UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 139 - July 08

Macro and Micro Dynamics of City Size Distributions: The Case of Israel

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Macro and Micro Dynamics of City Size Distributions: The Case of Israel

L. Benguigui[1], E. Blumenfeld-Lieberthal[2,A] and M. Batty[3]

July 2008

[1]Solid State Institute and Department of Physics,
Technion-Israel Institute of Technology, Haifa 32000, Israel

[2]Environmental Simulation Laboratory,
The Porter School of Environmental Studies and the Faculty of Arts, Tel-Aviv University, Tel-Aviv 69978, Israel.

[3]Centre for Advanced Spatial Analysis (CASA),
University College London, London WC1E 7HB, UK

[A]Corresponding Author: email: efrat.lieberthal@gmail.com

ABSTRACT

We study the distribution of sizes in the Israeli system of cities, using a rank-size representation of population distributions from 1950 to 2005. Based on a multiplicative model of proportionate growth, we develop a quantitative comparison relating the change in the rank-size curves to the change in the real data of Israeli cities during this period. At the level of macro dynamics, there is good agreement between the model and the real data. At the micro level, however, the model is less successful as the mean variation of the cities' rank during the period studied is much larger in the model than in the real data. To illustrate this difference, we use the rank-clock representation.

1 INTRODUCTION

The statistics of sizes is a topic studied in many disciplines across the natural and the social sciences [1-6]. One of the key problems is how to describe mathematically the function which describes the sizes of the objects or entities that compose such distributions. The two most common distributions in the literature are the power law and the lognormal distribution [7, 8, 9] but it is clear that in many cases, both of these distributions do not replicate the shape or form of functions in a satisfactory way. In fact, more accurate distributions lie somewhere between these two options [9, 10].

Among the difficulties concerning the choice of distribution is the problem of the lower tail. In the past, many applications have simply dealt with the largest entities usually because data has been available only for the largest, or sometimes because it is assumed that the most important entities are those that are the largest. The lower tail, or long tail as it sometimes called, of the frequency distribution of sizes is often disregarded – cut off, and it is clear that by changing the size of the lower tail of the entities' size distribution, the function which fits the distribution will also change. It is not easy to choose a criterion to define the cut off for the lower tail and very often it is chosen arbitrarily.

It is thus a major problem in exploring size statistics to relate the properties of a particular set of entities (for example, incomes, stock market values, populations of cities, frequencies of words and letters that comprise languages etc.) to a function that describes accurately their size distribution. Several models attempt to solve this problem by relating the entities' size distributions to some hypotheses concerning their behavior [10]. These models, however, usually examine the size distribution at a particular time, thus grounding the analysis in comparative statics, often beginning with some arbitrary initial distribution as the starting point for simulation and iterating the model until some equilibrium state is reached.

The purpose of this paper is twofold; in the first part, we investigate the *dynamics* of the size distribution (without searching for an equilibrium) and show how it is possible to relate the change in the distribution to the properties of the entities. For that, we use an approach which defines a distribution using a new exponent α [8]

presenting a simulation based on a model we have recently developed [9, 10]. In the second part, we study the dynamics of the Israeli system of cities, first at the macro level and then at the micro, comparing the micro-dynamics in the real data with that in the simulation.

Most of the existing work in the field looks for an agreement between the empirical distributions and that given by the models through measures of the accuracy of the model. This, in fact, seems sufficient in the case of a static model. When adding dynamics to the model, however, one has to consider not only the change in the distribution but the temporal variation of the entities as well. Recently, Batty [12, 13] proposed the rank-clock representation to study the micro dynamics of systems with entities that appear to generate homogeneous rank size distributions at the macro scale. This representation follows the dynamics of individual entities within the system and here we use it to check the validity of our work in terms of micro dynamics.

In this paper, we use the Israeli system of cities as our case study in examining the validity of our simulation. In the first part of the paper, we compare the macro dynamics of the Israeli system of cities from 1950 to 2005 at three snapshots of time – 1961, 1983 and 2005 – showing that the temporal change in the distribution is related to the variation of the number of cities in Israel through time. In the second part of the paper, we compare the micro dynamics of the simulation with the Israeli system of cities from 1950 to 2005. Sections 2–6 concern the macro dynamics of the simulation and the real system of cities while section 7 presents their micro dynamics.

