PREPRINT

Diffusion of broadband services: an empirical study

Ryszard Strużak
National Institute of Telecommunications, Wrocław, Poland
Email: r.struzak@ieee.org

ABSTRACT

The aim of this article is to broaden of our understanding of broadband telecommunication services diffusion process. Generic models of the process are created to describe the past and probable future of the process and specifically its limitations. The models can be set up for particular regions, user groups, or technologies to identify the growth momentum and intrinsic limits. They can serve as useful tools in planning, managing and monitoring of broadband services deployment and can be helpful in policy considerations.

Keywords: broadband, Internet, growth limit, logistic model, trend, access, information society,

INTRODUCTION

Broadband services are crucial for economic, social and human development – without them the future Global Information Society cannot be build [1]. With this in mind, numerous development programs have been launched around the globe to assure universal broadband access. As a result, the average broadband penetration rate has increased in all countries. However, it is sill far from being universal and significant broadband "gaps" exist, in spite of all the efforts made till now.

How does the present inequality affect the possible reduction of the divide in future? In some cases one observes the gaps o be increasing rather than decreasing: Could they ever close, or not? Or, in other words: Could the Global Information Society be ever possible?

Broadband services are relatively new. Their complexity is poorly understood. Their underlying business models have been in a state of formation. Various uncertainties remain to be settled.

The OECD Council has recommended recently "*Governments should focus their attention on improving metrics and analysis to better understand new usage trends, their impacts on the economy and society as well as policy*." [2].

This study aims at a better understanding of the broadband diffusion process. It bases on the real-world data, in contrast to computer simulations that deal with arbitrary data and "virtual reality". That brings an empirical component to the ongoing debate on the universal access, trends, markets, gaps, etc.

The paper discusses a generic model, which can easily be transformed into the specific model of the broadband penetration growth in a given country and/or user group. It explains how to build the model and how to estimate its parameters using published statistics. "Model" means here a formula that approximates the data and allows predicting their developments. It describes the dynamics of the diffusion process and its probable future using the historical data as a "window" to peer into its mechanism. The paper describes also how the models and empirical data were compared and what degree of agreement has been noted.

The models make it possible to answer important questions. Is the broadband penetration level X ever possible in country Y, or its infrastructure must be changed? If so, when could it be expected? Can the broadband gap between two given countries be ever closed, and if so, when? What potential market is to be expected in a given country and year? How to allocate rationally the available resources to reduce the broadband gaps among various countries or user groups? Note that none of these questions (and similar ones) could be answered without models.

Table 1. Basic indices of countries under study

Country	Broadband penetration rate, (2009), %	GDP (PPP), K$ per capita (20??)	Population, million (20??)	Population density, persons/km2	Urban population As % of total	Human development index (200??)	Gini coefficient (200??)
CH	35.64	41.4	7.6	188	73	0.960	34
DK	37.05	36.0	5.5	127.9	87	0.955	29
IS	32.82	39.6	0.3	3.1	92	0.969	28
KR	33.53	28.1	48.6	491	81	0.937	31
NL	37.10	39.5	16.8	400.2	82	0.964	31
NO	33.91	57.4	4.7	12.5	77	0.971	25
SE	32.36	36.6	9.1	20.6	85	0.963	23

The next section discusses the significance of the intrinsic limitations of the diffusion process. Then, statistical data from selected countries are analyzed and modeled. Logistic mapping, known from mathematical biology, fractals, and chaos theory, is used as the simplest approach. As the broadband inequality problem depends on fractals and chaos theory has been used for that purpose as the simplest approach. The focus is on how to identify the intrinsic limits of the process, and various methods are discussed and compared.

LIMITS TO GROWTH

The growth of broadband penetration has a limit, which is determined by the socio-ecological environment. The environment includes a set of world-views, resources, institutions, knowledge, technology, competition, price, contents offered, etc. The limit is unknown in advance, but has an immense impact on the development. For that reason, general discussions on the growth limits have started as early as in the 19-th century (Thomas Malthus, François Verhulst) and continue until now.

