

*fuzzy c-regression models,  
bacterial growth modelling,  
piecewise continuous functions*

Aleksander OWCZAREK\* Michał MOMOT\*,  
Ewelina SOBIERSKA\*\*

## FUZZY C-REGRESSION MODELS IN BACTERIAL POPULATION DYNAMICS MODELLING

The paper describes application of the fuzzy c-regression models algorithm in bacterial population dynamics modelling. Such a well-known algorithm, which provides simultaneous estimates of the parameters of c-regression models, together with a fuzzy partitioning of data, was used to determine bacterial growth curve given by the piecewise continuous function. The real as well as artificial data were used and promising results were obtained even for data with significant errors resulting from the measurement inaccuracy.

### 1. INTRODUCTION

The aim of cluster analysis is to divide a given set of  $N$  observations (data vectors)  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$  into  $c$  clusters (classes) [1]. This partition should have the following properties:

- *homogeneity* within cluster – data belonging to the same cluster should be as similar as possible,
- *heterogeneity* between clusters – data belonging to different clusters should be as different as possible.

The formed clusters are pairwise disjoint, nonempty and reproduce the original data set via union [2]. The result of clustering is usually presented as partition matrix  $U$ , which dimension is  $c \times N$ . In real applications there is very seldom sharp boundary between clusters so that fuzzy clustering is frequently better suited for the data.

The set of all possible fuzzy partitions of  $N$  data vectors into  $c$  nonempty subsets (classes) is defined as follow[3]:

$$P_{fc} = \left\{ \mathbf{U} \in \mathbf{R}^{c \times N} \left| \begin{array}{l} \forall_{\substack{1 \leq i \leq c \\ 1 \leq k \leq N}} u_{ik} \in [0,1], \\ \sum_{i=1}^c u_{ik} = 1, \\ 0 < \sum_{k=1}^N u_{ik} < N \end{array} \right. \right\} \quad (1)$$

Part of clustering methods is formulated as a minimization process of the suitable cost function [2,4]. The fuzzy c-regression models (FCRM) clustering algorithm provides simultaneous

\* Institute of Medical Technology and Equipment, 118 Roosevelta St., 41-800 Zabrze, Poland,  
tel./fax: +48 32 2716013, e-mail: aleck@itam.zabrze.pl

\*\* Institut für Kernenergetik und Energiesysteme, Universität Stuttgart, Pfaffenwaldring 31, 70569 Stuttgart

estimates of the parameters of c-regression models, together with a fuzzy clustering of data. In this method the cluster prototypes are functions instead of geometrical objects [5, 6].

Piecewise continuous functions have many applications in science, especially in microbiology where the bacterial growth curve is given by a nonlinear piecewise function containing four distinct regimes<sup>1</sup> (of which three are modelled).

In our work we used artificial as well as real data, among others *Escherichia coli*. This bacteria is an emerging cause of food-borne illness. Infection often leads to severe bloody diarrhea, abdominal cramp and occasionally to kidney failure (haemolytic uremic syndrome) or meningitis. Understanding the microbial population dynamics behaviour could help us for example to define a genetic diversity of the population and possibly to control the presence of pathogens [7].

The modelling of bacterial population dynamics based on FCRM clustering algorithm is described in this paper. The aim of our work is to determine function parameters of the bacterial growth curve in individual intervals as well as transition points<sup>2</sup>.

## 2. THE GROWTH CYCLE OF BACTERIAL POPULATION

When the bacteria are growing in a closed system the population of cells almost always exhibits growth dynamics presented on Figure 1. The growth curve can be divided into several distinct phases, such as [8]:

- *Lag phase* – the bacteria are becoming adjusted to their new environment,
- *Exponential phase* – the bacteria start dividing regularly, by the process of binary fission,
- *Stationary phase* – in a closed system, exponential growth cannot occur indefinitely as essential nutrients are depleted, or toxic products build up, so the growth is ceased,
- *Death phase* – the rate at which bacteria die is greater than the rate at which they divide - hence the number of viable cells decreases.

