Fuzzified Pipes Dataset to Predict Failure Rates by Hybrid SVR-PSO Algorithm

Jaber Soltani[1], Moosa Kalanaki[2] & Mohammad Soltani[3]

[1] Assistant Professor of Irrigation and Drainage Engineering Department, Aboureihan Campus, University of Tehran, Iran

[2] MSc. of Artificial Intelligence, Malek Ashtar University of Technology, Iran

[3] MSc. Student of productivity and system management Department, University of Mehralborz, Iran

Correspondence to: Jaber Soltani, Assistant Professor of Irrigation and Drainage Engineering Department, Aboureihan Campus, University of Tehran, Iran. E-mail: jsoltani@ut.ac.ir

Abstract

This paper proposes a Support Vector Regression (SVR) based on Fuzzified Input-output Variables which has good comprehensibility as well as satisfactory generalization capability. SVM provides a mechanism to predict data from training ones. Then, results from proposed Fuzzified SVR-PSO (FSVR-PSO) model are compared with other methods; comparative tests are performed using pipe failures data. The analysis and the experimental results show this method has high comprehensibility as well as satisfactory generalization capability.

Keywords: SVR, Fuzzified Input-Output, PSO, failure rates

1. Introduction

The operations of Fuzzy system rely strongly on inference procedures, and a standard engineering application of fuzzy systems comprises an inference framework integrating vague or incomplete measurements. One of the main plans of competent management and optimal operation of urban water distribution network is to adduce a model for forecasting the breakages of urban water distribution networks.

Since the ground truth or the correct class might be unknown or fuzzy, the so-called fuzzy SVM assigning memberships to several classes to single observations have been developed by (Huang & Liu, 2002).

Though, the output of those FSVM is still crisp and no fuzzy output is generated. Therefore, so-called fuzzy-input fuzzy-output SVM capable of receiving soft labeled data and producing soft outputs with memberships assigned over multiple classes have been developed (Thiel, Scherer & Schwenker, 2007) and (Borasca, Bruzzone, Carlin & Zusi, 2006). In most cases, accidents and pipe failures occur as a result of several factors some of which being measurable such as age, length, diameter, depth and pressure of the pipes (Tabesh, Soltani, Farmani & Savic, 2009).

Henceforth, to accomplish better results we require a thorough model to consider every one of these factors. Numerous studies have been done in this field with numerous sorts of methods over two decades some of them depend on traditional systems and others based on intelligent techniques containing ANN, ANFIS, fuzzy logic, and SVM in related field. SVM techniques utilized non-linear regression for environmental data and proposed a multi-target methodology, MO-SVM, for automatic configuration of the support vector machines in view of a based on a genetic algorithm. MO-SVM demonstrated more exact in forecasting performance of the groundwater levels than the single SVM (Giustolisi, 2006).

The comparison among NLR, ANN and ANFIS methods had been done in the other consideration. The consequences of the comparison between ANN and ANFIS indicated that ANN model was more trusted and faithful because ANN model is hypersensitive to pressure, diameter and age than ANFIS. (Tabesh et al., 2009).

A probabilistic measure of the failure rate was defined and formulated for cases where the pipe lifetimes follow parametric models. The resulting theoretical failure rates were time-invariant and the parametric models would be useful only if the failure rates of water distribution pipes were stationary random processes (Dehghan, McManus & Gad, 2008).

To discover the efficient parameters in pipe failure rates of water distribution system, a combined model (ANN -GA) was disclosed. Particularly, ANN model was developed for making a connection between the parameters of breakage and pipe failure rates (Soltani & Rezapour Tabari, 2012).

Input-Output data are fuzzified and taken into SVR-PSO model for forecasting at that time compared to the previous results in this article. This study is intended at enhancing parameters associated to SVR and choosing the excellent SVR-PSO parameters for better pipes failure rate forecasting.

2. Methodologies

2.1 SVR (Support Vector Regression)

SVR or Support vector machine regression is suggested as a method for evaluating the mapping function from Input space to the feature space depending on the training dataset (Vapnik & Chapelle, 2000).