The recent report of the authors of Club of Rome notes: *"Once the limits to growth were far in the future. Now they are widely in evidence"* and continues: *"one vivid example of global overshoot and collapse did actually take place around the turn of millennium: the 'dot-com" bubble in the stock market*" [3]. Despite of these discussions, the broadband diffusion limits have attracted little attention, even in the special report published recently by the World Economic Forum [4]. Apparently, the issue has been ignored.

Goals of development program have often been unrealistic, and neglecting natural limits might be one of reasons why so many programs failed to reach their targets.

One might believe that the problem of broadband gap is short-lived, but it is not. One might (falsely) expect that emerging markets will reach the same penetration level as advanced markets, because the latter saturate earlier and emerging countries seem to grow faster; all of them will reach the same penetration level. However, there is no such thing as a single limit to growth, common to all. For each user group there is a specific saturation level that may differ from one group to another. As a consequence, there can be irreducible gaps that never close in stationary conditions [5]. The issue is discussed below.

EMPIRICAL DATA

Empirical data studied here have been collected in seven countries during the decade 2000 to 2010. These countries are Denmark (DK), Iceland (IS), Korea (KR), the Netherlands (NL), Norway (NO), Sweden (SE), and Switzerland (CH). They have been selected because they are the broadband world-leaders, according to OECD. Table 1 gives basic indices of these countries.

Analyzed is the number of broadband subscribers per 100 inhabitants in each country, known also as the penetration rate, or diffusion rate. (An alternative measures in use are the proportion of households or enterprises with the broadband access.) The broadband services are meant here as those requiring the transmission speed of 254 kb/s or more to the customer, in accordance with the current international practice. Figure 1 shows how the penetration rates in these countries were changing with time. It increased from some 4% in 2001 up to 38% in 2009.

INCREMENTAL DISCRETE-TIME MODELS

This section discusses the first-order difference model of the diffusion process, and offers a simple way of derivation of its limit. The model describes changes of the penetration rate, or the (first-order) difference between two consecutive values of the penetration rate: $x_i = (p_{(i+1)} - p_i)$ as Figure 2 illustrates it on the plane (p, x). Only two countries are shown in the figure; the plots of remaining European countries are similar to that of the Netherlands. Note the clustering of points; if the process were random, the plot would have a shotgun pattern. Note also that no simple straight line could ever approximate the relationship between x and p for the countries shown.