We ignore the lag phase (and the lag time) for our study on population dynamics. Thus, the last three phases of bacterial growth are characterized by the following piecewise continuous function:

$$F(t) = \begin{cases} F_0 \cdot e^{\lambda t}, & 0 \leq t \leq t_s, \\ F_s, & t_s < t \leq t_D, \\ F_s + M \cdot (t - t_D), & t_D < t, \end{cases} \quad (2)$$

where:

- $\lambda$  is the exponential growth rate constant,
- $F_0$  is the initial cell number in the population,
- $F_s$  is the cell number during stationary phase,
- $t_s$  is the time at which stationary phase starts,
- $t_D$  is the time at which death phase starts,

---

<sup>1</sup> A regime - a section of a piecewise function that follows a single hypothesis.

<sup>2</sup> A transition point of a piecewise function - the point at which it makes a transition from one form to another.

$M$  is the rate at which cells disappear during the death phase.

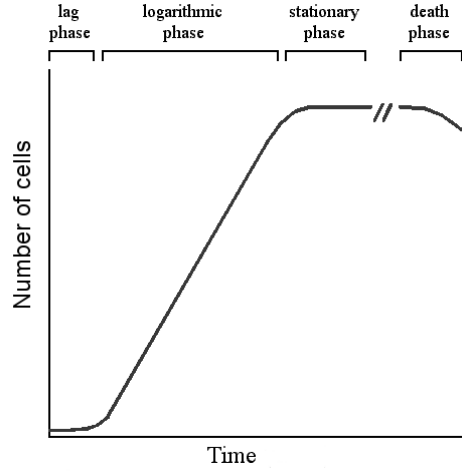


Fig.1 The typical bacterial growth curve.

### 3. FUZZY C-REGRESSION MODELS

The fuzzy c-regression models (FCRM) clustering algorithm yields simultaneous estimates of the parameters of c-regression models, together with a fuzzy c-partitioning of data [5]. The regression models take the following general form:

$$\forall_{1 \leq i \leq c} y_k = f_i(\mathbf{x}_k, \mathbf{w}_i) \quad (3)$$

where:

$\mathbf{x}_k = [x_{k1}, \dots, x_{kn}]$  – the k-th data sample  $k \in \{1, \dots, N\}$ ,

$f_i$  – the function parameterised by  $\mathbf{w}_i \in \mathfrak{R}^{q_i}$

$q_i$  – the number of parameters for the i-th model.

According to the minimizer property stated by Hathaway and Bezdek [5] we assume the following measure error function:

$$\forall_{1 \leq i \leq c} E_{ik}(\mathbf{w}) = (y_k - f_i(\mathbf{x}_k, \mathbf{w}_i))^2 \quad (4)$$

The family of fuzzy c-regression models objective functions is defined by:

$$E_m(\mathbf{U}, \{\mathbf{w}\}) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m \cdot E_{ik}(\mathbf{w}_i) \quad (5)$$

where  $m \in (1, \infty)$  denotes a weighting exponent which determines the level of fuzziness in the resulting clusters (usually  $m = 2$  is taken). The membership degree  $u_{ik}$  can be interpreted as a

weight representing extent to which the value predicted by the model  $f_i(\mathbf{x}_k, \mathbf{w}_i)$  matches  $y_k$ . The group coordinate minimization proposed by Hathaway and Bezdek [5] is one of the possible approach to the minimization of the objective function (4). It leads us to:

Fuzzy c–Regression Models Algorithm

**INITIALISATION**

- Given a set of data  $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ ,
- Set a weighting exponent  $m > 1$  and specify  $c$ ,
- Specify regression models (3) and the measure error function (4),
- Set a termination threshold  $\varepsilon > 0$ ,
- Initialise the partition matrix  $\mathbf{U}$  randomly.