In the SVR model, our goal is to evaluate and guess two parameters (w and b) to achieve the perfect results. The distinctions between actual data sets and guessed results are displayed by ε in SVR. Slack variables are (ξ) expressed to concede some errors that appeared by noise or other factors. Consequently, the algorithm cannot be estimated if we do not employ slack variables, and therefore some errors may occur. When Margin is defined as $margin = \frac{1}{\|w\|}$ Then, for maximizing the margin, through minimizing $\|w\|^2$, the margin becomes maximized. These operations are shown in (1) and (2) equations and these are concluded as a basis for SVR (Smola & Scholkopf 1998).

$$\text{Minimize:} \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{n}\xi_i \tag{1}$$

$$\text{subject to: } y_i(w^T x_i + b) \geq 1 - \xi_i, \; \xi_i \geq 0 \tag{2}$$

x_i and y_i represent respectively the input space and the feature space. w is the weight vector and b is the bias, which SVR will compute them in the training process. The trade-off between the margin size and the amount of error in training are determined by C parameter.

A kernel function (a linear separator) is based on inner vector products and defined as follows:

$$k(x_i, x_j) = x_i^T x_j \tag{3}$$

If data points by applying $\varphi: x \rightarrow \varphi(x)$ are moved to the feature space (higher dimensional space), their inner products turn into (4) (Vapnik, 2010).

$$k(x_i, x_j) = \varphi(x_i)^T . \varphi(x_j) \tag{4}$$

x_i is and x_j are respectively the support vectors and the training data.

Correspondingly, the following function becomes by employing kernel functions and determining derivatives of w and b, also employing Lagrange multiplier the SVR function $F(x)$.

$$F(x) = \sum_{i=1}(\overline{\alpha_i} - \alpha_i) K(x_i, x) + b \tag{5}$$

α_i as the vector of Lagrange multipliers represents the support vectors. If these indicated multipliers are not equal to zero, then they are multipliers; otherwise, they could represent the support vectors (Vapnik, 1992).

Loss function determines the way of penalizing the data during estimating. A Loss function implies to ignore the errors which are related to the points falling within a certain distance. If ε-insensitive loss function is utilized, errors between -ε and +ε are going to be ignored. If C=Inf is set, regression curve will follow the training data inside the margin which is determined by ε (Qi, Tian & Shi, 2013). The related equation can be observed in (6).

$$|\xi|_\varepsilon = \begin{cases} 0 & if\ |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & otherwise. \end{cases} \tag{6}$$

2.2 Fuzzy Basis Function Inference System

The most common fuzzy rule-based system consists of a set of linguistic rules in the following form IF premise (antecedent)

$$\text{Then conclusion (consequent)} \tag{7}$$

This form is referred as the IF-THEN rule. It typically expresses an inference such that if we know the fact (premise), then we can infer another fact, called the conclusion. In this work we consider the general case where the fuzzy rule base consists of M rules in the following form:

$$R_j: \text{IF x1 is } A_1^j \text{ and x2 is } A_2^j \text{ and x3 is } A_3^j$$

$$\text{THEN z is Bj for } j = 1,2,3\ldots M \tag{8}$$

where xi (i=1,2,..,n) are the input variables; z is the output variable of the fuzzy system; and A_i^j and B j are

linguistic terms characterized by fuzzy membership function $u_{A_i^j}(x_i)$ and $u_{B_i^j}(x_i)$, respectively.

In this paper, triangular fuzzy function has been utilized precise time; since triangular fuzzy number can indicates fuzzy data, to lessen these mistakes impacts on the last generalization capability of SVM, the left and right spreads of triangular fuzzy number are regarded into the formation of SVM (Wu and Law, 2011).

2.3 Particle Swarm Optimization Algorithm

And has a position represented by a position vector (Kennedy & Eberhart, 1995). A swarm of particles moves through the problem space, with the moving velocity of each particle represented by a velocity vector (Juang, 2004).

The particle swarm optimization (PSO) was formed by Kennedy & Eberhart, (1995). This algorithm resemble the arousing of social behaviour among people (particles) by a multi-dimensional search space, any particle serve as a probable solution.