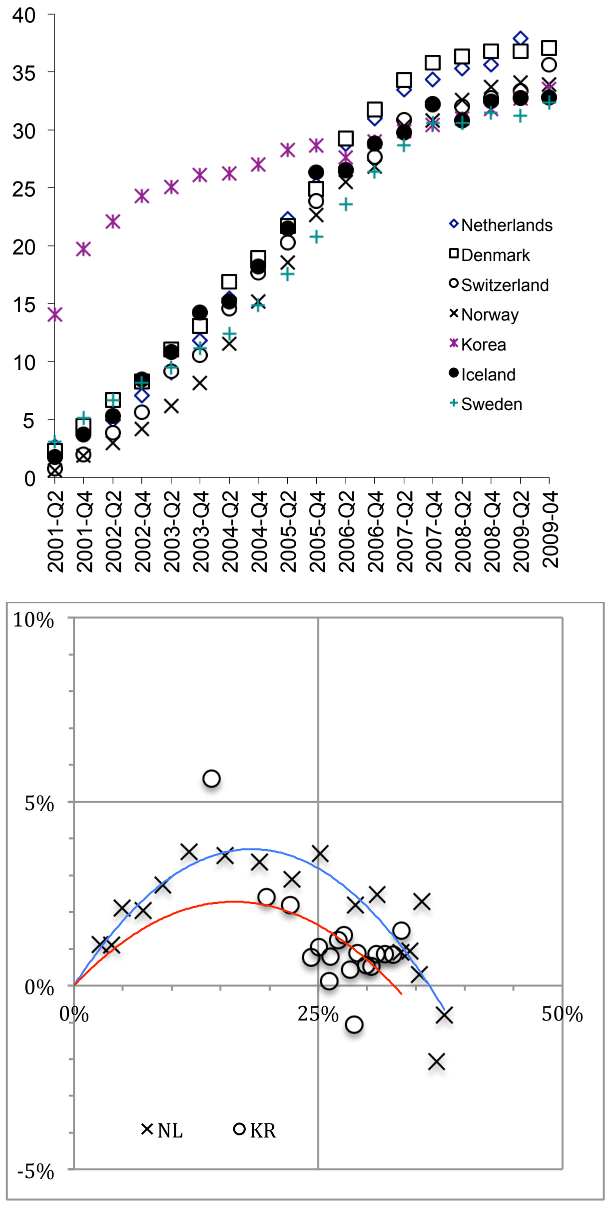

Figure 1. Number of broadband users per 100 people in seven leading countries according to OECD in the years 2001 – 2009 (left) and its increments in the Netherlands and Korea (right). The horizontal axis is the penetration rate (p); the vertical axis is its increment (x) in percentage points.

The simplest polynomial function that can do that is parabola:

$$x = \mathbf{a}p\,(\mathbf{b} - p). \qquad (1)$$

Parameters **a** and **b** decide on its form and position. The parabola was first used more than hundred years ago by François Verhulst to model the growth of human population.

When p approaches zero, the slope of the parabola tends to **c**=**ab**, and when p=**b**, the process reaches its limit.

Indeed p_{max}=**b**, because the increment x reaches zero at that point and the penetration rate does not increase further. The limit can be determined by inspection, by extrapolating the parabola in Figure 2 up to its crossing point with the p axis. It is close to 38% for the Netherlands and about 34% for Korea.

Alternatively, the limit can be calculated. For that purpose, the increments (x) of the original data (p) are to be calculated, and then the regression parabola $x(p)$ and its roots $x(p)$=0. The results of such calculations are shown in Table 2.

A second-order difference model can be constructed in a similar way. If the penetration rate (p) can be compared to the position of a body, the first-order difference (x) to its velocity, and the second-order difference (y) to its acceleration: $y = \mathbf{a}(\mathbf{b} - 2p)$. It is the slope of the parabola. At the beginning (p=0), the slope equals **c**=**ab**, and it zeroes at $p = \mathbf{b}/2$. Theoretically, the second-order difference model can be used to calculate parameters **a** and **b** in the same way as in he case of the first-order difference described above. However, practical significance of that method is limited due to high sensitivity to data errors.

MATCHING THE DATA

The generic model becomes specific one when its parameters get their numerical values. As mentioned, the parameters can be determined from observations. Given the generic function and set of data points, one select the function parameters in such a way that the sum of squared distances between the function and data points is minimal; it is the Least Squares method. In the case of parabola, two data points substituted to equation (1) create a system of two equations with two unknown that can be solved for **a** and **b**, which assures that the data points lie on the parabola. However, empirical data rarely are error-free and so is the solution. Increasing the number of data points and the Least Squares method reduces errors

The model – observations differences may be random, or systematic. While random ones are due to observation errors, systematic deviations may be due to wrong model, wrong observations, or wrong distribution of data points, or due to all these reasons jointly. For instance, the Korean data points in Figure 2 are inconveniently located. They are grouped in one place instead of being distributed more uniformly along the p-axis. The figure shows also that the current model does not explain exactly the real-life process: it does not allow negative increments (i.e. decrease of penetration rate) observed in reality. This may be the effect of the financial crisis that has pushed recently the global economy into its worst recession since the 1930s, but the model does not take it into account.

