An improved radiation model and its applicability for understanding commuting patterns in Hungary

Varga, Levente and Tóth, Géza and Néda, Zoltán

Department of Physics, Babeș-Bolyai University, Cluj-Napoca, Romania, Hungarian Central Statistical Office, Budapest, Hungary

February 2017

An improved radiation model and its applicability for understanding commuting patterns in Hungary

Levente Varga
Department of Physics,
Babeş-Bolyai University,
Cluj-Napoca, Romania
E-mail:
vargal@cs.ubbcluj.ro

Géza Tóth
Hungarian Central
Statistical Office,
Budapest, Hungary
E-mail:
geza.toth@ksh.hu

Zoltán Néda
Department of Physics,
Babeş-Bolyai University,
Cluj-Napoca, Romania
E-mail:
zneda@phys.ubbcluj.ro

Several empirical models aimed at describing human mobility have been proposed in the past. Most of them are based on an unjustified analogy, with a focus on gravity and physical vector or scalar fields. Recently, however, statistical physicists introduced a new category of models that are theoretically motivated by a few simple and reasonable socio-economic assumptions. *The Radiation Model* (Simini et al. 2012) and the *Radiation Model with Selection* (Simini–Maritan–Néda 2013) are such successful approaches. Here, we introduce a new version of the radiation model, the *Travel Cost Optimized Radiation Model*, and test its applicability for describing the commuting patterns in Hungary. We compare critically the performance of this model with the results of the previous radiation type models.

Introduction

Human mobility patterns form the focus of many studies in sociology and human geography (Varga et al. 2016), owing to their immediate impact on the economy (See literature review, for example, Ritchey 1976). Several electronically available large-scale and freely downloadable data sets help researchers unveil universalities and test their models (Barthélemy 2010, Bazzani et al. 2010). Recently, the statistical physics

community has shown an interest in the impact of human mobility patterns on the economy, exploiting the fact that many models and modelling methods that are well known to physicists can be used for understanding the spatiotemporal universalities observed in human mobility (Brockmann–Hufnagel–Geisel 2006, González–Hidalgo–Barabási 2008).

The continuous growth of big cities (Bettencourt–West 2010), spatial concentration of industries around them, and rapid development of quick and affordable travelling mediums make commuting an increasingly important social phenomena, affecting a large portion of the population (Cohen et al. 2008). An understanding of the statistical laws governing the commuting flux serves both practical and scientific purposes. Uncovering some universally observable features will facilitate better planning for new travel routes, traffic optimisation, and prevention of failures in mobility fluxes. From the viewpoint of fundamental research, a successful modelling paradigm will reveal important psychological and socio-economic factors governing the job and housing selection of individuals.

Physics motivated the most popular models that attempted to fit the distribution of the mobility fluxes between the settlements. In this context, one can refer to the gravity model (for a review, please consult Barthélemy 2010) that along with Newton's law of gravitation exploits a phenomenological analogy. According to the assumption of the gravity models, the commuter fluxes between two settlements are proportional to their populations and inversely proportional with the square of the commuting distance. Generalized versions of the model propose power-law dependencies with freely defined fitting exponents for the population of the host and target cities and the distance between them. The generalized potential model is also an attempt to use the classical methodology of physics for understanding commuting fluxes (Anderson 1956, Lukermann–Porter 1960). Other well-known models that are not necessarily motivated by physics are the intervening opportunities model (Stouffer 1940) and the random utility model (Block–Marschak 1960). The radiation model, which was recently proposed by Simini et al., is built on a new paradigm and it starts by using a sociological driving force for the individuals—an optimization to improve their current income (Simini et al. 2012). The initial model, which in principle is a one-parameter model, was formulated for a discrete settlement structure. It was later generalized (Simini et al. 2013) for a continuous population distribution and made more complex by allowing a realistic job selection for the individuals. The new version of the model, the radiation model with selection, is a two-parameter model that offered an improved fit for the commuting data in the USA.

In the present work, we offer a new generalization of the radiation model by taking into account the travel cost involved in the job selection process. Our model (the travel cost optimised radiation model) leads again to a compact analytical formula involving two fitting parameters. Here, we test the applicability of this model for the fluxes derived from a complete commuting database in Hungary. We also compare the performance of this new model with the previously known radiation type models.

