radiation model to different mobility systems and different spatial scales, Kang et al. (2015) extend the probability of the original radiation model (2) by a scaling exponent λ. In addition, we present four variants of the model based on combinations of search and trigger directions. This enables an integration of the different pull and push factors into the model. It also follows the introduction of the normalization factor for finite systems based on Masucci et al. (2013). Overall, the authors present four variants of the generalized radiation model: production-constrained intervention-based radiation model (PIK), production-constrained competition-based radiation model (PCR model), attraction-constrained intervention-based radiation model (AIR) and attraction-constrained competition-based radiation model (ACR):

$$T_{ij}^{PIK} = \frac{T_i}{\sum_{k\neq i}^{N}\left[\frac{m_i n_k}{(m_i+s_{ik})(m_i+n_k+s_{ik})}\right]^{\lambda}} \left[\frac{m_i n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})}\right]^{\lambda} \qquad (4)$$

$$T_{ij}^{PCR} = \frac{T_i}{\sum_{k\neq i}^{N}\left[\frac{n_k m_i}{(n_k+s_{ki})(n_k+m_i+s_{ki})}\right]^{\lambda}} \left[\frac{n_j m_i}{(n_j+s_{ji})(n_j+m_i+s_{ji})}\right]^{\lambda} \qquad (5)$$

$$T_{ij}^{AIR} = \frac{T_j}{\sum_{k\neq j}^{N}\left[\frac{m_k n_j}{(m_k+s_{kj})(m_k+n_j+s_{kj})}\right]^{\lambda}} \left[\frac{m_i n_j}{(m_i+s_{ij})(m_i+n_j+s_{ij})}\right]^{\lambda} \qquad (6)$$

$$T_{ij}^{ACR} = \frac{T_j}{\sum_{k\neq j}^{N}\left[\frac{n_j m_k}{(n_j+s_{jk})(n_j+m_k+s_{jk})}\right]^{\lambda}} \left[\frac{n_j m_i}{(n_j+s_{ji})(n_j+m_i+s_{ji})}\right]^{\lambda} \qquad (7)$$

Another typical parameter-free mobility model is the population-weighted odds (PWO) model. It aims to compensate for underestimations on a city scale. The model's special feature is that journeys within a city as well as intercity traffic can be precisely predicted on different spatial scales (Yan et al., 2014). S_ij is therefore the extension of s_ij. S_ij additionally considers the number of students at home and future university location. There is also a correction with 1/G. G defines the total number of all students in Germany. The basic assumption for this model is that the probability of moving to a university location is proportional to its attractiveness. Therefore, the attractiveness of a region with a university is inversely proportional to the number of students. The overall formula is:

$$T_{ij}^{PWO} = T_i \frac{n_j\left(\frac{1}{S_{ji}} - \frac{1}{G}\right)}{\sum_{k\neq i}^{N} n_k\left(\frac{1}{S_{ki}} - \frac{1}{G}\right)} \qquad (8)$$

Alternatively, we use the Opportunity Priority Selection (OPS) model. The main idea is that the transition probability to a destination is proportional to the probability that the destination has higher utility than regions with a smaller distance from the initial location (Liu & Yan, 2019). The definition is:

$$T_{ij}^{OPS} = T_i \frac{\frac{n_j}{S_{ij}}}{\sum_{k\neq i}^{N}\frac{n_k}{S_{ik}}} \qquad (9)$$

The Soerensen index (SI) makes it possible to determine the best possible model. It is a popular tool for measuring fluctuation. The similarity measurement indicates the correctly reproduced proportion of commuting flows in simulated networks and varies between 0 and 1. A value of 0 means no correspondence with the original mobility movements, whereas a value of 1 in the empirical and simulated networks is identical. Compared to other similarity measurements such as Euclidean distance, the Soerensen index maintains sensitivity in more heterogeneous data sets and is less sensitive to outliers (McCune & Grace, 2002). The following formula calculates the measurement indicator where 〖T_ij〗^empric represents the empirical data and T_ij^model stands for the different models 1-9 (Soerensen, 1948):

$$SI = \frac{2\sum_{i=1}^{N}\sum_{j=1}^{N} min\left(T_{ij}{}^{empric}, T_{ij}^{model}\right)}{\sum_{i=1}^{N} T_{ij}{}^{empric} + \sum_{j=1}^{N} T_{ij}^{model}} \tag{10}$$

4. Results

When comparing the different models, Soerensen's fluctuation index (Soerensen index) shows that there is a better match between empirical and simulated student mobility for the younger age groups for the period of investigation (see Figure 3). The model achieves the best fit in 2012. Aside from the Soerensen index, the correlations calculation determines the relationship between the models and the empirical data with another indicator. It is also noticeable that the correlation coefficients at the beginning of the investigation period are significantly lower than the calculations of the measurements of association between 2009 and 2016.

When comparing the models with each other, it becomes apparent that the gravity model has an average Soerensen index of 0.5894 (Figure 3). The correlation coefficient varies between 0.64 and 0.72 (Figure 4). The direct comparison of the models with the empirical data makes it clear that the gravity model has a marginal deviation only for commuter flows with approx. 2,000 persons (Figure 5). When analyzing the site size, the gravity model provides overestimates in the range between 1,000 and 2,000 students (Figure 6) and in the case of very large sites. In addition, Figure 6 shows that the gravity model underestimates small distances up to 30 kilometers.

A comparison with the probabilities of the number of new students from the previous year (see Figure 5) shows that the prognosis overestimates for study locations with low enrollment numbers. For locations with 1,000 or 2,000 students in particular (Figure 6), the model's probability is significantly higher than the results from the empirical data. In addition, this model underestimates the probabilities for sites with very large enrollments. When looking at the probabilities of attending a university location as a function of distance (cf. Figure 7), we see a clear underestimation of the mobility flow in the direct environment up to 10 kilometers.

With the gravity model, in contrast, the estimated probability between approx. 30 and 130 kilometers (Figure 7) is clearly below the results from the empirical data, while from 170 kilometers (Figure 7) upwards, the model again determines higher probabilities than can be seen in the empirical results.