In the next section, we present the new exponent α and its correlation with the distributions. In section 3, we then cover the data concerning the cities of Israel followed by an outline of the model (in section 4) and its application to Israeli cities (in section 5). In section 6, we present the results of the model at the macro level while section 7 concerns the micro dynamics of the simulation in comparison to the real system of Israeli cities. Finally we discuss the results of the model, introducing the rank clock analysis which makes comparisons between the model and the real data. Surprisingly, the model provides a rather good description of the distributions at the macro level, but fails to give a sufficiently accurate analysis of the individual changes in the entities at the micro level.

2 A NEW EXPONENT

Recently, we proposed a phenomenological approach [9] when analyzing the size distribution of entities. We based our approach on the three equivalent representations of the size distribution [11]:

- The density function $D(S)$ which gives the number of entities with size between S and $S+dS$.
- The cumulative function $P(S)$ which gives the number of entities with a size larger or equal (or smaller) than a given S. These two functions $D(S)$ and $P(S)$ can also be expressed in relative terms. The two relative functions do not give the number of entities but rather the percentage (or probability) of the total number of entities.
- The rank size representation which is transformed into the logarithm of the size ($y = \ln Size$) and plotted as a function of the logarithm of the rank ($x = \ln Rank$). The function $y(x)$ is called the rank-size curve. When the relationship between the size and rank of the entities can be expressed by the function $y \sim x^n$ where n is a power of the function, the distribution is referred to popularly as Zipf's law after Zipf [14] who presented a graphical and rather forceful summary of such relationships. This relationship is represented pictorially by a linear equation plotted as a double logarithmic graph .

Our proposed equation concerns the function $y(x)$. We propose to analyze the rank-size curve for a system of entities using the following expression:

$$y = y_0 - H(1-\alpha)\mu[b + H(1-\alpha)x]^{\alpha} \quad , \qquad (1)$$

where $y = y(x)$. $H(1-\alpha)$ is the Heaviside function, equal to -1 if $\alpha < 1$ and to 1 if $\alpha \geq 1$. y_0, b, and μ are parameters and the exponent α can be smaller, equal or larger than 1. In the case $\alpha = 1$, Zipf's law is recovered, i.e. there is a linear relation between x and y. When $\alpha \neq 1$, the curve $y(x)$ has different shapes following the value of α which we call the "shape exponent". This means that each value of α defines a particular distribution.

3 THE CITIES OF ISRAEL

We have analyzed the rank size distribution of all the settlements in Israel from 1950 to 2005 [15] and found that qualitatively, they all have the same shape. In Fig.1 we present the rank-size curve of all settlements in the years 1961, 1983, and 2005. These three curves have the same shape and the only observed differences are the shifts of the curves upwards with time. The distribution demonstrates discontinuity around the value of population equal to 1,000. In other words, the distributions can be divided into two parts around this value. Based on the above, we defined the lower tail of the distribution for all settlements with population smaller than 1,000 for the entire period. The inset of Fig.1 presents the rank size curves for the years 1961, 1983, and 2005, after the exclusion of their lower tails (based on settlements with populations smaller than 1,000).

Similar to several other cases [9], the fit of the distributions to equation (1) is very good as demonstrated in Fig.1. Fig.2 presents the variations of the exponent α with time. The exponent α is larger than 1 in 1961 and changes to values lower that 1 in the following years. In Fig.3 we show the change in the number of cities (settlements with population larger than 1000) as a function of time. The data fits a quadratic equation with good precision.

4 THE MODEL

The model we have used is based on a computer simulation that we have recently presented [9, 10]. We begin with N_0 cities with population equal to 1. At the first stage, each city grows by random multiplicative growth: one city is chosen at random and its population is changed from step T to step $T+1$ as:

$$S(T+1) = \gamma\, S(T) \qquad (2)$$

where γ is a random variable uniformly distributed between $\gamma_m < 1$ and $\gamma_M > 1$ such that the mean $(\gamma_m + \gamma_M)/2$ is fractionally larger than 1. At the second stage, if the population of a city decreases below 1, the city disappears from the system. At the third stage, a new city is added to the system with the population equal to 1 after K steps. As for the initial cities, if the size of the new city decreases below 1, it disappears from the system.