The model

The **original radiation model (RM)** starts from the assumption that commuting fluxes reflect, in some manner, optimisation reasoning in the jobseekers' choice for employment. The main hypothesis is that a jobseeker will accept the closest job offer that has a better salary than the one that it is currently available in his/her location. Owing to this requirement each individual travels to the nearest location where she/he can improve her/his current benefits. Let us assume that the cumulative distribution function of the incomes (benefits) is given by $p_{\leq}(z)$ (the probability that the income of one person is smaller or equal than z). Mathematically written, the main assumption of the original radiation model is that the probability $P_{>}(z \mid a)$ that an individual with benefit z refuses the closest a number of job offers is:

$$P_{>}(z \mid a) = p_{\leq}(z)^{a},$$ (1)

In writing this probability, we assumed that the rejection of the a number of closest job offers with benefits less or equal to z are independent events. This is true in case there are in total more number of job offers than the number a.

Using the incomes probability density function, $p(z) = \dfrac{\delta p_{\leq}(z)}{\delta z}$ we can now calculate the $P_{>}(a)$ probability of not accepting the closest a number of job offers by integrating over the benefits:

$$P_{>}(a) = \int_{0}^{\infty} P_{>}(z \mid a) p(z) dz =$$

$$\int_{0}^{\infty} P_{>}(z \mid a) \frac{\delta p_{\leq}(z)}{\delta z} dz = \int_{0}^{1} p_{\leq}(z)^{a} dp_{\leq}(z) = \frac{1}{a+1}$$ (2)

If we assume now that the jobseekers are selective in their choices and they are willing to accept better offers only with a probability $1-\lambda$, we get (Simini et al. 2013):

$$P_{>}(z \mid a) = \left(1 - (1-\lambda)(1 - p_{\leq}(z))\right)^{a} = \left(\lambda + (1-\lambda) p_{\leq}(z)\right)^{a}$$ (3)

Alternatively, the assumption behind this generalization can be interpreted as a fact that the jobseekers are aware only of a fraction $1-\lambda$ of the available job offers. In such conditions, we obtain:

$$P_{>}(a) = \int_{0}^{\infty} \left[\lambda + (1-\lambda) p_{\leq}(z)\right]^{a} p(z) dz =$$

$$\int_{0}^{1} \left[\lambda + (1-\lambda) p_{\leq}(z)\right]^{a} dp_{\leq}(z) = \frac{1 - \lambda^{a+1}}{(a+1)(1-\lambda)}$$ (4)

This yields a more complex formula for the commuting fluxes, since an addition λ parameter appears. This model is known as the **radiation model with selection (RMwS)**. For $\lambda = 0$, we recover the original radiation model. Simini et al. in 2013

showed that this model describes the distribution of the commuting fluxes for the USA better than the simple radiation model.

Subsequently, we introduced another extension for the original radiation model, which we named as the **travel cost optimized radiation model (TCORM).** The radiation model assumes that commuters will accept jobs that offer gains, relative to their present income, independently, on the distance between their residence and workplace. It presents a very weak condition for commuting, and our proposal is that individuals must choose to commute if there is a net gain in their salaries after subtracting the travel costs. However, travel costs depend on the distance travelled. Therefore, in principle, the radiation model has to be modified so that, in addition to the transited job offers *(a)*, the travel distance *(r)* is considered as an important factor. In this view, the working paradigm of the radiation model is reconciled with basic assumptions in the radiation and potential models. According to this assumption and *following the reasoning of the radiation model,* we can modify the starting hypothesis (1) as:

$$P_> (z|a) = p_\le (z)^{a_0} \cdot p_\le (z + \alpha r_1)^{a_1} \cdot ... \cdot p_\le (z + \alpha r_q)^{a_q} \tag{5}$$

So that $\sum_{j=1}^{q} a_j = a$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (6)

In the above hypothesis, α is the parameter that governs the distance-dependent travel costs. We denoted the job openings inside settlement j by a_j and distance between originating point to the destination point inside the settlement by r_j.

