

RESIDUAL BASED FLOW TIME ESTIMATION ALGORITHM FOR LABOR INTENSIVE PROJECT TYPE PRODUCTION SYSTEMS

(EMEK YOĞUN ÇALIŞILAN PROJE TİPİ ÜRETİM SİSTEMLERİ İÇİN ARTIK
TERİM TABANLI AKIŞ ZAMANI TAHMİN ALGORİTMASI)

Aslan Deniz KARAOĞLAN¹

ABSTRACT

Regression-based methods are widely used for flow time estimation of customer orders. However, for the customer orders that will be produced for the first time in a labor intensive project type production system with new design parameters, it is hard to make thoroughly accurate flow time prediction at the quotation stage. This is caused by having so many uncontrollable factors in a production system, that are not placed in the mathematical models. These uncontrollable factors cause high differences between the observed and expected flow time. In this study, a new algorithm - that combines the regression analysis and the artificial neural networks - is proposed. By this way, the prediction performance of fitted regression model is improved and the lack-of-fit is decreased.

Keywords: Regression analysis, neural network applications, flow time, project type production

ÖZ

Regresyon tabanlı metodlar, müşteri siparişlerinin akış zamanının hesaplanmasında yaygın olarak kullanılmaktadır. Ancak emek yoğun proje tipi üretim yapan işletmelerde, ilk defa yeni tasarım parametreleri ile üretilecek olan ürünlerin akış zamanını, üretime başlamadan önce müşteriye fiyat teklifi verme aşamasında tahmin etmek zor bir problemdir. Bu durum, üretim sisteminin matematiksel modellerde yer verilemeyen pek çok kontrol edilemeyen değişken içermesinden kaynaklanır. Bu kontrol edilemeyen değişkenler ise beklenen akış zamanı ile regresyon tabanlı denklemlerle tahmin edilen zamanlar arasında hatırı sayılır bir tahmin hatasını ortaya çıkarır. Bu çalışmada, regresyon denkleminin tahmin hatasını minimize etmek üzere, yapay sinir ağları ile regresyon analizini birleştiren bir algoritma önerilmiştir. Bu yolla regresyon denkleminin tahmin performansı artırılmış ve tahmin hataları minimize edilmiştir.

Anahtar Kelimeler: regresyon analizi, yapay sinir ağları, akış zamanı, proje tipi üretim

¹ Balıkesir Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, BALIKESİR, deniz@balikesir.edu.tr

1. INTRODUCTION

Accurate flow time prediction and due date assignment for a customer order are important in the real production environment. There are so many studies presented in the literature to estimate the flow time. Some of these studies presented remarkable results calculated by regression-based methods and artificial neural networks (ANN) which are related to the subject of this paper. Ragartz and Mabert [1] estimated the flow time by using linear and nonlinear regression analysis. They used the standard deviation of lateness, mean lateness as response while the general shop characteristics (such as the number of jobs waiting in the queue, the number of work-stations etc.) were considered as factors. Vig and Dooley [2] also used regression analysis to estimate the flow time and the due date. Arizono et al. [3] studied the prediction and minimization of the total actual flow time by using Gaussian machine model. They used job sequences, processing times, setup time, starting time as input to predict the actual flow time. Enns [4] proposed a forecasting approach for flow time prediction in a job shop. In Enns [4], the flow time and due date of the orders are determined by considering the forecasting errors and the confidence interval of the delivery performance level. Harris [5] modeled the relations between the factors (transporter speed, buffer capacity, dispatching rule, shop load and routing configuration) and the responses (mean flow, makespan, the variance of flow times and deadlocks) by using fractional factorial design. Veral [6] employed response surface methodology (RSM) via regression analysis to model operation flow time characteristics which are nonlinear depending on a dispatching rule. Goving and Roeder [7] used RSM to predict the flow time of the orders that have variable routing during the production process. They used mean and variance of remaining time from the starting point as a factor and flow time as a response. Li et al. [8] combined the back propagation network (BPN) model and genetic algorithms (GA) for completion time prediction. GA was adopted in the BPN to determine the BPN's parameters and to improve the accuracy of completion time prediction. They used factors such as the number of parallel resources, the number of mean process steps for each job, the number of overlaps and the workload to predict the completion time. Alenezi et al. [9] developed a support vector regression model for predicting real-time flow time in multi-resource and multi-product systems. Asadzadeh et al. [10] presented a flexible algorithm based on ANN, fuzzy regression (FR) and conventional regression for forecasting lead time. Chen and Wang [11] estimated the job cycle time using principle component analysis, fuzzy logic and BPN. Kumru and Kumru [12] used ANN to forecast operation times in the metal industry. Pan and Dong [13] proposed an improved migrating birds optimization (MBO) to minimize the total flowtime for a hybrid flowshop scheduling problem. Ribas et al. [14] presented a high performing Discrete Artificial Bee Colony algorithm for the blocking flow shop problem with flow time criterion. To find the best configuration of the algorithm, they used design of experiments (DOE).