**REPEAT** for  $r = 1, 2, 3, \dots$

**STEP 1**

Calculate values for the  $c$  model parameters  $\mathbf{w}_i$  that minimize the objective function  $E_m(\mathbf{U}, \{\mathbf{w}\})$ .

**STEP 2**

Update the partition matrix  $\mathbf{U}$  :

$$\forall_{1 \leq i \leq c} u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{E_{ik}}{E_{jk}}\right)^{\frac{1}{m-1}}} \text{ if } E_{ik} > 0 \tag{6a}$$

$$\text{otherwise: } u_{ik} = 0 \text{ if } E_{ik} > 0 \text{ and } u_{ik} \in [0,1] \text{ with } (u_{1k} + \dots + u_{ck}) = 1, \tag{6b}$$

until  $\|\mathbf{U}^{(r+1)} - \mathbf{U}^{(r)}\| > \varepsilon$ .

#### 4. THE BACTERIAL POPULATION DYNAMICS MODEL

Various attempts has been made to obtain parameters of the bacterial growth curve model. In our work we used the fuzzy c-regression model (FCRM) clustering algorithm to solve the problem. Including random initialisation of the partition matrix  $\mathbf{U}$  in a standard algorithm the regression models (3) take the following form:

$$\forall_{1 \leq i \leq 3} y_k = w_{i1} \cdot e^{w_{i2} \cdot x_k} + w_{i3} \cdot x_k + w_{i4} \tag{7}$$

for three classes (regimes), according to equation (2). In this case it is not possible to calculate the model parameters for the assumed measure error function (4).

However, we can simplify our problem by taking an advantage of the continuity of the considered model and initialising the partition matrix  $\mathbf{U}$  as follow:

$$\mathbf{U} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix} \quad (8)$$

$$u_{i,1} \quad u_{i,M} \quad u_{i,M+1} \quad u_{i,2M} \quad u_{i,2M+1} \quad u_{i,N}$$

where:  $M = \left\lfloor \frac{N}{3} \right\rfloor$ .

Such approach leads us to the following regression models:

$$\begin{aligned} y_k &= w_{11} \cdot e^{w_{12} \cdot x_k}, \\ y_k &= w_{21}, \\ y_k &= w_{31} \cdot x_k + w_{32}, \end{aligned} \quad (9)$$

and now we can readily calculate values of the model parameters  $w_i$  proposed above which minimize the objective function  $E_m(\mathbf{U}, \{\mathbf{w}\})$ .

The second model can be calculated as follow:

$$w_{21} = \frac{\sum_{k=1}^N (u_{2k})^2 \cdot y_k}{\sum_{k=1}^N (u_{2k})^2}. \quad (10)$$

For the third model we have:

$$w_{31} = \frac{\sum_{k=1}^N (u_{3k})^2 \cdot \sum_{k=1}^N (u_{3k})^2 \cdot x_k \cdot y_k - \sum_{k=1}^N (u_{3k})^2 \cdot x_k \cdot \sum_{k=1}^N (u_{3k})^2 \cdot y_k}{\sum_{k=1}^N (u_{3k})^2 \cdot \sum_{k=1}^N (u_{3k})^2 \cdot x_k^2 - [\sum_{k=1}^N (u_{3k})^2 \cdot x_k]^2} \quad (11a)$$

$$w_{32} = \frac{\sum_{k=1}^N (u_{3k})^2 \cdot x_k^2 \cdot \sum_{k=1}^N (u_{3k})^2 \cdot y_k - \sum_{k=1}^N (u_{3k})^2 \cdot x_k \cdot \sum_{k=1}^N (u_{3k})^2 \cdot x_k \cdot y_k}{\sum_{k=1}^N (u_{3k})^2 \cdot \sum_{k=1}^N (u_{3k})^2 \cdot x_k^2 - [\sum_{k=1}^N (u_{3k})^2 \cdot x_k]^2} \quad (11b)$$

The first non-linear model has to be transformed to linear one [9]:

$$\ln(y_k) = \ln(w_{11}) + w_{12} \cdot x_k \Rightarrow y' = \alpha_{12} \cdot x_k + \alpha_{11} \quad (12)$$

so we can calculate parameters  $\alpha_{12}$  using equation (11a) and  $\alpha_{11}$  using equation (11b).