According to this, in order to proceed we had to make some assumptions about the nature of the $p_\le (z)$ cumulative income distribution. Definitely, $p_\le (z)$ has to be a monotonically increasing function, with the condition $p_\le (0) = 0$ and $p_\le (\infty) = 1$. Empirical data on an estimated income distribution for Hungary shows that a kernel of the form $p_\le (z) = 1 - e^{-\gamma z}$ gives a reasonable approximation (see Figure 1 where we have plotted $1 - p_\le (z)$ on lognormal scale, suggesting a linear dependence). This form also satisfies the above-imposed conditions. Working with such kernel, one can write:

$$p_\le (z + \alpha r_i)^{a_i} = (1 - e^{-\gamma(z + \alpha r_i)})^{a_i} \tag{7}$$

We assumed that the cost of travel is small in comparison to the salary, and hence $\alpha r_i << z$. It also means that $\alpha << 1$. In such case, we can make a Taylor expansion around the value of z:

$$(1 - e^{-\gamma(z + \alpha r_i)})^{a_i} \approx (1 - e^{-\gamma z})^{a_i} \left(1 + \alpha \gamma r_i a_i \frac{e^{-\gamma z}}{1 - e^{-\gamma z}} \right) \tag{8}$$

According to this, keeping only up to first order terms in α, we obtained:

$$P_> (z|a) \approx \left(1 - e^{-\gamma z}\right)^a \left(1 + \alpha\gamma \frac{e^{-\gamma z}}{1 - e^{-\gamma z}} \sum_{j=1}^{q} a_j r_j\right) = \left(1 - e^{-\gamma z}\right)^a \left(1 + \alpha\gamma a\langle r(a)\rangle \frac{e^{-\gamma z}}{1 - e^{-\gamma z}}\right) \quad (9)$$

Where, $\langle r(a)\rangle$ denotes the average distance of the a number of job openings from the centre.

Based on the above derivation, we expressed the $P_>(a, r)$ probability as:

$$P_> (a) = \int_0^1 p_\leq(z)^a \left(1 - \alpha\gamma a\langle r(a)\rangle\left(1 - \frac{1}{p_\leq(z)}\right)\right) dp_\leq(z) =$$
$$\int_0^1 x^a \left(1 - \alpha\gamma a\langle r(a)\rangle\left(1 - \frac{1}{x}\right)\right) dx \quad (10)$$

Performing the integral, we obtained:

$$P_> (a) = \frac{1 + \alpha\gamma\langle r(a)\rangle}{(a + 1)} \quad (11)$$

As it was considered in previous work (Simini et al. 2012), we also assumed that job offers in a territory is linearly proportional with the population, i.e. $a = \mu w$, where w is the population and μ is a proportionality constant. According to this, we obtain again a two-parameter model, yielding the following with $\beta = \alpha\gamma\mu$:

$$P_> (w) = \frac{1 + \beta\langle r(w)\rangle}{(\mu w + 1)} \quad (12)$$

Thus, this gives another two-parameter model extension to the original radiation model.

In case of assuming a uniform distribution of the population in the considered territory, we obtained the following:
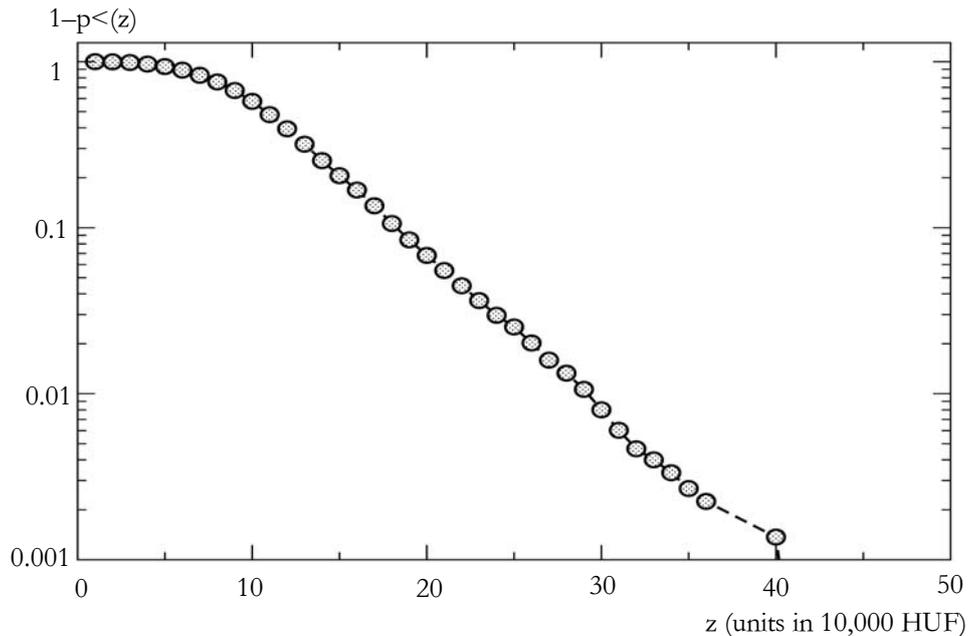
$$\langle r(w)\rangle \infty \sqrt{w} \quad (13)$$