PSO operates in three steps at first Define each particle as a potential solution to a problem and best positions have been selected. Each particle has a velocity and by selecting, the best of them these velocities will update.

$$v_i(t+1)=wv_i(t)+c_1r_1\left[\hat{x}_i(t)-x_i(t)\right]+c_2r_2[g(t)-x_i(t)] \tag{9}$$

$v_i(t)$ is the particle's velocity at time t and $x_i(t)$ is the particle's position at time t and $\hat{x}_i(t)$ is the particle's individual best solution in time t. g(t) is the swarm's best solution in time t and w is inertia weight.

2.4 Case Study

A part of a water dissemination organization of a city clinched alongside Iran is recognized as those contemplated zone. This city is a standout amongst the urban areas continuously every now and again visited eventually by perusing travellers (see Figure 1). Those territory of this locale may be 2,418 hectares, for a populace about 93719 people, supplied with 579,860 meters about conveyance pipes including steel pipes 800, 700 and 600 millimetres over diameter, asbestos cement and cast iron pipes 400, 300, 250, 200, 150, 100 and 80 millimetres in diameter.

The installation and execution of the network pipelines in this area were generally started in 1981. According to statistical records, this region has the highest failure rate especially on asbestos cement. In this study, due to incomplete data on steel and cast iron pipes, asbestos cement pipes are only used in the modelling process (Soltani, 2009).
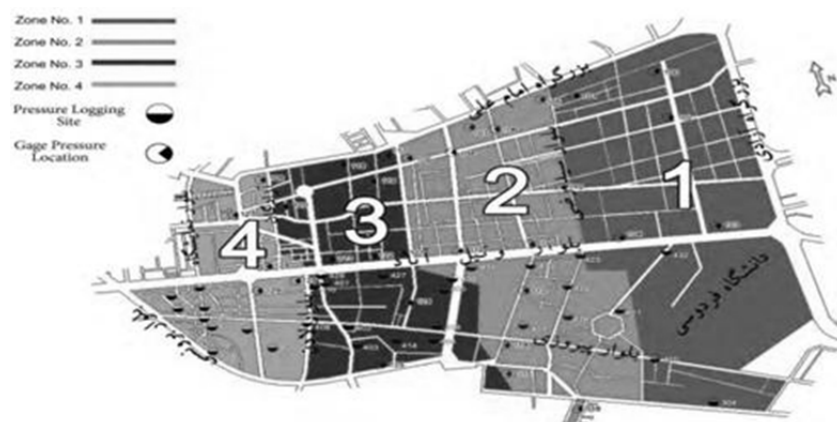

Figure 1. Schematic of study area and pressure measurement points (Soltani, 2009)

In order of modelling the failure rate of the asbestos cement pipes, the daily events have been recorded from 2005 to 2006 and analysed as to 2438 record data including some information such as diameter, year of implementation,installation depth, total accidents happened and the average of hydraulic pressure. These data have been collected from local water and water waste company.

2.5 Material and Methods

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis for decision-making. The process of fuzzy inference involves all of the pieces like membership functions, fuzzy logic operators, and if-then rules. The Fuzzification module transforms the crisp inputs-Outputs into fuzzy values. After this these values are processed in the SVR model.

(a)   (b)   (c)

(d)   (e)   (f)
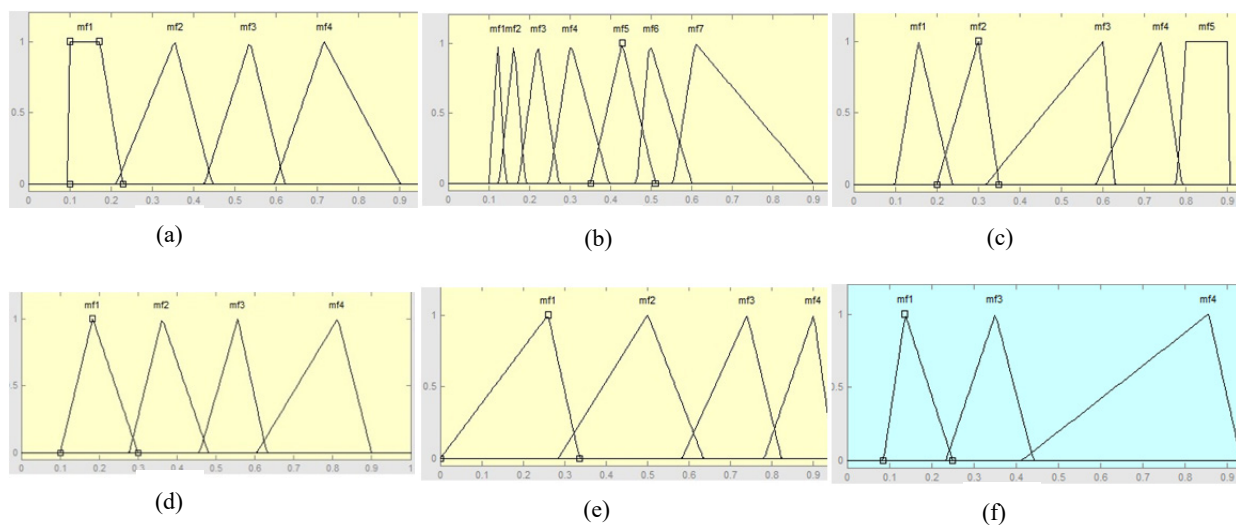
Figure 2. Fuzzified Input-Output data