The classical radiation model shows the lowest match to the real distribution with an average of 0.5221. Pearson's correlation measurement yielded values between 0.62 and 0.73 (Figure 4) for the radial model, while the values for the finite radial model vary between 0.63 and 0.74 (Figure 4). In general, both models overestimate small student migration flows. Particularly good fits are visible for mobility flows between 100 and 1,000 students (Figure 6). Concerning location size, both the radial model and the finite model produce overestimates in the range between 1,000 and 2,000 students (Figure 6), while large locations remain systematically underestimated. Over-all, the probabilities for the radial and finite radial models of students attending a university lo-cation depending on the distance show only marginal differences. The model probabilities correspond to the empirical data of both models up to a radius of approx. 60 kilometers (Figure 7). From about 70 kilometers (Figure 7) upwards, the probabilities of the models is slightly overestimated.

Generalized models achieve significantly higher agreement between model and empirical data. The ACR and AIR models provide the best results among the four model variants between 1996 and 2016, with an average Sörensen index (Figure 3) of 0.7088 and 0.7006 (standard deviation: 0.0041 and 0.0056) respectively. The average for the PIK and PCR model is 0.5894 and 0.6471 (Figure 3) respectively (standard deviation: 0.0108 and 0.0108). In addition, a high correlation between 0.75 and 0.82 (Figure 4) exists for the ACR and AIR models. The correlation coefficient for the PIK and PCR models is between 0.72 and 0.78 and between 0.69 and 0.76 (Figure 4) respectively. Examining the student flows, it is noticeable that low flow numbers remain underestimated in all four models (Figure 5). The adjustment is optimal at around 1000 students per mobility flow (Figure 6). For the probability of attending a university location, very small deviations can be observed in the number of students, especially for the AIR and ACR models. In the PIK and PCR models, on the other hand, the probability overestimates for small locations and underestimates for large locations. Concerning distance, particularly precise estimates can be reported for the PIK model. The model slightly underestimates the probability of choosing a college location only in a radius of 30 to 50 kilometers (Figure 7). The ACR model provides similar results. The AIR model slightly underestimates the determined probability within the first ten kilometers. The model continues to show a slight overestimation between 30 and 70 kilometers (Figure 7). The PCR model underestimates the probability of a university located within a radius of 30 kilometers (Figure 7). From about 50 kilometers (Figure 7) upward, marginal differences exist between the model and the real data.

Generalized models involve an examination of parameters over time (Figure 8). The parameters are selected using a search algorithm. This maximizes the Soerensen index. The parameter for the ACR model is stable at 1 from 1996 to 2016. For the PIK and AIR models, the model is weighted with a factor of 1 until 2006 and 2008 respectively. A weighting of 0.9 followed. The maximizing parameter is 1 until 2001. This is followed by three years with a coefficient of 0.9 and another five years with a parameter of 0.8. From 2011 onward, 0.7 is chosen as the optimal parameter.

For the parameter-free PWO model, the Soerensen index varies between 0.6741 and 0.6951 (Figure 3). On average, the fit is 0.6849 (standard deviation: 0.0061). The correlation coefficient ranges from 0.76 to 0.8 (Figure 4). Furthermore, Figure 6 illustrates the underestimation of low student mobility flows. The deviations are also smaller concerning the size of the university lo-cation than in the non-parameterized models. In addition, Figure 7 shows that the PWO model underestimates small distances up to 30 kilometers.

The OPS model obtains a similarly good agreement. In this case, the Soerensen index averages 0.6360 (standard deviation: 0.0086). Overall, it varies between 0.6213 and 0.6478 for this model (Figure 3). In addition, Figure 6 illustrates that for the OPS model, underestimation of the models occurs with low enrollment numbers of students, while mobility flows with large enrollment numbers are systematically overestimated. Similar to the nonparametric models, this model underestimates the probability of attending a college site as a function of its size. However, the variances are smaller. Figure 7 shows that the OPS model underestimates the student enrollment numbers for small distances up to about 30 kilometers. In the OPS model, the underestimation is stronger than in the PWO model. In the range between 90 and 400 kilometers, the probabilities are slightly overestimated.