which yielded the following:

$$P_> (w) = \frac{1 + \lambda\sqrt{w}}{(\mu w + 1)} \quad (14)$$

Where, λ is again proportionality constant.

Figure 1

**Estimated cumulative distribution of incomes in Hungary in the year 2014
(please note the logarithmic vertical axes)
The linear trend suggests an exponential decay**


Source: HCSO.

Data analyses

As a basis for our experimental investigation, we processed a complete commuter and population database from Hungary. We analysed the 2011 population census data using Q=**3.176** settlements (nodes) (see white filled circles in Figure 2) and **81.664** commuter routes (edges) (see cyan lines between white filled circles in Figure 2) on the territory of Hungary. In our calculation, we also used data with a square kilometre resolution for population distribution in the census year 2011. The total population of Hungary, according to the 2011 census year, is **W=9.972.000**.

During our data processing, we select one by one the settlements i from where commuting starts and construct the boundaries of the disks with radius $d(i, j)$, reaching up to those settlements j, where the commuting is done (see Figure 2). We counted the population $w_i[j]$ inside the specified disk with radius $d(i, j)$ and recorded the number of commuters $n_i(j)$ starting from settlement i and traveling to settlement j.

Figure 2

Settlements in the commuting network of Hungary. Here, we illustrate the data processing method. Disks of different radius *d(i, j)*, starting from a given settlement and reaching the other *j* settlements, are constructed. The population *wᵢ[j]* inside these disks and the commuter number, starting from settlement *i* and traveling to settlement *j*, *nᵢ(j)*, is recorded.
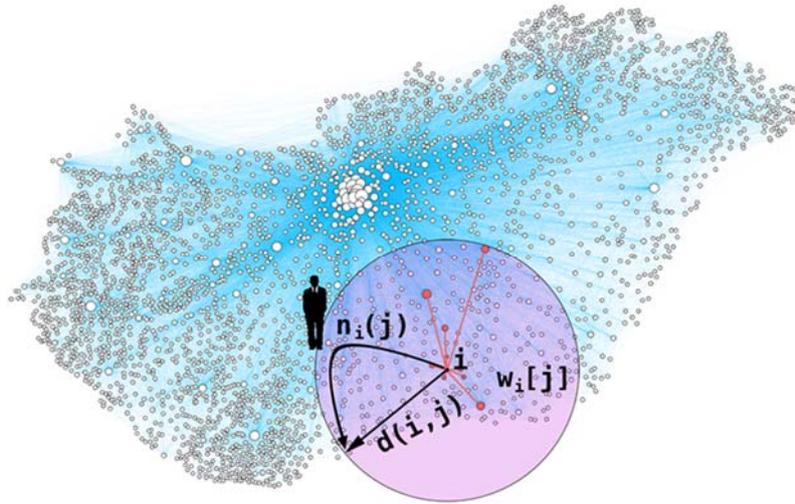
Source: Edited by the authors.