After that, to obtain parameters for first non-linear model we have to recalculate parameters as follow:  $w_{11} = \alpha_{12}$  and  $w_{12} = \exp(\alpha_{12})$ .

### 5. RESULTS

All the experiments were run in the MATLAB environment. For our purposes we used:

- artificial data with simulated errors resulted from measurement inaccuracy,
- real data obtained from Escherichia coli and Spingomonas aromaticivorans cultures.

The noise were obtained with MATLAB ‘RANDN’<sup>#</sup> function. Iterations were stopped as soon as the Euclidean norm of a difference between two successive **U** matrices dropped below  $10^{-5}$ .

Figure 2 presents results of modelling for selected noised artificial data. The numerical parameters [according to equation (2)] obtained for those data with and without noise are presented in Table 1.

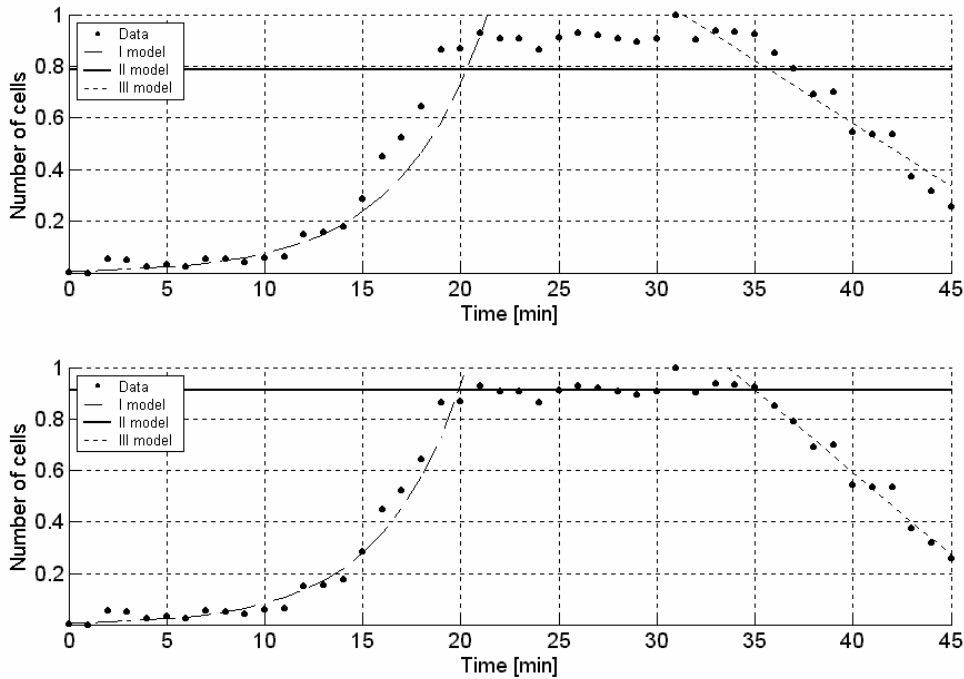


Fig. 2 The result of modelling for noised artificial data. Initial state (top) and the final result (bottom).

Tab.1 The results of modelling for artificial data with and without noise.

Parameter	$\lambda$	$F_0$	$F_S$	$M$
Simulated	0.27	1.6000	290.7134	-20.000
Model without noise	0.27	1.5982	290.7156	-20.000
Model with noise	0.24	2.4293	291.5946	-20.547

<sup>#</sup> RANDN produces pseudo-random numbers, with mean zero and standard deviation one.