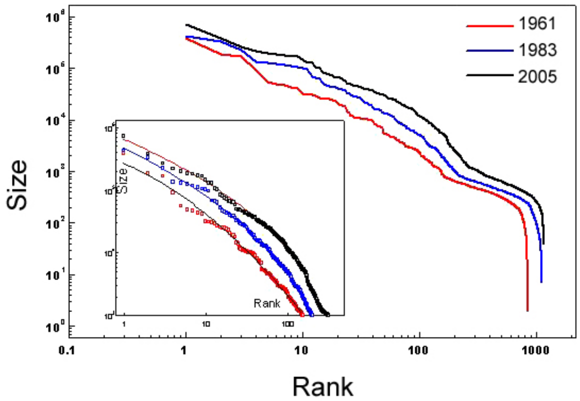

Figure 1: *The rank-size distribution of all the settlements in Israel in the years: 1961, 1983, and 2005. In the inset: the rank size distribution of all the settlements in Israel with populations larger than 1,000.*

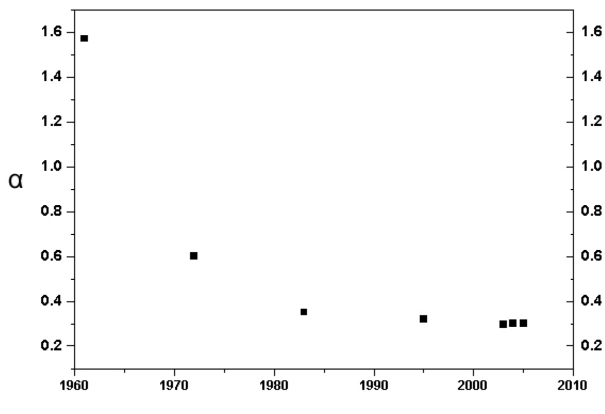

Figure 2*: The variation of the 'shape exponent' α between the years 1961 and 2005 for the Israeli system of cities.*

The distributions that this model yields are dependent on the number of steps T (or in other words, the total growth time) and on the rate of introducing new cities K (for a given γ distribution). If $K > 300$ and is constant, the exponent α changes as a step function when, for small T, it is smaller than 1 and for large T it changes to values larger than 1. For smaller values of K, the exponent α is larger than 1 for small T and becomes equal to 1 for larger T. A significant property of the model involves its statistical basis. For given values of T and K, the results of the model do not always yield the same parameters. When fitting the resulting rank size curves of the model to equation (1), a distribution of such parameters emerges.

An important issue in the model is the definition of the time t. It is not equal to the number of steps (T) since the number of steps (on average) separating two consecutive choices of the same city in the growth process is dependent on the number of cities. The unit of time is chosen as the mean number of steps separating two consecutive choices of the same city. For a number steps ΔT, the interval of time Δt is equal to $\frac{\Delta T}{N}$ or if we consider the continuum limit:

$$dt = \frac{dT}{N} \quad . \tag{3}$$

If we add to this the rate of creating new cities, we get:

$$\frac{dN}{dT} = \frac{1}{K} \quad . \tag{4}$$

If K is constant, it is not difficult to show from equations (3) and (4) that the variation of the number of cities with time is exponential, and is given by $N = p_s N_0 \exp\left(\frac{t}{K}\right)$ where p_S is the probability of a new city surviving, with p_S approximately equal to 0.27.

5 APPLICATION OF THE MODEL TO ISRAELI CITIES

In this section, we show how we use the model to interpret the system of cities in Israel. Our goal is to use the model in order to determine the city size distribution or more precisely the rank-size function for Israel's real system of cities. Since we know the function $N(t)$ is quadratic, we have to find the function $K(T)$ which fits this result. For that, the following system of equations needs to be solved:

$$\frac{dt}{dT} = \frac{1}{N} \quad , \tag{5a}$$

$$\frac{dN}{dT} = \frac{1}{K} \quad , \quad \text{and} \tag{5b}$$

$$N = N_0 + N_1 (t - t_1)^2 \quad . \tag{5c}$$

Analytic solutions to this system of equations is complicated, and thus we propose to find an approximate solution using a heuristic approach. After some systematic trial and error iteration, we generated the following expression: $K(T) = K_0 + K_1 T^{1/3}$, which yields good results, as indicated below. Based on this expression, it is possible to show that $N(t)$ is given by:

$$N = \left(N' - \frac{6K_0^2}{K_1^3} \right) + \left(\frac{3}{2K_1} \right) \left(T^{\frac{1}{3}} - \frac{K_0}{K_1} \right)^2 + \left(\frac{3K_0^2}{K_1^3} \right) \ln \left(K_0 + K_1 T^{\frac{1}{3}} \right) , \tag{6}$$

where N' is an integration constant. Considering that the log term in equation (6) changes very slowly in its dependence on T, it is possible to add it to the constant term and get a good approximation using the following simple expression for N:

$$N = N_0 + N_1 \left[T^{\frac{1}{3}} - (T_1)^{\frac{1}{3}} \right]^2 \quad , \tag{7}$$

where $T_1^{1/3} = K_0 / K_1$. The constants N_0 and N_1 in equation (7) are dependent on the coefficients K_0 and K_1 and also on the integration constant.