With *d(i, j)*, *nᵢ(j)*, and *wᵢ[j]* data for all the settlement pairs *(i, j)*, where $n_i(j) \neq 0$, we proceeded to the construction of the *P>(w)* probabilities. The number of workers that have their residence in settlement *i* are denoted by N_i. We have expressed this as follows:

$$N_i = \sum_{j=1}^{Q} n_i(j) \tag{15}$$

We denoted the total population in the considered territory as *W*.

We ordered the settlements according to their distance relative to *i*. Let $h_i^{[k]}$ be the index of the settlement that is the kth one in this row (for example, $h_i^{[1]}$ is the index of the settlement that is the closest to settlement *i* and $h_i^{[2]}$ is the index of the settlement that is the second closest to *i*). Here, we introduced the *s(i, w)* integer number in the following manner:

$$s(i,w) \in \{1,2,3,...,Q\} \tag{16}$$

which is defined in the following manner for $w \leq W$:

$$w_i\left[h_i^{[s(i,w)-1]}\right] < w \leq w_i\left[h_i^{[s(i,w)]}\right] \tag{17}$$

(We record that Q denotes the total number of settlements and W the total population in the studied territory.) If no such number exists, then we will consider *s(i, w)=Q*.

In layman terms, *s(i, w)* denotes the smallest number of settlements that have to be considered so that the total population is larger (or equal) than W inside the disk that contains these settlements (centred to *i*).

The probability that the commuters from *i* are a transitioning population W can be written as:

$$P_{>}^{i}(w) = \frac{1}{N_i}\left[N_i - \sum_{k=1}^{s(i,w)} n_i\left(b_i^{[k]}\right) \right] \tag{18}$$

The $\langle r_i(w) \rangle$ curve can also be constructed as:

$$\langle r_i(w) \rangle = \frac{\sum_{k=1}^{s(i,w)} \left[w_i\left[b_i^{[k]}\right] - w_i\left[b_i^{[k-1]}\right] \right] \cdot \frac{1}{2}\left[d(i, b_i^{[k]}) + d(i, b_i^{[k-1]}) \right]}{w_i[s(i,w)]} \tag{19}$$

After obtaining the quantities for all settlements, we constructed the desired curves as follows:

$$P_{>}(w) = \frac{1}{Q}\sum_{i=1}^{Q} P_{>}^{i}(w) \tag{20}$$

$$\langle r(w) \rangle = \frac{1}{Q}\sum_{i=1}^{Q} \langle r_i(w) \rangle \tag{21}$$

After constructing a data analyses algorithm based on the above calculations, we proceeded to construct experimentally the *P>(w)* and $\langle r(w) \rangle$ curves.

Results and discussions

The values for *P>(w)* computed from the census data are compared with the best fits obtained with the original radiation model (RM), the Radiation model with selection (RMwS), and the travel cost optimized radiation model (TCORM). The boundary effects become important for large *w* values (the disks centred on the settlements are incomplete due to the fact that they extend over the borders of Hungary). In addition, to minimise these effects we considered the data only up to w=1.000.000. This cut does not affect the commuters originating from Budapest, since the capital is not considered as a single settlement in this study, but is divided into sectors.

In order to use the TCORM model, we first had to investigate experimentally the $\langle r(w) \rangle$ function and check whether our assumption (13) holds. This is calculated by following the formulae (19) and (21) and it is plotted in Figure 3. The results reveal

that the approximation (13) works well for distances that are larger than 3 km and smaller than 100 km (which is a relevant distance scale in commuting). If one would like to get an even better theoretical verification for formula (12), then the numerically computed $\langle r(w) \rangle$ could be used instead of (13).