https://doi.org/10.48088/ejg.m.lit.14.1.21.34

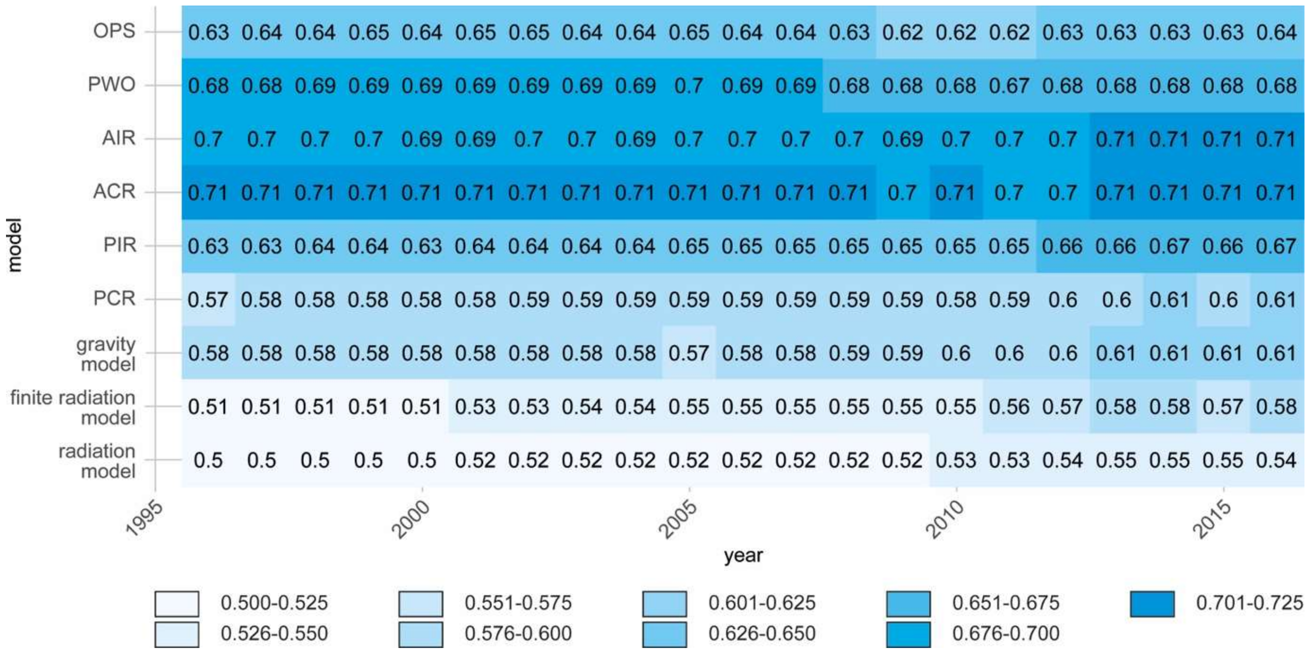

Figure 3. The Soerensen index over time. Description: The Soerensen index has grown slightly over time for all models. This means, the fits between the predicted fluxes and the empirical data become better. Overall, the AIR and ACR models achieve the best results. Source: own calculation, FDZ 1992-2017

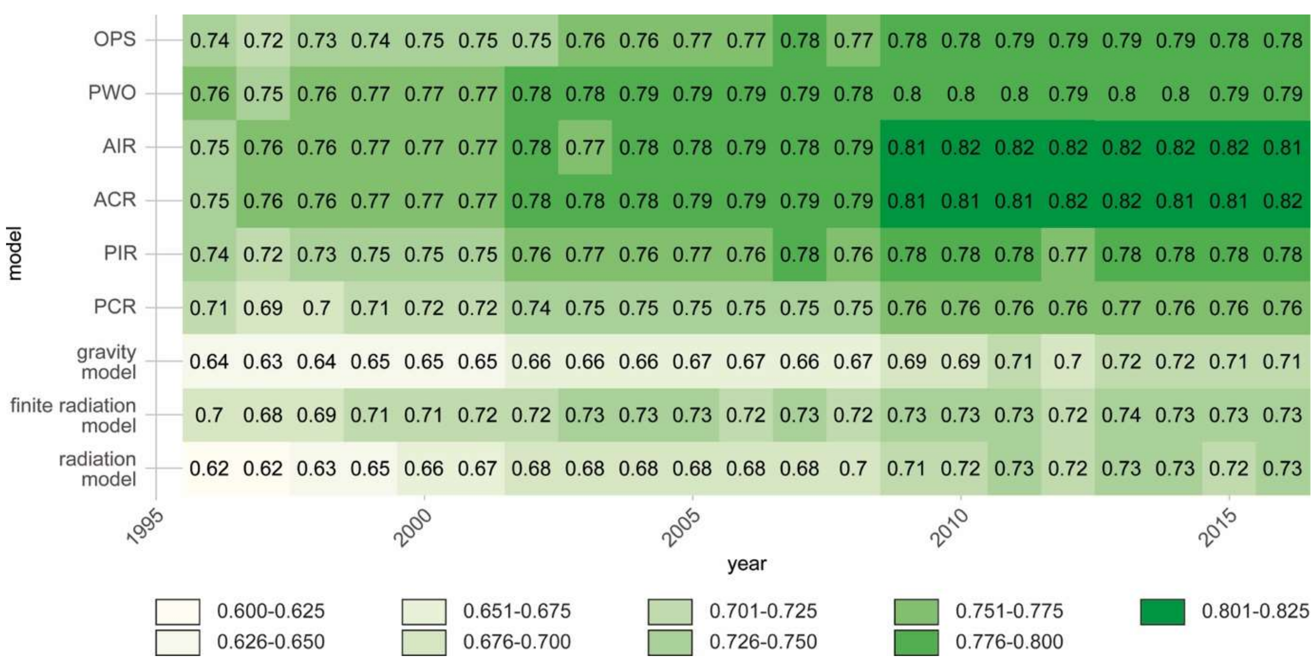

Figure 4. The Pearson correlation coefficient over time. Description: The Pearson correlation coefficient also increased slightly over time for all models. The deviations between the predicted flows and the empirical data became smaller. Overall, the AIR and ACR models show the strongest correlations between model and empirical data. Source: own calculation, FDZ 1992-2017