To find the relation between the time t and the number of steps T, we integrate equations (5a) and (7) to get:

$$t = \frac{u}{N_1} + B\theta^{-1}\left[u\sqrt{\frac{N_1}{N_0}}\right] + C\ln\left(N_0 + N_1 u^2\right) + t' \quad , \tag{8}$$

where $u = \left(T^{\frac{1}{3}} - T_1^{\frac{1}{3}}\right)$. In equation (8), t' is an integration constant, and the coefficients B and C are dependent on K_0, K_1, and N_0, N_1.

It is possible to consider the second and third terms on the right hand side of equation (8) as constants: the log term because it changes very slowly in its dependence on N and hence on T; and the θ^{-1} term because its argument is larger than 1 (considering the real values of the parameters). The final result suggests that t is linearly dependent on $T^{1/3}$. One can choose the integration constant such the model can be written as follows:

$$t = D\left[T^{\frac{1}{3}} - (T_1)^{\frac{1}{3}}\right] \quad . \tag{9}$$

From equations (7) and (9), it can be deduced that the dependence of N on t is indeed the quadratic equation we expected.

In the following steps, we consider K as constant, choose an initial state for which the number of cities is roughly 150 (see Fig.3) and select the exponent α as approximately 1.6 (see Fig.2). We found that this corresponds to a state with 50 initial cities, $T = 45{,}000$ steps and $K = 80$. For larger values of T, we used the following function to describe K as a function of T: $K = K_0 + K_1 T^{1/3}$ such that the value of

$T_1 = (K_0 / K_1)^3$ is near the value of T in the initial state. Then, we ran the simulation for several values of T where for each T we found the value of N. We also determined the time by graphical integration of the function N^{-1} versus T, and we plotted the rank-size curve determining the shape exponent α by fitting the curve using equation (1).

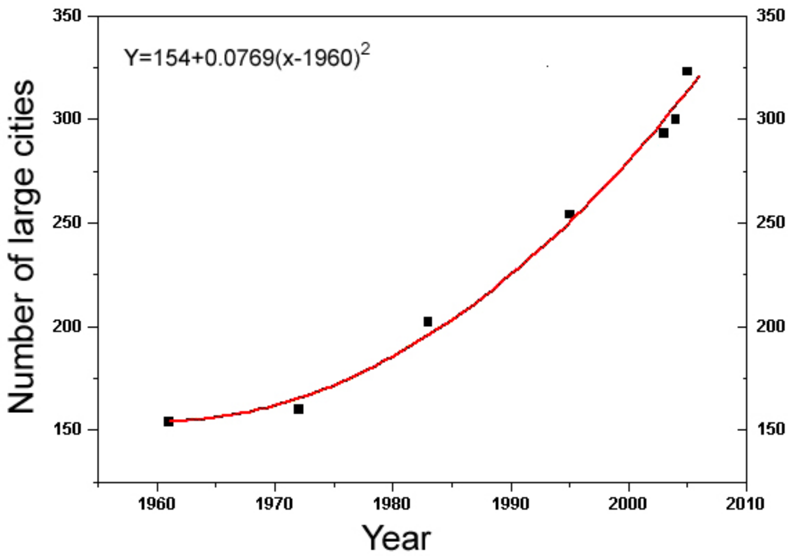

***Figure 3**: The change in the number of cities (with populations larger than 1000) between the years 1961 and 2005 for the Israeli system of cities.*

6 MACRO ANALYSIS OF THE MODEL: THE RANK SIZE CURVE AND THE NUMBER OF CITIES

The outputs of the model are presented in figures 4–8; Fig. 4 presents the relation N versus $T^{1/3}$ where it is clear that the relation in equation (7) is indeed verified. The relation t versus $T^{1/3}$ is presented in Fig. 5 where a linear relationship between time and $T^{1/3}$ is also verified with the same value of $T_1^{1/3}$ found in Fig.4. Fig. 6 shows N versus t where the quadratic relation is also found. Note that the values of N and of the coefficients N_0 and N_1 found in the model, are very close to the ones in the real data. In Fig. 7 we present the exponent α versus t. Here too, there is a very good quantitative agreement between the model and the real data. Finally, we show the rank-size curves of the model in Fig.8. These curves are qualitatively very similar too to the ones that result from the real data. Based on the above, we think the model provides a good description of the evolution of the Israeli system of cities (but clearly only when considering cities with populations larger than 1,000). More particularly, the qualitative change in the distribution (i.e. in the exponent α) is a direct consequence of the variation in the number of cities with time.