In Figure 2. (a)-(e) are fuzzified Input and (f) is fuzzified output.

Generally, earlier previous studies have appropriated SVR parameters achieved from the trial-and-error methods. Each parameter is evaluated for approaching the appropriated values in trial-and-error methods. Nonetheless, this method has been very time-consuming and is not inadequately accurate. Accordingly, an integrated (SVR-PSO) model is recommended for searching the possible solutions in this research.

This research has been developed by MATLAB (version 7.12(R 2011a)) and for solving all of these problems, MATLAB SVM Toolbox and parameters were localized by PSO. In order to normalize the Input values to the models, Equation (10) is used.

$$x_n = 0.8 \frac{(x-x_{min})}{(x_{max}-x_{min})} + 0.1 \qquad (10)$$

x is the original value, x_{min} is the minimum value and x_{max} is the maximum value between input values, and also x_n indicates the normalized values here (Johnson, 1999). So that, input results are between [0.1, 0.9].

Also, the root of mean squared error (RMSE), normal root of mean squared error (NRMSE) and determination coefficient (R^2) are utilized as assessment criteria of the model reliability in this paper.

$$R^2 = \frac{\left(\Sigma_{i=1}^n (y_{actual}-\bar{y}_{actual})(y_{pred}-\bar{y}_{pred})\right)^2}{\Sigma_{i=1}^n (y_{actual}-\bar{y}_{actual})^2 \, \Sigma_{i=1}^n (y_{pred}-\bar{y}_{pred})^2} \qquad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^n \left(y_{actual_i}-y_{pred_i}\right)^2} \qquad (12)$$