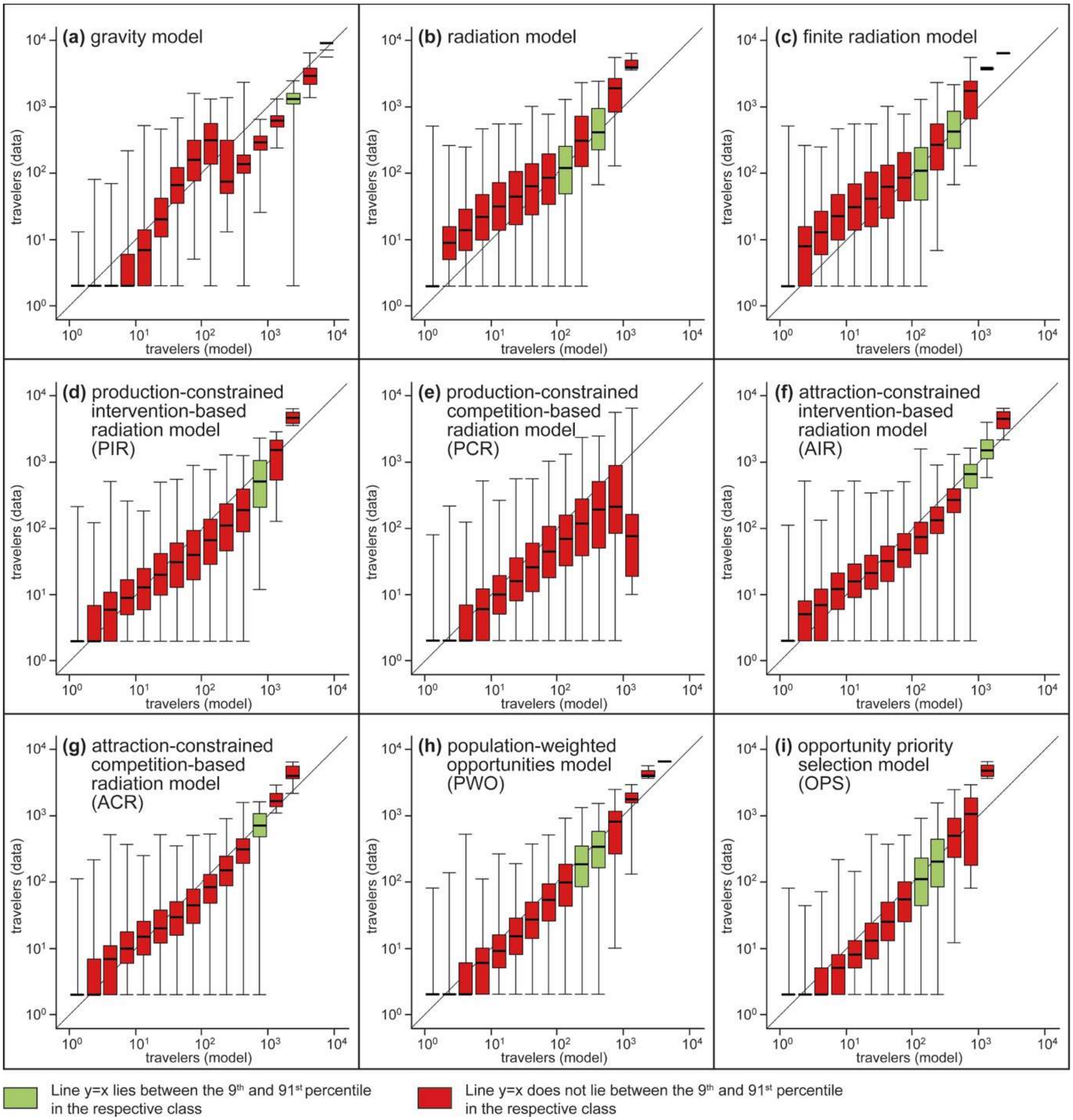

Figure 5: Deviation between the empirical data and the models. Description: The boxplots are coloured green if the line y=x lies between the 9th and 91st percentile in the respective class. Otherwise, the colour is red. Especially for mobility flows with more than 100 people, very good adaptations can be observed. Source: own calculation; FDZ, 1992-2017

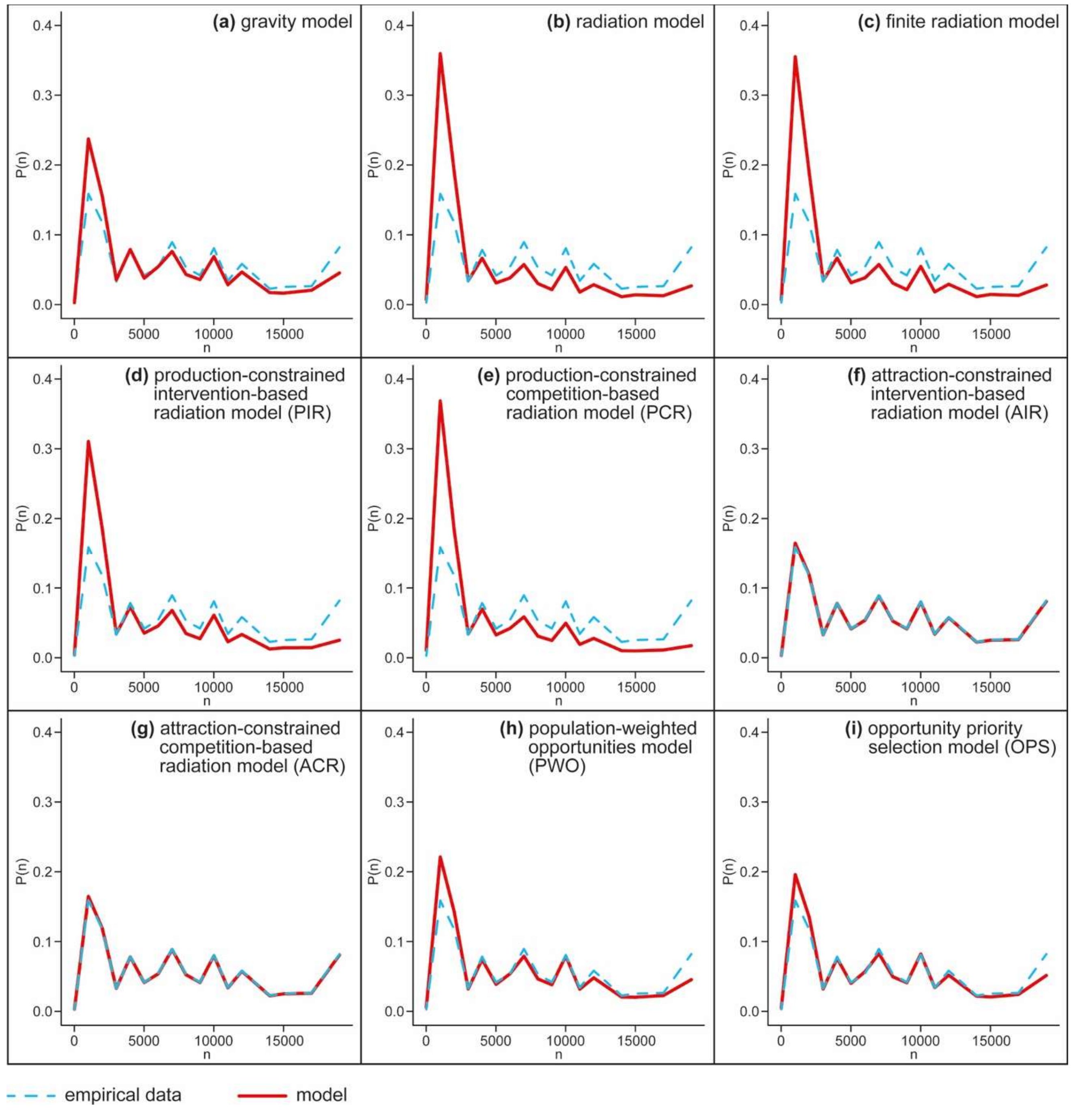

Figure 6: Probability in relation to the size of the higher education location. Description: The figure shows the probability of student mobility between two counties depending on the size of the chosen university location. The ACR and AIR models illustrate particularly small deviations. Source: own calculation, FDZ, 1992-2017