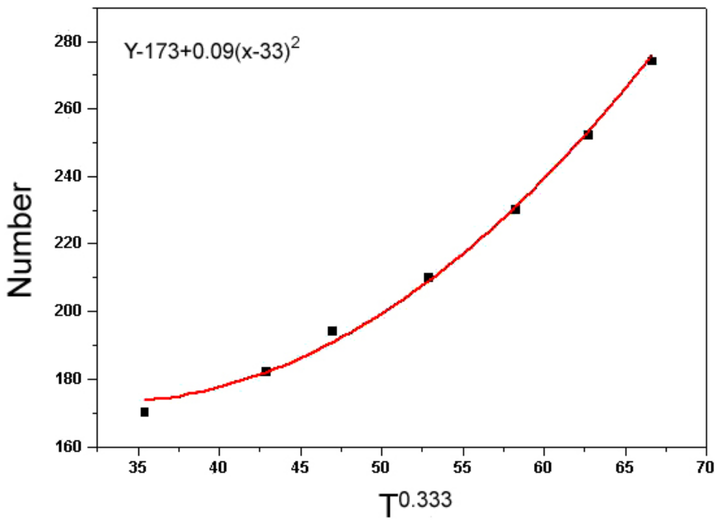

Figure 4: *Results of the model: the number of cities* N *versus* $T^{1/3}$, *where* T *represents the number of steps.*

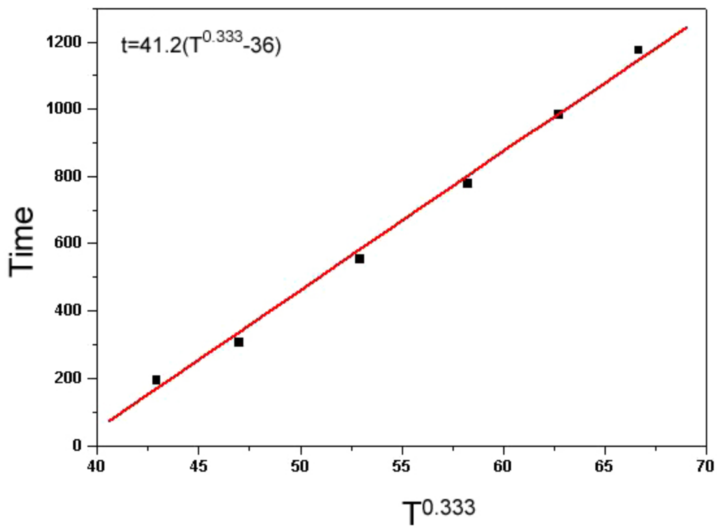

Figure 5: *Results of the model: time* t *versus* $T^{1/3}$, *where* T *represents the number of steps.*

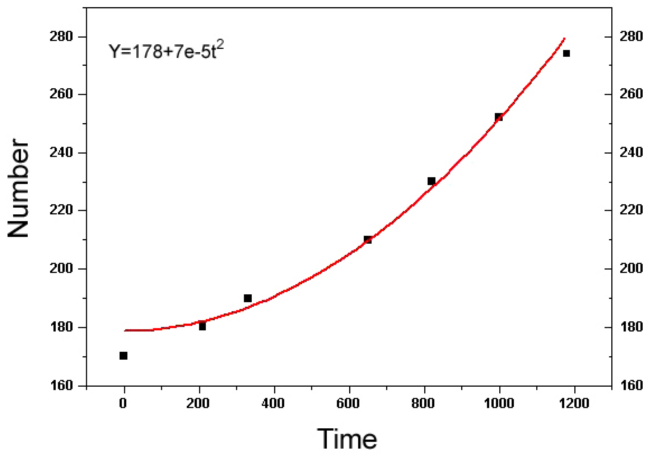

Figure 6*: Results of the model: the number of cities* N *versus the time* t. *Note that the data fits a quadratic equation.*