Figure 3

The mean distance of a population w, in a disk-like geometry, relative to the centre of the disk (the mathematically defined $\langle r(w) \rangle$ as a function of w). The blue line is calculated using formulae (19) and (21) from census data for Hungary in year 2011. The red line shows the region where the power-law dependence is very good. The dashed light blue line shows a power-law trend corresponding to an exponent 0.53 (please note the logarithmic axes).
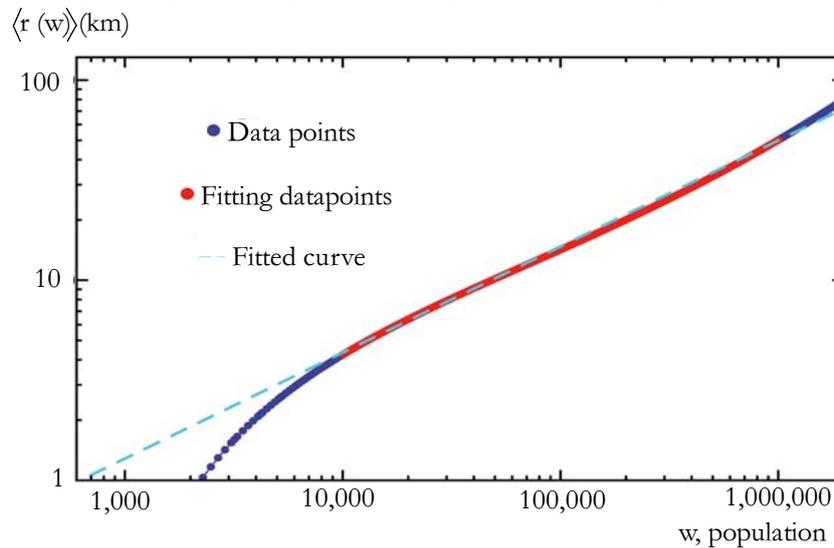
Source: Edited by the authors.

In order to verify the applicability of different variants of the radiation model we first compute the $P_>(w)$ experimental curve from the 2011 census data by following the steps specified in (18) and (20). We compare than visually the best fits to this line by using the:

RM prediction: $P_>^{RM}(w) = \dfrac{1}{\mu w + 1}$ (22) (one parameter fit)

RMwS prediction: $P_>^{RMwS}(w) = \dfrac{1 - \lambda^{\mu w + 1}}{(\mu w + 1)(1 - \lambda)}$ (23) (a two parameter fit)

TCORM prediction: $P_>^{TCORM}(w) = \dfrac{1 + \lambda\sqrt{w}}{(\mu w + 1)}$ (24) (a two parameter fit)

The best fit to the experimental curve is done numerically and here we rely only on a visual comparison. The results are plotted on Figure 4. In Figure 4, we use a lognormal scale in order to visualize better the differences at small values.

Figure 4

Comparison of radiation type models prediction with experimental data for Hungary. The thick blue line $P_>(w)$ is calculated for commuters' data in Hungary using the 2011 census data and following the algorithm summarized by equations (18) and (20). These results are visually compared with the best fits obtained for the RM model (eq. 22), RMwS model (eq. 23) and TCORM model (eq. 24). Please note the logarithmic vertical axes.

$P_>(w)$, probability
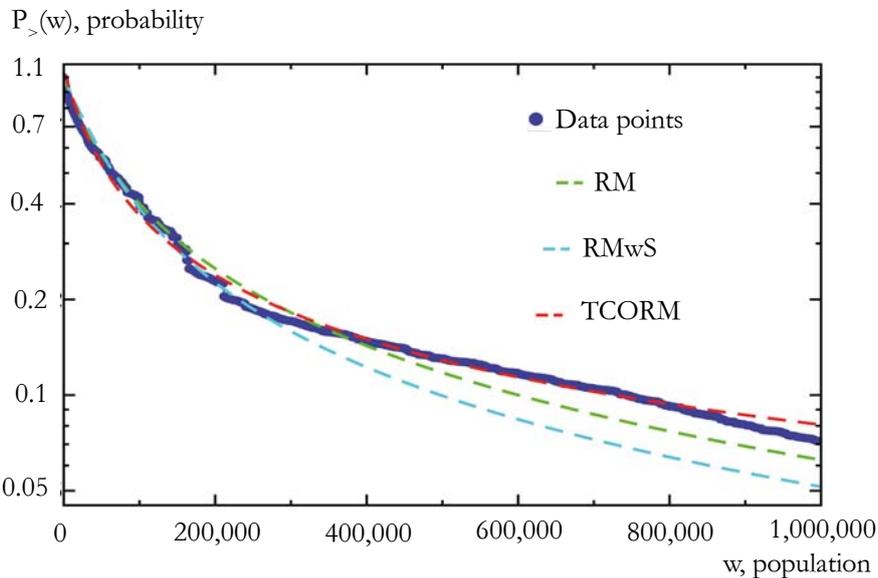
Source: Calculated by the authors.