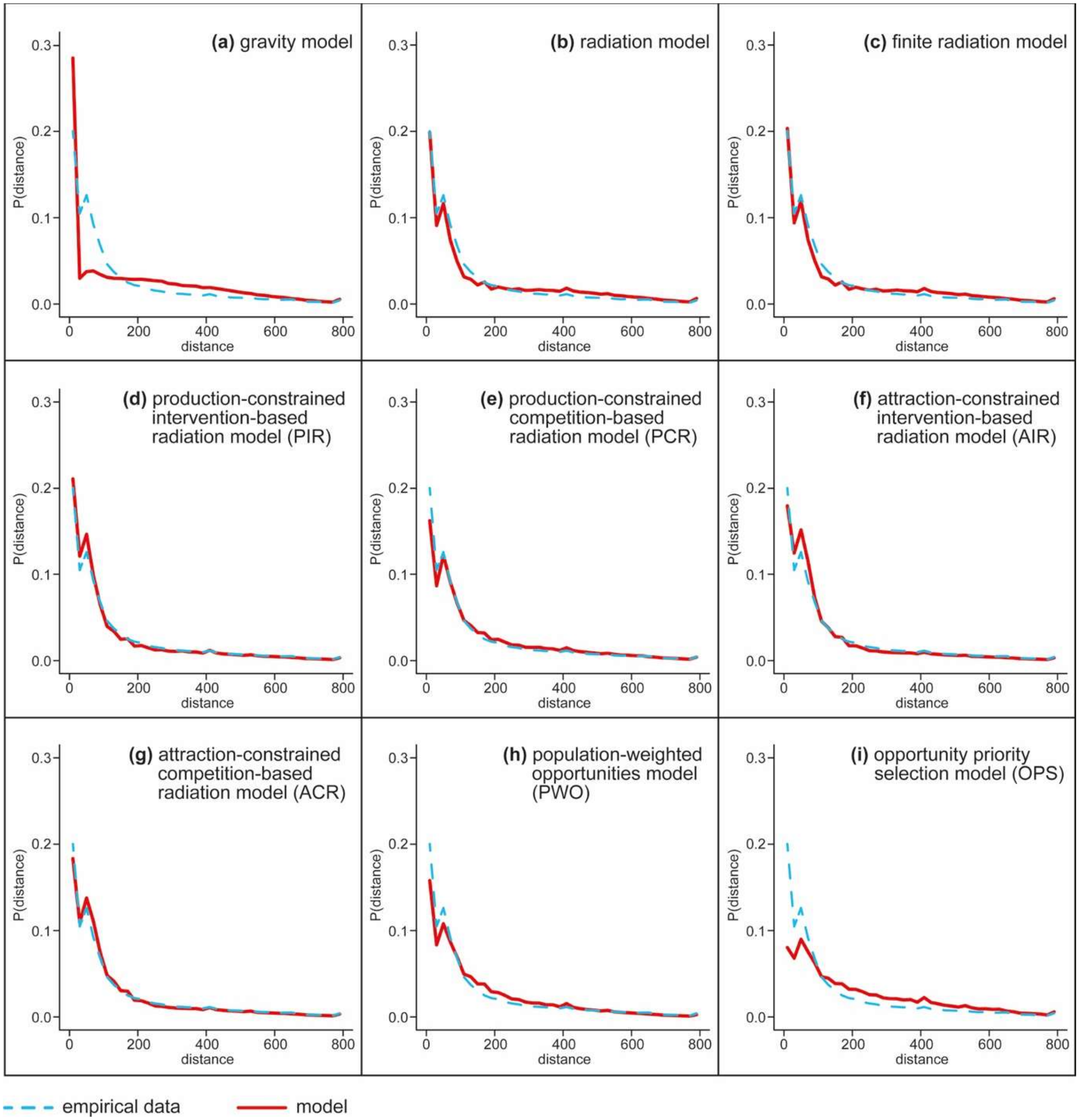

Figure 7. Probability in relation to the distance. Description: The figure illustrates the probability of student mobility between two counties as a function of distance. The ACR model shows particularly small deviations. Source: own calculation, FDZ 1992-2017