$$NRMSE = \frac{RMSE}{var(y_{actual})} \qquad (13)$$

$$MAE = \frac{\Sigma \left| y_{actual_i} - y_{pred_i} \right|}{n} \tag{14}$$

Where y_{actual} is the actual (observed) data, $y_{prediction}$ as the predicted data, $y_{average}$ as the average of data and n indicates the number of observations and Also, var(y_{actual}) represents the variance of actual data. The associated parameters are selected only by author's experiments. Number of iterations of all algorithms are set to 30 and initial population equalled to 25. Boundaries parameters that relate to SVR also are set as: $0 < \varepsilon \le 1$, $1 \le \gamma \le 10$ and $10 \le C \le 200$. The inertia weight was 0.9; acceleration constants C_1 and C_2 were considered 0.9 and 1.7 respectively according to PSO parameters. The flowchart of this research has been shown in Figure 3. In this Figure, data sets are fuzzified then entered into the model after that by using the PSO algorithm better parameters that related to the SVR model is produced (Smola & Scholkopf, 1998) & (Kennedy & Eberhart, 1995).
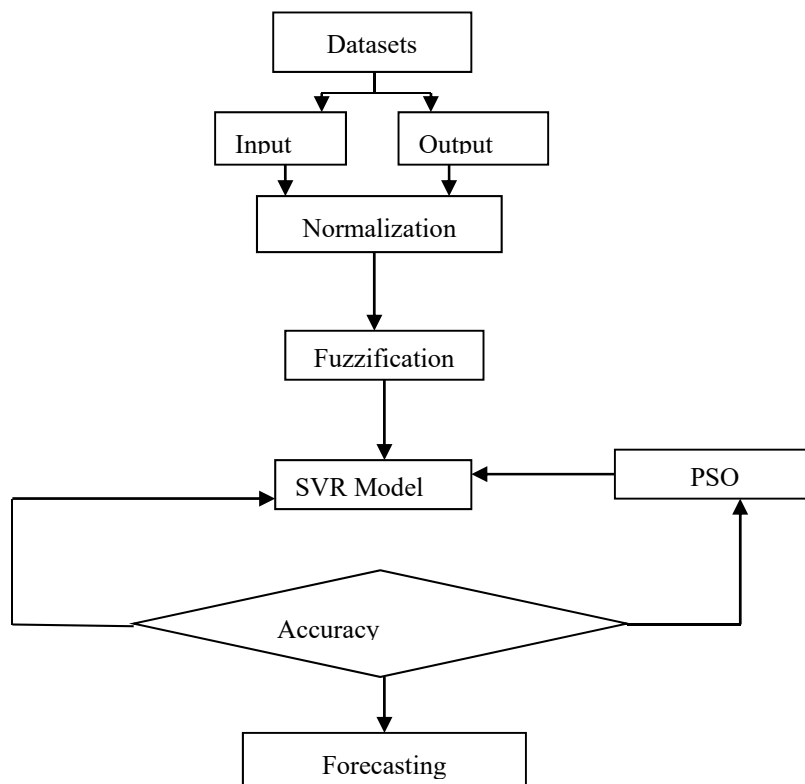

Figure 3. Prediction process by FSVR-PSO model

Table 1. Predicted result with RBF kernel function

Kernel Type	Time(s)	R^2	RMSE	NRMSE	MAE	C	P1	ε
RBF	471.237	0.9998289	0.00432	0.09369	0.00238	137.732	2.71635	0.000038

Table 1. Notes FSVR-PSO model with RBF kernel function and related SVR and error parameters. Table 1. Results has been shown in Figure 4.
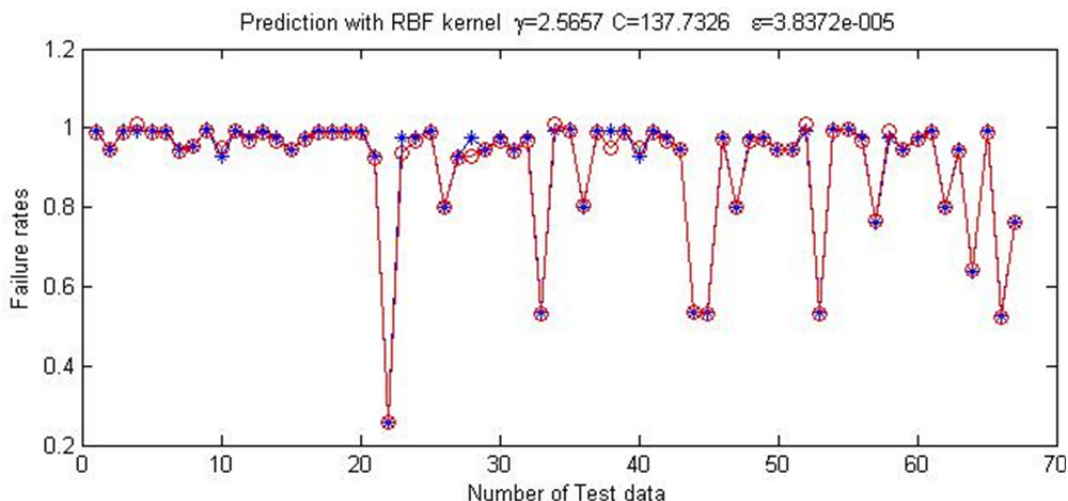
Figure 4. Predicted result with RBF kernel function

In this paper RBF kernel function has been used, because it shows the most appropriate results.

Table 2. Comparison results among other models.