Figure 3 presents results of *S. aromaticivorans* dynamics population modelling and Figure 4 presents results of *Escherichia coli* dynamics population modelling.

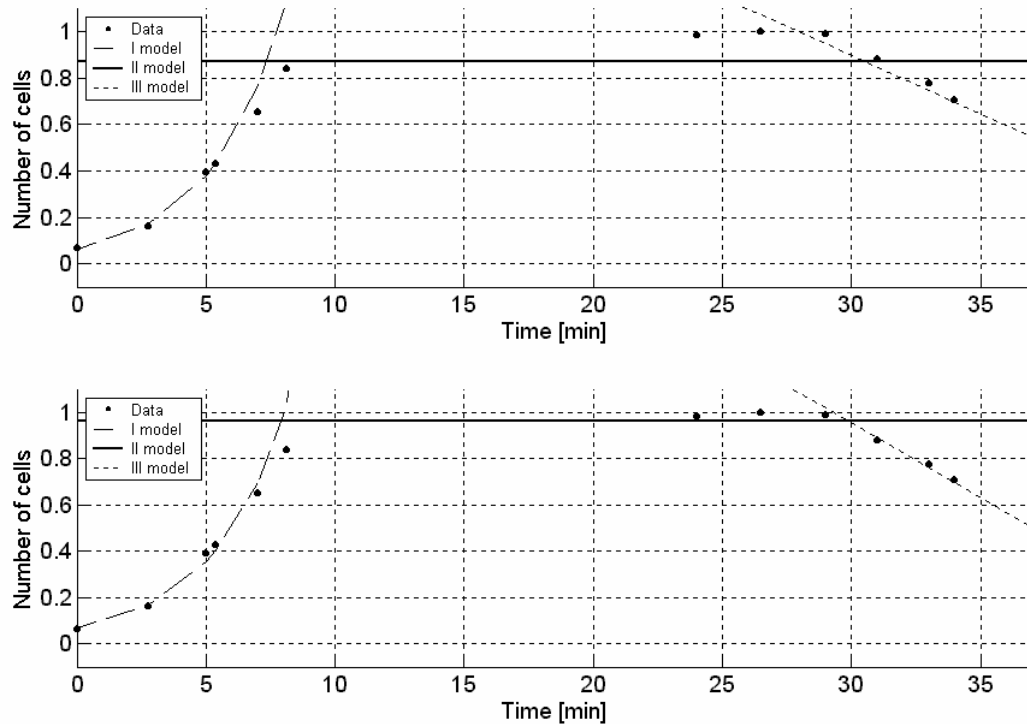


Fig.3. The result of *S. aromaticivorans* bacteria modelling. Initial state (top) and the final result (bottom).

## 6. CONCLUSIONS

In this paper an application of the fuzzy c-regression models algorithm to the bacterial population dynamics modelling is presented. The bacterial growth curve has been given by the piecewise continuous function. The promising results (the parameters of c-regression models estimates and transition points) for artificial as well as real data were obtained with small computational effort. Simultaneously, fuzzy clustering of used data were made.

In the future we are planning to use our method for more complex models and problems.