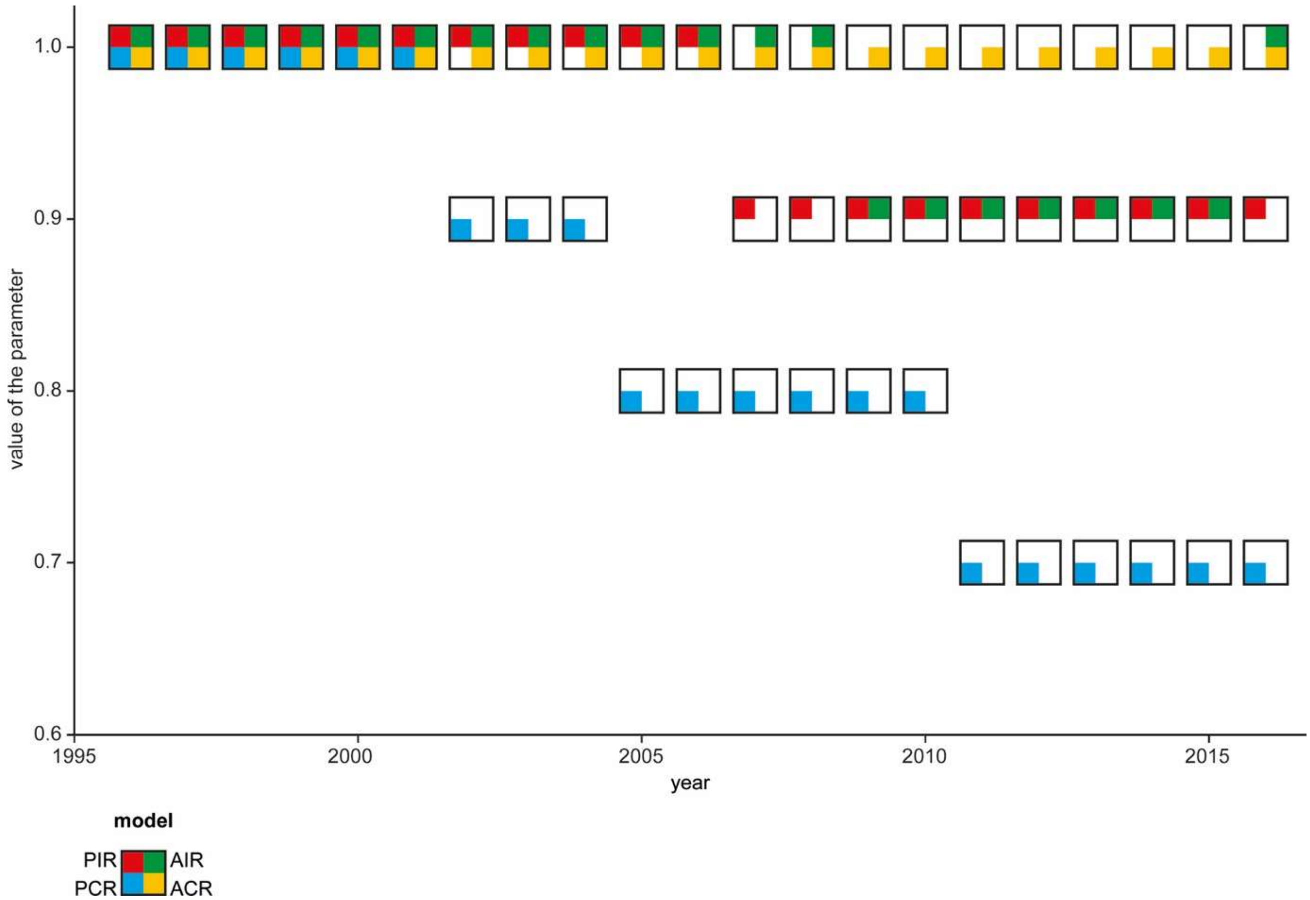

Figure 8. Parameter development over time. Description: The figure visualises the relatively stable development of the respective parameter over time. If the parameter takes on the respective value, the corner of the model is coloured. Only for the PCR model does the parameter fluctuate between 1 and 0.7. Source: own calculation, FDZ 1992-2017

5. Discussion

Overall, the analysis of the correlation coefficient and the Soerensen index shows particularly good results for the AIR and ACR models. These observations are also stable over time. The reason for this could be the historically developed nationwide coverage of higher education institutions in Germany. This enables a large number of students to begin their studies at locations close to home.

Among the non-parameterized models, the OPS model delivers the best results. Models that focus on the size of the university location produce comparatively good results. This finding reflects the results of various studies. Potential students often migrate to student-dominated regions or regions with a high proportion of academics due to similar lifestyles and amenities (Buenstorf et al., 2016; Haussen & Uebelmesser, 2018). However, the non-parameterized models underestimate student flows, especially for large university locations. The underestimations could be due, among other things, to regions with a high degree of urbanization such as Berlin, Hamburg or Munich. They are considered very attractive and appealing to first-year students due to di-verse amenities (Cullinan & Duggan, 2016; Sá et al., 2004; Weisser, 2019).

Overall, it should be noted for all the models presented that, compared to the forecasts of the Standing Conference of the Ministers of Education and Cultural Affairs of the States in the Federal Republic of Germany, significantly less data is required for the calculation. Only the first semester numbers of the previous year, the high school graduates and the distance between the university and the home district are included in the calculation. This means that the models presented are significantly more data-efficient than the models of the Standing Conference of the Ministers of Education and Cultural Affairs of the States in the Federal Republic of Germany (KMK, 2012).

Another advantage of the models presented is the possibility of carrying out the predictions on the district. This gives universities the chance to obtain a reliable forecast for the coming semester and is of particular importance for the basic financial security of universities (HMWK, 2015). In addition, the predicted student traffic flows can be used to negotiate medium to long-term public transport connections with transport providers. In this context, further studies on mobility flows at the community level would be conceivable.

Despite these advantages, the models presented have weaknesses. The main problem in this context is that normalization (cf. Kang et al., 2015; Masucci et al., 2013; Yang et al., 2014; Liu & Yan, 2019) only allows an estimate of students who have already started their studies in the previous year.

When looking at the results, it should also be critically noted that the instruments presented can-not be used to react to political decisions. For example, decisions such as the suspension of compulsory military service in 2011 cannot be integrated into the forecasts. Therefore, forecasts are always subject to an uncalculated error. Against the backdrop of the current Covid-19 pan-demic and the resulting digital teaching offer of the universities, the current student numbers should be used to examine the extent to which the spatial mobility patterns of students will still be valid in the future.

In conclusion, the generalised radiation models provide the best fits to the empirical data and should be used to predict student mobility.

Acknowledgement

We would like to thank Lisett Diehl for her cartographic support.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, 28(6), 728–738. https://doi.org/10.1016/j.econedurev.2009.06.008

Azad, M., Abdelqader, D., Taboada, L. M., & Christopher R. Cherry, C (2021): Walk-to-transit demand estimation methods applied at the parcel level to improve pedestrian infrastructure investment. *Journal of Transport Geography* 92. https://doi.org/10.1016/j.jtrangeo.2021.103019

Balcan, D., Colizza, V., Goncalves, B., Hud, H., Ramasco, J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences,* 106(51), 21484-21489. https://doi.org/10.1073/pnas.0906910106

Bartzokas-Tsiompras, A. & Photis, Y. (2019): Measuring rapid transit accessibility and equity in migrant communities across 17 European cities. *International Journal of Transport Development and Integration* 3 (3), 245–258. https://doi.org/10.2495/TDI-V3-N3-245-258.

Bergstrand, J. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics* (67), 474–481. https://doi.org/10.2307/1925976

Buenstorf, G., Geissler, M., & Krabel, S. (2016). Locations of labor market entry by German university graduates: is (regional) beauty in the eye of the beholder?. *Review of Regional Research*, 36(1), 29-49. https://doi.org/10.1007/s10037-015-0102-z

Breznik, K., & Skrbinjek, V. (2020). Erasmus student mobility flows. European Journal of Education, 55(1), 105-117.Carey, H.C. (1858). *Principles of Social Science*. Lippincott.

Champion, T. (2022). The United Kingdom's 'going away to university'migration and its changing impact on local populations in the twenty-first century. *Local Economy*, 37(4), 279-296.