Model Type		RMSE	R^2	NRMSE	Elapsed-time(s)
This Research	FSVR-PSO	0.004323	0.9998289	0.093699	471.27
Soltani & Rezapour Tabari, (2012)	ANN-GA	0.009775	0.999232	0.026433	1526.2
Tabesh et al., (2009)	ANFIS	0.020785	0.99802	0.17101	469.632

Table 2 indicates comparable results among three models that applied in the past for forecasting the same case study. The results indicate that FSVR-PSO model at its best performance is present in all factors; ANN-GA model has a good performance but has spent a long time. The elapsed time at ANFIS model is acceptable and sufficient but its allocated performance doesn't have adequate accuracy.

In this paper, we tried to make a relationship between the failure rate parameters in pipes with the number of events and failure of pipes which were considered as the main component of urban infrastructure, water supply and hygiene and health. Also, the optimal kernel function type and SVR related parameters have been extracted by utilizing Particle Swarm Optimization algorithm. These parameters have more accuracy and results obtained presented a better performance between actual and predicted data here.

3. Conclusions

In this paper, a Fuzzified-SVR-PSO (FSVR-PSO) model exhibited for anticipating the failure rates of pipes of water circulation networks, so as from claiming diminishing those number of breaks. In the suggested model, it might have been endeavored for creating a relationship among the parameters of failure rate in pipes which have been considered as the basic infrastructures, the for the number of occasions and disappointment from claiming pipes.

References

Borasca, B., Bruzzone L., Carlin, L., & Zusi, M. (2006). A fuzzy-input fuzzy-output SVM technique for classification of hyperspectral remote sensing images. In: Proceedings of the 7th Nordic Signal Processing Symposium, 2–5.

Dehghan, A., McManus, K. J., & Gad, E. (2008). Statistical analysis of structural failures of water pipes. *Proceedings of the Inst of Civil Eng - Water Management, 161*, 207 –214.

Giustolisi, O. (2006). Using a multi-objective genetic algorithm for SVM construction. *Journal of Hydroinf, 8,*

125-139.

Huang, H. P., & Liu, Y. H. (2002). Fuzzy support vector machines for pattern recognition and data mining. *International Journal of Fuzzy Systems, 4,* 826–835.

Johnson, R. (1999). Rainfall-Runoff Modelling Using Artificial Neural Networks. *J Hydrologic Eng., 4,* 232-239.

Juang, C. F. (2004). A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design. *IEEE TRANSACTIONS ON SYS, 34*, 997-1006.

Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *IEEE Int. Conf on Neural Networks,* 1942–1948.

Qi, Z., Tian, Y., & Shi, Y. (2013). Robust twin support vector machine for pattern classification. *Pattern Recognition, 46,* 305-316.

Smola, A., & Scholkopf, B. (1998). A tutorial on Support Vector Regression. *J. Statistics and Computing, 14,* 199-222.

Soltani, J. (2009). An Intelligent Model for Reliable Optimal Rehabilitation and Replacement of Water Distribution Networks Considering Pressure Dependency of Demand, Ph.D. Dissertation. University of Tehran, University College of Engineering School of Civil Engineering (in persain).

Soltani, J., & Rezapour Tabari, M. (2012). Determination of effective parameters in pipe Failure rate in water distribution system using the combination of Artificial Neural Networks and Genetic Algorithm. *Journal of Water & Wastewater, 83*, 2-18 (in persain).

Tabesh, M., Soltani, J., Farmani, R., & Savic, D. A. (2009). Assessing Pipe failure Rate and Mechanical Reliability of Water Distribution Networks Using Data Driven Modeling. *Journal of Hydroinf, 11*, 1-17.

Thiel, C., Scherer, S., & Schwenker, F. (2007). Fuzzy-input fuzzy-output one-against-all support vector machines. *11th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 3*, 156–165.

Vapnik, V. N., & Chapelle, O. (2000).Bounds on error expectation for support vector machines. *Neural Computation, 12*, 2013-2036.

Vapnik, V. N. (1992). Principles of risk minimization for learning theory. *Advances in Neural Information Processing Sys, 4*, 831-838.

Vapnik, V. N. (2010). The Nature of Statistical Learning Theory, Springer Verlag, US.

Wu, Q., & Law, R. (2011). The complex fuzzy system forecasting model based on fuzzy SVM with triangular fuzzy number input and output. *Expert Systems with Applications, 38*(10), 12085–12093.

Copyrights