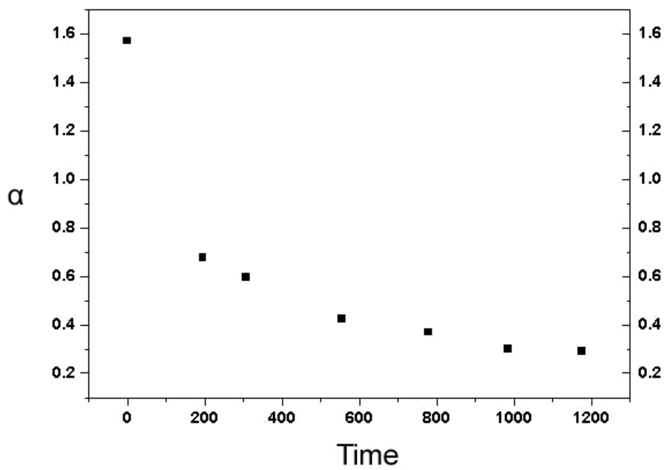

Figure 7*: Results of the model: the 'shape exponent'* α *versus the time* t

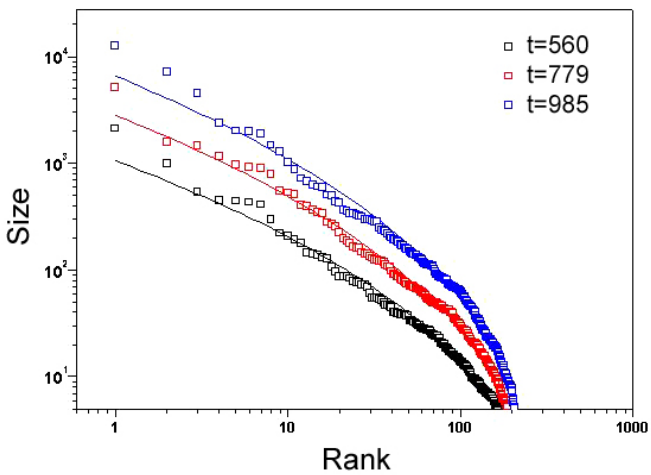

Figure 8*: Results of the model: the rank size distribution of the cities in the model on a log-log graph.*

7 MICRODYNAMICS: THE RANK-CLOCK ANALYSIS

So far, we have focused only on the macro dynamics of the Israeli system of cities. This means that we ignored the changes that appear within the positions (or ranks) of individual cities with time, and looked only at the rank size distribution of the entire system. In this section, we follow Batty [12, 13] and present an initial analysis of the micro dynamics of both the Israeli system of cities and the above simulation model.

We start with the real data of Israeli cities (and settlements) with populations larger than 5,000 from 1950 to 2005 [14]. The number of these settlements increased from 34 in 1950 to 172 in 2005. Fig. 9 presents the rank clock of Israeli cities for these years. The cities are colored according to their rank and the time they first entered the system with red representing cities which enter first through a spectrum of colour – red to yellow to green to blue – for the cities that enter last. The micro dynamics of this system of cities presents little irregularity, is mostly stable in structure and shows a system that is rapidly growing with cities rising rapidly up the ranks but few cities falling out. We can conclude all this in rather impressionistic terms by simply viewing the colour balance of the clock and comparing this to a system like the US where there is much greater volatility into and out-of of the top ranked cities [12]. There are only a few cases where cities move toward the center of the rank clock over this period while the cities that existed in the early stages of the development from 1950 remain the largest cities in the system (with cities entering earlier nearer the center of the rank clock). Few cities that were introduced in later years manage to increase significantly and move to the center of the clock. We can define a number of measures or parameters that characterize the clock, hence the system of cities. First, the rank shift is a parameter which indicates the stability of the dynamics of individual entities in the system. It is defined [12] as:

$$d_i(t) = \left| r_i(t) - r_i(t-1) \right| \quad , \qquad (10)$$

where $r_i(t)$ represents the rank of city i at time t, and $r_i(t-1)$ represents the rank of city i at time $t-1$. Obviously, this expression is valid only if the examined city is in the system at both times t and $t-1$. The average shift for the entire system is:

$$d(t) = \frac{\sum_i d_i(t)}{N} \quad , \qquad (11)$$

where N is the number of entities in the system. The average shift for the entire period studied is defined as:

$$d = \frac{\sum_t d(t)}{T} \quad , \qquad (12)$$

where the sum is the average shift of the system at different times and T is the entire period.