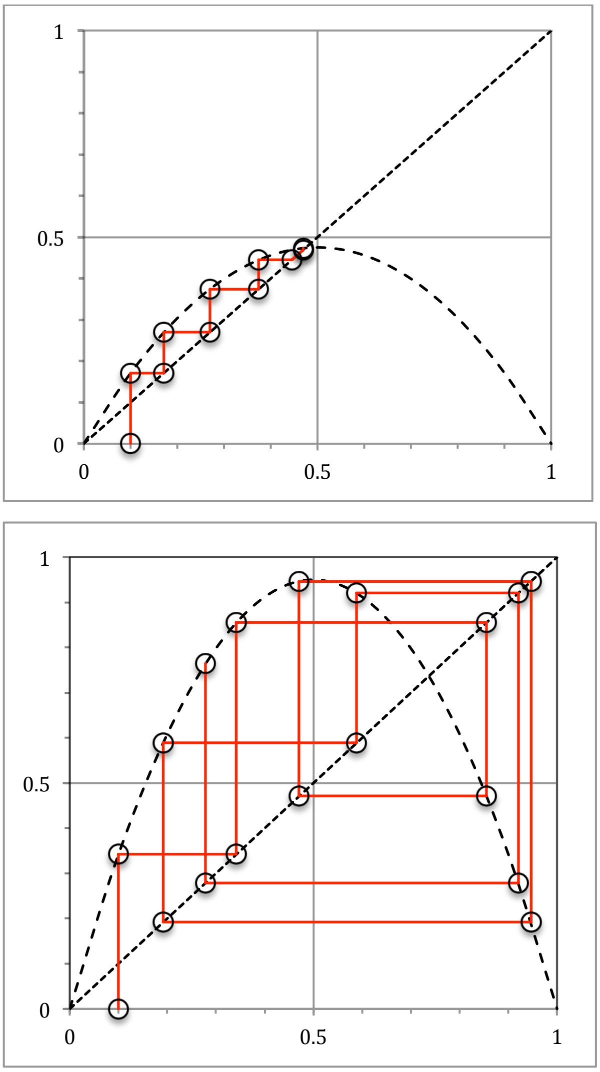

Figure 2. Examples of graphical integration of recurrence equation (2) for various values of growth parameter r: 1.9 (top), 3.8 (bottom)

A standard statistical measure of the model – observations match is the coefficient of determination, or R-squared. Its numerical value (contained between zero and one) says how adequately the model explains the data variations. If $R^2 = 1$, the model perfectly matches the data and explains all their variability. In such a case, every data point lies exactly on the diagram. However, when some observations are corrupted, they are at a distance from the line and R^2 is lower than one in such cases.

If e.g. R^2=0.9, the model explains approximately 90% of the variations observed, and with $R^2 = 0$, there is no statistical relationship between the model and observations. The determination coefficient and the method of least squares have been applied throughout the study, using the Microsoft Excel and its Solver and other standard tools.

DISCRETE-TIME ITERATED MODELS

This section introduces iterated (discrete-time) models of the process. Earlier we noted that the "next" value $p_{(i+1)}$ of the process is the sum of its "previous" value p_i and increment x_i as described by parabola $p=p+\mathbf{a}p(\mathbf{b}-p)$. The parabola equation can be simplified by substituting $p=z(1+\mathbf{ab})/\mathbf{a}$ and $\mathbf{r}=1+\mathbf{ab}$:

$$z_{(i+1)} = \mathbf{r}z_i(1 - z_i) \qquad (2)$$

Here, variable z is the normalized penetration rate. This relation is known in Chaos Theory as quadratic iterator or logistic map. In biological applications, **r** is interpreted as combined rate of reproduction and starvation.

To determine the process over a longer period, equation (2) can be solved through graphical iteration, as shown in Figure 2 [6]. The figure is a unit square with the horizontal axis z, and the vertical axis v, an auxiliary variable. Note that z=1 corresponds to $p=\mathbf{r}/\mathbf{a}$. The parabola shown there is the graph of function (2), with maximum of $\mathbf{r}/4$ at z=1/2. The square's diagonal is an auxiliary line and the stepwise line connects the results of iterations.

The first iteration starts from an initial value with a vertical line that goes up to the parabola. Its length is the function increment. Then, the line changes direction and continues horizontally up to the diagonal $v = z$. Here, the subsequent iterations begin: a new vertical line goes to the parabola, changes the direction, etc. etc. In Figure 2, the process converges for r=1.9 at the point-attractor at the intersection of the parabola and the diagonal, no matter where it starts. One can verify that at that point $p_a=\mathbf{b}$. Such a converging case coincides with the continuous model discussed in the previous paper [5] and in the following section.

The smooth, monotonic convergence to the point attractor is only one of possible growth-paths, which is observed for **r** between 1 and 2 [6]. For **r** between 2 and 3, the process converges too, but with transient oscillations that disappear after some time. For **r** between 3 and about 3.45, the process stops to settle down and the oscillations are permanent. The oscillations can have a complicated form, with more than two alternating values. With parameter **r** increasing further, the periodicity gives way to chaos – to fluctuations that do not settle down at all, see Figure 2. Chaotic variations appear first at **r** greater than about 3.57 and then the chaotic regions are interleaved with those of regular behavior in a complex arrangement [6]. In this study, however, neither chaos, nor periodicity, was noticed in the observations.