Choudaha, R. (2017). Three waves of international student mobility (1999–2020). *Studies in Higher Education*, 42(5), 825-832.

Cooke, T. J., & Boyle, P. (2011). The migration of high school graduates to college. *Educational Evaluation and Policy Analysis*, 33(2), 202-213. https://doi.org/10.3102/0162373711399092

Cullinan, J., & Duggan, J. (2016). A School-Level Gravity Model of Student Migration Flows to Higher Education Institutions. *Spatial Economic Analysis*, 11(3), 294-314. https://doi.org/10.1080/17421772.2016.1177195

Deutscher Bundestag (2011). *Aussetzung der allgemeinen Wehrpflicht beschlossen*. Retrieved June 10, 2019, from https://www.bundestag.de/dokumente/textarchiv/2011/33831649_kw12_de_wehrdienst-204958

Erlander, S., & Stewart, NF. (1990). *The gravity model in transportation analysis: theory and extensions. Topics in transportation.* VSP.

Faggian, A., & Franklin, R. S. (2014). Human capital redistribution in the USA: the migration of the college-bound. *Spatial Economic Analysis*, 9 (4), 376–395. https://doi.org/10.1080/17421772.2014.961536

Fagiolo, G. (2010). The international-trade network: gravity equations and topological properties. *Journal of Economic Interaction and Coordination* 5, 1–25. https://doi.org/10.1007/s11403-010-0061-y

FDZ (1992-2017). Froschungsdatenzentrum: Statistik der Studenten. Retrieved June 6, 2019, https://www.forschungsdaten-zentrum.de/de/bildung/studenten.

Framhein, G. (1983). Alte und neue Universitäten. Einzugsbereiche und Ortswahl der Studenten, Motive und Verhalten. In Der Bundesminister für Bildung und Wissenschaft (Eds.), *Schriftenreihe Hochschule,* (44). Bock.

Gareis, P. / Broekel, T. (2022): The Spatial Patterns of Student Mobility Before, During and After the Bologna Process in Germany. *Tijdschrift voor economische en sociale geography*. https://doi.org/10.1111/tesg.12507.

Gösta, G., & von Stuckrad, T. (2007). *Die Zukunft vor den Toren. Aktualisierte Berechnungen zur Entwicklung der Studienanfängerzahlen bis 2020.* Retrieved November 13, 2020, from https://www.che.de/download/che_prognose_studienanfaengerzahlen_ap100-pdf/?wpdmdl=11190&ind=5d1a0a27b4528

Griffith, D.A. (2009). Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems,* 11, 117–140. https://doi.org/10.1007/s10109-009-0082-z

Haussen, T., & Uebelmesser, S. (2018). No place like home? Graduate migration in Germany. *Growth and Change*, 49(3), 442-472. https://doi.org/10.1111/grow.12249

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal, 53*(4), 987–1021. https://doi.org/10.3102/0002831216653204

HMWK (2015). *Hochschulpakt 2016 – 2020.* Retrieved July 12, 2019, from https://wissenschaft.hessen.de/si-tes/default/files/media/hmwk/hsp_2016-2020.pdf

Hong, I., & Jung, W.S. (2016). Application of gravity model on the Korean urban bus network. *Physica A: Statistical Mechanics and its Applications*, 46, 48–55. https://doi.org/10.1016/j.physa.2016.06.055

INKAR (2020): Indikatoren und Karten zur Raum- und Stadtentwicklung. Retrieved May 11, 2020, https://www.inkar.de

Jung, W.S., Wang, F., & Stanley, H. (2008). Gravity model in the Korean highway. *Europhysics Letters Association*, 81(4). https://doi.org/10.1209/0295-5075/81/48005

Kaluza, P., Kölzsch, A., Gastner, M.T., & Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7, 1097–1103. https://doi.org/10.1098/rsif.2009.0495

Kang, C., Liu, Y., Guo, D., & Qin, K. (2015). A Generalized Radiation Model for Human Mobility: Spatial Scale, Searching Direction and Trip Constraint. *PLoS ONE ,10*, Article e0143500. https://doi.org/10.1371/journal.pone.0143500.g003

Kauder, B., & Potrafke, N. (2013). *Government Ideology and Tuition Fee Policy: Evidence from the German States.* CESifo Working Paper, No. 4205, Center for Economic Studies and ifo Institute (CESifo), Munich.

Kluge, L. / Schewe, J. (2021): Evaluation and extension of the radiation model for internal migration. *Phys. Rev. E* 104, 054311. https://doi.org/10.1103/PhysRevE.104.054311

KMK (2005). *Prognose der Studienanfänger, Studierenden und Hochschulabsolventen bis 2020*. Retrieved October 23, 2020, from https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_10_01-Studienanfaenger-Absolventen-2020.pdf

KMK (2011). *Bestandsaufnahme und Perspektiven der Umsetzung des Bologna-Prozesses.* Retrieved May 6, 2019, from https://www.kmk.org/fileadmin/veroeffentlichungen_beschlu-esse/2011/2011_03_10-Bestandsaufnahme-Bologna-Prozess.pdf

KMK (2012). *Vorausberechnung der Studienanfängerzahlen 2012-2025. Fortschreibung. Stand, 24, 2012.* Retrieved April 1, 2019, from https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Vorausberechnung_der_Studienanfaengerzahlen_2012-2025_01.pdf

Kramer, J.W. (2005). Zur Prognose der Studierendenzahlen in Mecklenburg-Vorpommern bis 2020, Wismarer Diskussionspapiere, No. 12/2005, HochschuleWismar, Fachbereich Wirtschaft, Wismar.