(a) (b)

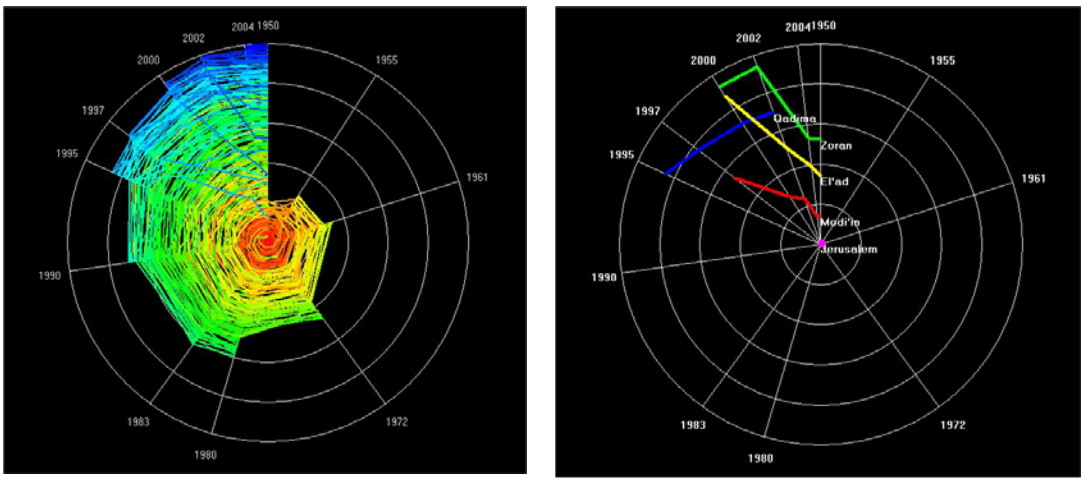

Figure 9*: The rank clock representation of the Israeli system of cities (a) between the years 1950 and 2005 and (b) with some specific cities also plotted.*

In the Israeli system of cities, d was found to be 5.4 which resembles the values of the same parameter for the USA and the UK [12]. This means that on average, each city in the system changes its location in the rank list by 5.4 places during the studied period. When analyzing these changes in rank, we can see that most cities in Israel presented very few changes in their ranks (see Fig. 9), while a small number of cities changed their ranks considerably. These cities can be divided into 3 groups, based on the reasons for their growth; the first group consists of orthodox-religious settlements, characterized by high annual growth rates (approximately 10%), which can be explained by the high volume of birth in the orthodox community. El'ad is an example for such a settlement. It was introduced to the system only in the year 2000 and within the next 4 years changed its rank from 144 to 64 ($d_i(t) = 20$).

The second group includes several settlements that were united (by government decision) into one municipality. Following this unification, some of these settlements disappeared from the system, while others increased systematically and moved towards the center of the rank clock showing rapid change reflected in the steepness of their slope. Zoran which was united with Qadima in the year 2003, changed its rank from 152 to 87 within one year ($d_i(t) = 65$). Lastly, the third group consists of the city of Modi'in, the only city in Israel which was completely planned (in terms of its located population) to be a large city between Tel Aviv and Jerusalem. Modi'in's rank changed from 88 to 23 between the years 1997 and 2005($d_i(t) = 9.3$).

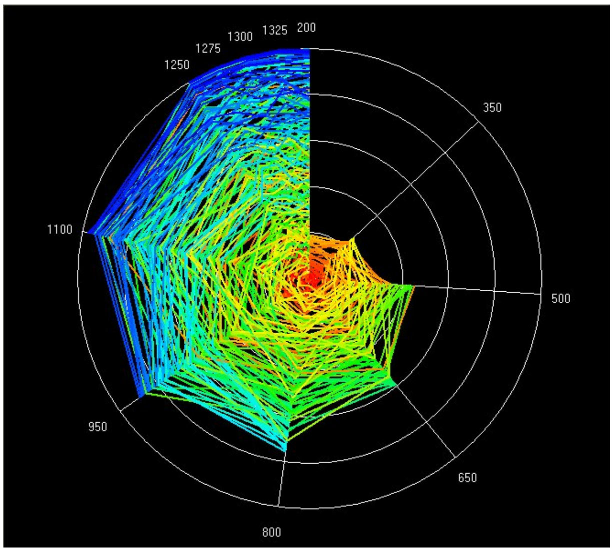

***Figure 10**: The rank clock representation of the model results at equivalent time points to the real evolution in Figure 9.*

When analyzing the results of the model for the Israeli system (see section 5), one finds $d = 14.8$ which is considerably larger that the value of d for the real data describing the Israeli system of cities. Fig. 10 presents the rank clock for this model. and based on this, the findings suggest that the model does not provide a good description for the micro dynamics of the system. Even before calculating the value of d , we can see that the rank clock is different from the one of Israel's real data as the colors in the clock are mixed and present no organized pattern. However, the calculated value of $d = 14.8$ is similar to the case of the top city populations in the world data set from classical times to the modern day [12]. This value is relatively high and indicates that the individual entities of the system presented great changes in their sizes and ranks over time.