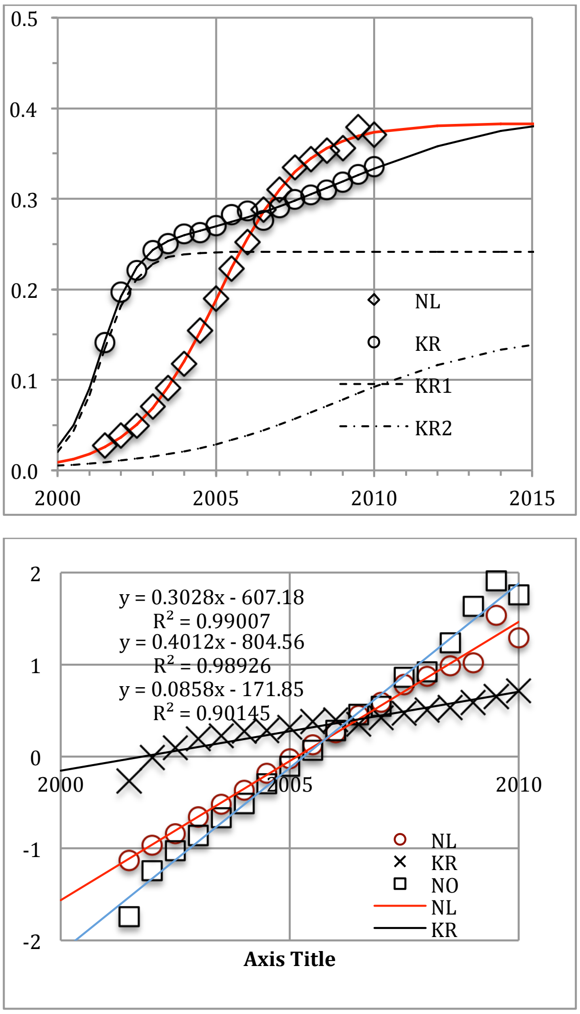

Fig 3. Continuous-time models. Top: Time history for the Netherlands (simple model) and Korea (two-component model). Middle: Regression lines for transformed data w=Ln{[**b**/p(t)]-1} for the Netherlands, Norway and Korea.

Continuous-time models

The continuous-time model is closely related to the iteration model for **r** < 2, as mentioned above. When the increments (x) of penetration rate (p) become infinitesimal (dp), difference equation (2) becomes an ordinary, first-order differential equation dp = **a**p(**b** – p)dt in the time domain. Its solution is the well-known logistic growth function:

$$p(t) = \mathbf{b}/\{1 + \exp\,[\mathbf{c}(\mathbf{d} - t)]\} \qquad (3)$$

This function, known from the previous sections and article [5], involves three parameters. Parameter **b** represents the limit of the process, and **c = ab**, is the steepness of the growth process. The third parameter, **d**, describes the time behavior: it is a specific time, at which the penetration rate reaches the half of its maximal value. According to the model, the growth starts exponentially with time and later slows down. Figure 3 shows an example. The time scale is extended up to the year 2015, under the assumption of the constant parameters. It means that the existing socio-economic environment will not change in that time.

The parameters of the models are shown in Table 2.

Model parameters can also be evaluated using linearization of the right side of formula (3) by substitution: w=Ln[(**b**/p)-1]. This transforms (3) into a straight-line w=**c**(**d**-t), as shown in Figure 3. The figure shows data from the Netherlands, Norway and Korea. A reasonable agreement with the model is noticeable in the case of the Netherlands. The departures from straight line in the case of Korea and Norway indicate that the simple logistic model does not reflect exactly all the complexity of the process. However, practical significance of that method is limited. The logarithmic function requires its argument to be greater than zero, which introduces complications in calculations of **b** using the least square method.