The best fits parameters obtained are as follows: $\mu=0.000015$ for RM (eq. 22), $\mu=0.000028$ and $\lambda=0.33$ for RMwS (eq. 23), and $\mu=0.00003$ and $\lambda=0.0015$ for TCORM (eq. 24).

A visual examination also clearly indicates that the TCORM model performs better than the simple RM and RMwS model. The TCORM model has the capability of offering a visually good fit for the whole interval, while the simple RM model and RMwS are only capable of describing a portion of the $w \in [1, 1.000.000]$ interval. It is apparent that a that a two-parameter fit is more successful than a one-parameter fit. Accordingly, it is also natural that TCORM would perform better than the simple RM. It is important to note that the TCORM seems to be a much better approximation than RMwS for describing commuter data for Hungary, although both the models are two-parameter regressions. However, future commuter data for other

geographical regions must also be checked in order to affirm that TCORM is a better approximation for commuter fluxes than RMwS.

Conclusion

We studied the distribution of commuter fluxes between the settlements of Hungary using the census data from 2011. In order to explain the observed data, we derived a novel extension for the classical radiation model (RM). In our travel cost optimized radiation model (TCORM), we take into account the travel cost for commuting in the benefit optimisation. Realistically, where the travel cost is much smaller than the benefits, we obtained an analytical solution and a compact formula for the commuter fluxes. The complexity of this description is comparable with the complexity of the radiation model with selection (RMwS), wherein both the approaches focus on fitting a two-parameter model to experimental data. We find that our novel TCORM model performs better than the RMwS model in describing the commuting data for Hungary. This observation has to be thoroughly examined in the future for other geographical regions as well.

Acknowledgement

REFERENCES

ANDERSON, T. R. (1956): Potential models and the spatial distribution of population *Papers in Regional Science* 2 (1): 175–182.

BARTHÉLEMY, M. (2010): Spatial networks *Physics Reports* 499: 1–101.

BAZZANI, A.–GIORGINI, B.–RAMBALDI, S.–GALLOTTI, R.–GIOVANNINI, L. (2010): Statistical laws in urban mobility from microscopic GPS data in the area of Florence *Journal of Statistical Mechanics: Theory and Experiment* 2010: P05001.

BETTENCOURT, L.–WEST, G. (2010): A unified theory of urban living *Nature* 467: 912–913.

BLOCK, H.–MARSCHAK, J. (1960): Random orderings and stochastic theories of responses *Contributions to probability and statistics* 2: 97–132.

BROCKMANN, D.–HUFNAGEL, L.–GEISEL, T. (2006): The scaling laws of human travel *Nature* 439: 462–465.

COHEN, J. E.–ROIG, M.–REUMAN, D.C.–GOGWILT, C. (2008): International migration beyond gravity: A statistical model for use in population projections *Proceedings of the National Academy of Sciences* 105 (40): 15269–15274.

GONZÁLEZ, M.C.–HIDALGO, C.A.–BARABÁSI, A. L. (2008): Understanding individual human mobility patterns *Nature* 453: 779–782.

LUKERMANN, F.–PORTER, P.W. (1960): Gravity and potential models in economic geography *Annals of the Association of American Geographers* 50 (4): 493–504.

RITCHEY, P. N. (1976): Explanations of migration *Annual Review of Sociology* 2: 363–404.

SIMINI, F.–GONZÁLEZ, M.C.–MARITAN, A.–BARABÁSI, A.L. (2012): A universal model for mobility and migration patterns *Nature* 484: 96.

SIMINI, F.–MARITAN, A.–NÉDA, Z. (2013): Human Mobility in a Continuum Approach. *PLoS ONE* 8(3): e60069.

STOUFFER, S.A. (1940): Intervening opportunities: a theory relating mobility and distance *American Sociological Review* 5 (6): 845–867.

VARGA, L.–KOVÁCS, A.–TÓTH, G.–PAPP, I.–NÉDA, Z. (2016): Further We Travel the Faster We Go *PLoS ONE* 11 (2): e0148913.