The importance of this comparison is that it shows that even though the model provides a very good description for the macro dynamics of the Israeli system of cities, it does not explain their micro dynamics. It appears that further work needs to be done in order to develop a model that will provide a good description for both the macro and micro dynamics of a system and this would probably have to include many more specific spatial factors which characterize the urban development of Israel during these years. The reason for that lies in the fact that the micro dynamics of the Israeli system of cities is affected by various factors (such as government regularities, as described earlier) that cannot be imitated in the current model. The model, which is a complex system, is based on many random variables that are all dependent on one another. In its current version, it is very difficult to depict the exact variable that controls each aspect of the changes in the micro level of the system. Thus further work is needed in order to calibrate the model to fit the micro dynamics of this system of cities.

8 CONCLUSIONS

We have presented an adaptation of a simulation model for the growth of entities in the Israeli system of cities so that we might examine the dynamics of the distribution through time. Our approach is different from most other approaches to city size distributions in particular and size distributions in general in that our focus is upon the micro dynamics as well as the macro statics of cross sectional city size distributions. We also applied our multiplicative growth model to the cities of Israel and were partially successful. When considering the ensemble of cities at the macro scale, i.e. their rank size distributions, we get a convincing explanation of the time variation of these distributions which is dependent upon the rate of creating new cities. However, when studying the micro dynamics of the system, i.e. the evolution of individual cities over time, using the rank-clock representation, it is clear that the relative variation in size and rank of the cities (average shift) is considerably larger in the model than in the real data. Moreover, the general pattern of this dynamics reveals considerably more irregularity than the real system. Hence, we believe there is an inconsistency between the macro and micro aspects of the analysis.

We believe the model presents good results for the description of the macro dynamics of the system but fails to describe its micro dynamics, and thus it needs to be extended. One option, suggested by Havlin [16], is to consider the growth of cities with interactions or correlations among themselves. Such extensions would take the model to one dealing with systems of cities and their interactions which have meaning in terms of trade and other transportation flows. In the current model the growth of each city follows equation (2) alone, i.e. each city grows independently of every other. The proposal which we will follow in future research is to introduce enough but not too rich a set of interactions between cities such that the growth of any one city will be dependent on the growth of others.

Finally, we wish to emphasize that even the known "static models" present some evolution until they obtain the desired distribution. Until recently [12, 13], the evolution of the individual entities was hardly investigated at all but it seems necessary to do so in order to understand the evolution of systems of entities as a whole. In other words, we believe that dynamics has to be introduced in all models that study the distribution of sizes.

9 AKNOWLEDGMENTS

The authors thank George Kun from the Israel Central Bureau of Statistics for providing the data concerning the cities of Israel and for useful discussions.

10 REFERENCES

1. R. L. Axtell, Science **293**, 1818-1820 (2001).
2. S. V. Buldyrev et al., Physica A **330**, 653-659 (2003).

3. R. Carvalho and A. Penn, Physica A **332**, 539-547 (2004).
4. G. Duranton, Regional Science and Urban Economics **36**(4), 542-563 (2006).
5. K. A. Chattopadhyay and S.K. Mallick, Physica A **377**(1), 241-252 (2007).
6. J. Laherrere and D. Sornette, Euro.Phys. Jour. B **2**, 525-539 (1998).
7. A. Blank and S. Solomon, Physica A **287**, 279-288 (2000).
8. E. Limpert, W. A. Styahel, and M. Abbt, BioScience 51, 341-352 (2001).
9. L. Benguigui and E. Blumenfeld-Lieberthal, Int. Journal of Modern Physics C, **17**(10), 1429-1436. L. Benguigui, E. Blumenfeld-Lieberthal, Computers Environment and Urban Systems, in press http://dx.doi.org/10.1016/j.compenvurbsys.2006.11.002
10. L. Benguigui and E. Blumenfeld-Lieberthal, Physica A **384**(2), 613-627, (2007)
11. D. Sornette, *Critical Phenomena in Natural Sciences,* Springer, Berlin (2000)
12. M. Batty, Nature, **444,** 592 – 596, (2006)
13. M. Batty, CASA working papers 112, (2007) http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=112
14. G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge, MA (1949)
15. Israel Bureau of Statistics www.cbs.gov.il
16. S. Havlin, Private communication.