Combined models

This section bases on observation that a linear combination (sum and/or difference) of two or more logistic function can match the data better than the single function does. An example is shown in Figure 5, which shows the two-component model of the broadband diffusion in Korea. Before 2002, the process was developing in that country following the model component KR1. KR1 has the limit of some 24 broadband subscribers per 100 inhabitants and saturates about the year 2005. Then the process develops further following the sum of two components, KR1 plus KR2. The second component adds some fifteen extra subscribers and saturates after the year 2020. The replacement of the simple model by the combined one has increased the coefficient of determination from R^2=0.376 to R^2=0.993.

Discussion

The parameters of the models determined using different methods are shown in Table 2. The discrete-time difference model is simplest, but its coefficient of determination is the lowest: from 0.306 in Switzerland to 0.804 in the Netherlands. The two-component model offers the best approximation, but requires more data points. The simple continuous-time logistic model is somewhere in between.

Figure 4 compares continuous-time model predictions with the data. A good agreement between the empirical data and model predictions is seen in the figure. The coefficient of determination and coefficient of proportionality are both close to the ideal value of 1.

The models uncover underlying structure of the process that may be of interest for those desiring to gain insight into the engineering, scientific, business and other aspects of the processes behind the data.

Table 2. Model parameters

	Incremental model		Simple continuous-time logistic model			Double continuous-time logistic model					
	a	**b**	**b**	**c**	**d**	**b1**	**c1**	**d1**	**b2**	**c2**	**d2**
CH	0.392	0.347	0.347	0.795	2005.0	0.261	0.858	2005.7	0.087	1.478	2003.3
DK	0.362	0.380	0.391	0.688	2005.0	0.391	0.688	2005.0	0.121	1.089	2019.5
IS	0.406	0.325	0.333	0.766	2004.6	0.334	0.760	2004.6	0.121	1.089	2048.7
KR	0.329	0.314	0.320	0.489	2001.2	0.234	1.807	2001.4	0.178	0.321	2009.3
NL	0.408	0.363	0.382	0.744	2005.0	0.281	0.691	2004.9	0.101	0.937	2005.3
NO	0.440	0.339	0.347	0.847	2005.3	0.247	0.784	2005.7	0.104	1.092	2004.7
SW	0.312	0.342	0.349	0.593	2005.3	0.235	1.093	2006.0	0.088	1.538	2001.9

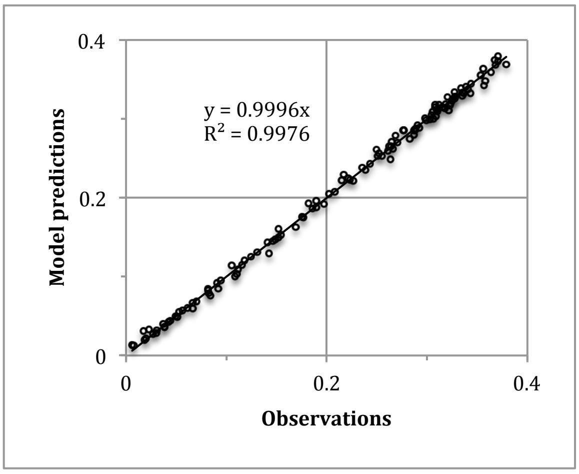

Figure 4. Empirical data points and model predictions show a good agreement (all countries, 126 points).

For instance, it can be seen in Figure 3 that the diffusion process in Korea has changed, adapting to alternations in the socio-economic and technical environment that took place in that country. However, the data studied have not included any clues that would allow identifying what specifically influenced the process. Was it technological progress, or something else?

Many changes took place around the world during the decade studied. The transmission bandwidth has increased. New technologies, new services, and new business models have been successfully introduced, new regulations, etc. However, when looking at the models (Figure 3), we see that in European countries the broadband penetration process was developing following a steady pattern that did not change during the whole decade: all the changes in technology, regulatory infrastructure etc. had unnoticeable impact on the process of broadband services development. From the broadband penetration rate viewpoint they were insignificant.

The models presented here make it possible to answer various important questions that could not be answered otherwise; examples of such questions are given in Introduction. If the model parameters are known, answering them is elementary. In this section we discuss only two of them.

The first one is: How the present inequalities impact the possible closing the broadband diffusion gap in future? The answer is simple: the present situation does not count; what counts is the difference between the growth limits. In other words: if two countries have the limits equal one to another, the gap between them will close; the models can indicate when it is to be expected. If, however, the limits differ one from the other, the gap is irreducible and never will close in the existing environment. The models also say that if the penetration rates in two countries are identical at some specific moment of time, they may differ significantly in a later time.