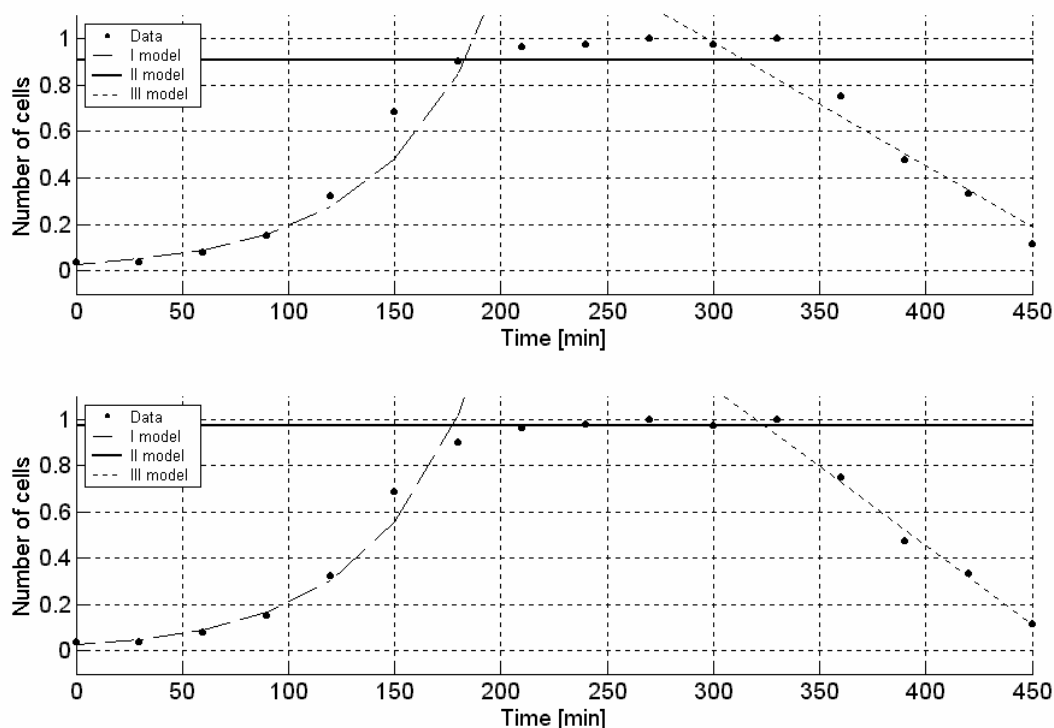


Fig.4 The result of E. coli bacteria modelling. Initial state (top) and the final result (bottom).

#### BIBLIOGRAPHY

- [1] CZOGAŁA E., ŁĘSKI J., Fuzzy and Neuro-Fuzzy Intelligent Systems. Studies in Fuzziness and Soft Computing, Vol. 47, Springer-Verlag, Haidelberg, 2000.
- [2] CZABAŃSKI R., Automatic Context Determination For Fuzzy Conditional Clustering. SYMBIOSIS 2001 Conference, Szczyrk, pp. 56-58, September 2001.
- [3] ŁĘSKI J.,  $\varepsilon$ -Insensitive Fuzzy c-Regression Models: Introduction to  $\varepsilon$ -Insensitive Fuzzy Modelling. IEEE Transaction on System, Man, and Cybernetics, in print, 2003.
- [4] HÖPPNER F., KLAWONN F., KRUSE R., Fuzzy Cluster Analysis. WileyEurope, 1999.
- [5] HATHAWAY R. J., BEZDEK J. C., Switching regression models and fuzzy clustering, IEEE Transaction on Fuzzy Systems. Vol.1, No. 3, pp. 195-204, August 1993.
- [6] ABONYI J., BABUSKA R., CHOVAN T., SZEIFERT F., Incorporating Prior Knowledge in Fuzzy c-regression Models Application to System Identification. Proc. of the Intelligent Systems in Control and Measurement Symposium, INTCOM 2000, pp. 99-110, Veszprem, Hungary, 2000.
- [7] OWCZAREK A. MOMOT M. SOBIERSKA E., Bacterial Growth Modelling Based on Fuzzy c-Regression Models. IX Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie, Piwniczna, in print, 2003.
- [8] BROCK T., MADIGAN T., MARTINKO J., PARKER J., Biology of Microorganism. 7-th Edition, Prentice Hall, New Jersey, pp. 327 – 330, 1994.
- [9] BRANDT S., Analiza danych. Wyd. 2, Wydawnictwa Naukowe PWN, Warszawa, 1999.