Krings, G., Calabrese, F., Ratti, C., & Blondel, V.D. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics Theory and Experiment*, 2009(7). https://doi.org/10.1088/1742-5468/2009/07/L07003

Lenormand, M., Huet, S., Gargiulo, F., & Deuant, G. (2012). A universal model of commuting networks. *PLoS One,* 7, Article, e45985. https://doi.org/10.1371/journal.pone.0045985

Lenormand, M., Bassolas, A., & Ramasco, J. J. (2016). Systematic comparison of trip distributionlaws and models. *Journal of Transport Geography*, 51, 158-169. https://doi.org/10.1016/j.jtrangeo.2015.12.008

Liang, X., Zhao, J., Dong, L., & Xu, K. (2013). Unraveling the origin of exponential law in intraurban human mobility. *Scientific Reports*, 3, Article 2983. https://doi.org/10.1038/srep02983

Liu, Y, Sui, Z, Kang, C, & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS One, 9* Article e86026. https://doi.org/10.1371/journal.pone.0086026

Liu, E., & Yan, X. (2019). New parameter-free mobility model: Opportunity priority selection model. *Physica A*, 526. https://doi.org/10.1016/j.physa.2019.04.259.

Masucci, A., Serras, J., Johansson, A., & Batty, M. (2013). Gravity versus radiation models: on the importance of scale and heterogeneity in commuting fows. *Physical Review*, 88(2). https://doi.org/10.1103/PhysRevE.88.022812

Maris, M., Jovacik, M., & Fazikova (2019). Commuting Trends and Patterns Behind the Regional Imbalances in Slovakia. *European Journal of Geography* 10 (1).

Marshall, J.M., Wu, S.L., Sanchez C., Kiware, S.S., Ndhlovu, M., Ouédraogo, A.L., Touré, M.B., Sturrock, H. J., Ghani, A. C., & Ferguson, N. M. (2018). Mathematical models of human mobility of relevance to malaria transmission in Africa. S*cientific Reports, 8, Article*. https://doi.org/10.1038/s41598-018-26023-1

McCune, Bruce & Grace, James (2002): *Analysis of Ecological Communities.* MjM Software Design. Gleneden Beach.

Multrus F., Majer, S., Bargel, T., & Schmidt, M. (2017). Studiensituation und studentische Orientierungen. In Bundesministerium fuer Bildung und Forschung (BMBF). *13. Studierendensurvey an Universitäten und Fachhochschulen.* Bonn, Berlin.

Murat, C.H. (2010). Sample size needed for calibrating trip distribution and behavior of the gravity model. *Journal of Transport Geography*, 18(1), 183–190. https://doi.org/10.1016/j.jtrangeo.2009.05.013

Nutz, M. (1991). *Räumliche Mobilität der Studierenden und Struktur des Hochschulwesens in der Bundesrepublik Deutschland.* Köln.

Odlyzko, A. (2015). The forgotten discovery of gravity models and the inefficiency of early railway networks. *OEconomia,* 5, 157–192. https://doi.org/10.4000/oeconomia.1684

Pan, R.K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports*,2, Article 902. https://doi.org/10.1038/srep00902

Rees, P. (1986). A geographical forecast of the demand for student places. *Transactions of the Institute of British Geographers*, 5-26.

Ren, Y., Ercsey-Ravasz, M., Wang, P., González, M. C., & Toroczkai, Z. (2014). Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications*, 5, Article 5347. https://doi.org/10.1038/ncomms6347

Sá, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: a gravity model approach. *Regional Studies*, 38(4), 375-392. https://doi.org/10.1080/03434002000213905

Sallah, K., Giorgi, R., Bengtsson, L., Lu, X., Wetter, E., Adrien, P., Rebaudet, S., Piarroux, R., & Gaudart, J. (2017). Mathematical models for predicting human mobility in the context of infectious disease spread: introducing the impedance model. *International Journal of Health Geographics, 1*6, Article 42. https://doi.org/10.1186/s12942-017-0115-7

Simini, F., Gonzalez, M.C., Maritan, A., & Barabasi, A.L. (2012). A universal model for mobility and migration patterns. *Nature*, 484, 96-100. https://doi.org/10.1038/nature10856

Soerensen T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biologiske Skriftler* 5, 1-34.

Statistisches Bundesamt (2008-2017). *Fachserie 11 Reihe 4.1. Bildung und Kultur. Studierende an Hochschulen. Wintersemester 2007/2008 – Wintersemester 2016/2017.* Wiesbaden.

Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C.M., Blondel, V., Smoreda, Z., González, M., & Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Computational Biology*, 10, Article e1003716. https://doi.org/10.1371/journal.pcbi.1003716

Thomas, T., & Tutert, S. (2013). An empirical model for trip distribution of commuters in the netherlands: transferability in time and space reconsidered. *Journal of Transport Geography*. 26, 158–165. https://doi.org/10.1016/j.jtrangeo.2012.09.005
Tuite, A. R., Thomas-Bachli, A., Acosta, H., Bhatia, D., Huber, C. Petrasek, K., Watts, A., Yong, J.H.E., Bogoch, I. I., & Khan, K. (2018). Infectious disease implications of large-scale migration of Venezuelan nationals. *Journal of Travel Medicine*, 25(1). https://doi.org/10.1093/jtm/tay077
Viboud, C., Bjørnstad, O.N., Smith, D.L., Simonsen, L., Miller, M.A., & Grenfell, B.T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science,* 312(5772), 447–451.
Weisser, R. (2019). How Personality Shapes Study Location Choices. *Research in Higher Education*,61, 88-116. https://doi.org/10.1007/s11162-019-09550-2
Wilson, A., (1970). *Entropy in urban and regional modelling.* Pion, London.
Wissenschaftliche Dienste des Deutschen Bundestages (2006). Der Studentenberg –Kollaps der Universitäten oder Illusion? Ein kritischer Beitrag zur aktuellen Diskussion. Retrieved April 2, 2021, from https://www.bundestag.de/resource/blob/418880/251afebe1c84c24d81ad39c8bbf34334/WD-8-212-06-pdf-data.pdf
Yan, X-Y., Zhao, C., Fan, Y., Di, Z., & Wang, W-X. (2014). Universal predictability of mobility patterns in cities. *J. R. Soc. Interface*, 11(100). http://doi.org/10.1098/rsif.2014.0834
Yang, Y., Herrera, C., Eagle, N., & Gonzalez, M.C. (2014). Limits of predictability in commuting ows in the absence of data for calibration. *Scientific Reports*, 4, Article 5662. https://doi.org/10.1038/srep05662
Zipf, G.K. (1946). The P1 P2/D hypothesis: on the intercity movement of persons. *American Sociological Review*, 11, 677–686.

European Journal of Geography