The second question relates to allocation of resources to reduce the broadband gaps among various countries or user groups: How to allocate them rationally? An answer has been proposed in the previous paper [5]. It is based on an identification of bottlenecks where the resources are needed most. For that purpose evaluated are and compared are the inherent limits, potential markets, and efforts needed to reach the common goal.

CONCLUDING REMARKS

The article has introduced simple models of broadband diffusion process. They match quite well the empirical data collected in world-leading countries. A few alternative ways have been presented on how to identify the model parameters; they all give similar results. The data indicate that the broadband penetration process in each country studied, except Korea, was following a steady model constant during the whole decade; changes in technology and in socio-economic environment that took place in the meantime did not alter it.

This article complements the previous one [5] with new elements: (1) new data; (2) new models; (3) new methods of limit determination; (4) new model-data comparison; (5) new conclusions about irregular (chaotic) behavior; (6) new conclusion about the impact of changes that took place in the past decade in the European countries under study.

No unpredictable behavior of the data has been noticed, but the study has shown that the diffusion process can be periodical or chaotic in specific conditions. It is due to super sensitivity of the quadratic recurrence relation to perturbations; even minuscule perturbations can lead to huge effects. Edward Lorenz, a pioneer of chaos theory, first noted that phenomenon in his weather prediction studies in 1960s.

For many years, most of science grew up on the tacit assumption that the world is ordered, and chaos is something unusual. Edward Lorenz and his followers are convinced that it is the opposite: it is rather the deterministic order that is exceptional. Whether further data on broadband diffusion will genuinely be chaotic, remains an open question.

The models can be used in studying, planning and monitoring of the diffusion of broadband services, in trend predictions and in policy considerations. They can also serve as an empirical reference in simulation studies. The approach presented here is generic enough to be applicable not only to various countries but also to various regions, services, and user groups within a country.

ACKNOWLEDGMENTS

This study was possible thanks to the grant by the National Institute of Telecommunications (NIT), Poland. The work was done partly at the Marconi Laboratory, International Centre for Theoretical Physics (Trieste, Italy). The author would like to take this opportunity to thank Marcin Strużak for the assistance offered.

REFERENCES

[1] Touré H.: Broadband Challenges; Journal of Telecommunications and Information Technology (Special Edition, PP-ITU, Guadalajara), October, 2010; p. 3-7.

[2] Broadband Growth and Policies in OECD Countries; OECD Publications (2008); ISBN 978-92-64-04668-9-No. 56221

[3] Meadows D.H., Rangers J, Meadows D. (2004): Limits to Growth – The 30-year Update; Chelsea; ISBN 1-931498-51-2

[4] Dutta S., Mia I. (Eds.): The Global Information Technology Report 2010–2011; World Economic Forum, 2011, ISBN-10: 92-95044-95-9

[5] Struzak R.: Broadband Internet in EU Countries: Limits to Growth; IEEE Communications Magazine Vol. 48, No. 4, April 2010, pp.52-57

[6] Peitgen H-O., Jurgens H., Saupe D.: Chaos and Fractals: New Frontiers of Science. Springer Verlag. 2004, ISBN 0-387-20229-3

BIOGRAPHY

RYSZARD STRUZAK [LF] (r.struzak-at-ieee.org) is a full professor at the National Institute of Telecommunications and co-director of the ICTP School Series on Networking. He is the author/co-author of some 200 publications and 10 patents. He is the former acting assistant director of ITU-CCIR and head of its Technical Department, Editor-in-Chief and Editorial Board Chair of *Global Communications*, and a consultant to ITU, UN-OCHA, UNESCO, WB, IUCAF, and other entities. He is co-founder and former Chair of the International Wroclaw Symposium on EMC (EMC Europe after 2010). He was elected to leading positions in ITU-RRB, ITU-CCIR, URSI, and CISPR. He is the recipient of the ITU Silver Medal, two International Symposia awards, IEEE EMCS Special Symposium Recognition Award, and numerous other awards. He was elected member of the New York Academy of Science and an Academician of the International Telecommunication Academy.