These studies generally focused on the characteristics related with job or job-shop (such as processing times, cycle times, variances, the number of work-stations, workload, the number of jobs in the queues etc.) to predict the flow time. However, these studies were presented under the assumption that the processing times or their probability distributions were known or could be calculated from the previous data.

The literature review indicates that there is no presented study that uses directly the product specifications or order specifications demanded by the customer to predict the processing times and flow time instead of using the general job and shop characteristics. In

other words, the processing times and flow times are not predicted by using the properties of the product and the customer specifications. However using the product specifications or order specifications demanded by the customer directly (instead of using job or shop characteristics as presented in the literature) may be necessary for some real industrial problems. One example is the manufacturers those produce their products by labor intensive project type manufacturing in a job shop environment. In these systems the processing times and flow times are directly influenced from the technical specifications of the orders and it is not realistic to predict these data by only discussing the status of the shop floor. For example, as presented in Section 3; the power, voltage, and weights of the used raw materials directly affect the processing times and the flow time although a single product is produced just like transformers. Using general job and shop characteristics are valid when the processing times are known before starting the production. However in labor-intensive project type manufacturing this information is not known. This is reasoned from performing different effort for the orders when the specifications are varied.

In project type manufacturing, most of the products are produced for the first time at the production system according to the customer demands and there are no processing times measured for this order's new design parameters. These unknown processing times may cause problems while calculating the labor hour, labor cost and bidding for an order at the quotation stage. Especially in labor intensive production systems, this situation causes remarkable deviations between expected and observed flow time which causes ambiguity at giving price offer and due date to the customers.

This study proposes a new method which uses the product (or order) specifications based on the customer demands. In this way, it is possible to estimate the processing times and flow times of the products that will be produced the first time. This proposal is tested by conventional regression analysis method. However, high prediction errors occurred because of many unconsidered product or production parameters. To cope with this problem, a residual based flow time estimation algorithm (RBFEA) for the labor intensive project type production is proposed. By using RBFEA, it is aimed to increase the prediction performance of regression models for flow time estimation.

The proposed algorithm is described in the following section. The case problem and the production system of the selected transformer producer are described in Section 3. The prediction results of conventional regression analysis, RBFEA and performance comparisons are given in the same section. Results and discussions are given in Section 4. Finally, the conclusions are summarized in Section 5.

2. PROPOSED RESIDUAL BASED FLOW TIME ESTIMATION ALGORITHM

In this study, a Residual Based Flow Time Estimation Algorithm (RBFEA) for the labor intensive project type production is proposed. In the conventional flow time prediction; the characteristics related to the job or shop are used to predict the flow time. If the products vary according to the customer demands in a project type manufacturing environment, the processing times or probability distributions are unknown. In such a situation where the most of the products are being produced for the first time; using the technical specifications of the products instead of using job or shop characteristics to predict the flow time is more realistic. However, a production system includes numerous factors those affect the processing times and flow time. Therefore, this may cause high prediction errors between the observed and

expected flow time. In the literature, regression-based approaches are used effectively for flow time estimation. However, due to the nature of the problem, the systematic prediction error must also be considered and have to be added to the expected flow time, which is fitted from a mathematical equation.

The flow time is predicted by a regression equation that uses technical specifications demanded by the customer as a factor. The effect of unknown factors (or production parameters) are estimated by artificial neural networks and these effects are added to the previously predicted flow time values. By this way, the lack-of-fit is decreased and more accurate flow time estimation is obtained. Pseudo-code of *RBFEA* is given below:

START

(Step 1) DETERMINE the factors of regression equation those have effect on flow time

(Use the technical specifications of the product those are demanded by customers)

(Step 2) FIT the regression equation by considering the previously observed production data:

$$\mathbf{Y} = \beta \mathbf{X} + \varepsilon \quad (1)$$

where \mathbf{Y} is the *output matrix* that is composed of flow time values, \mathbf{X} is the *input matrix* that is composed of technical specifications of the order, and ε is the *residuals*. The least square estimator of β matrix that is composed of coefficients of the regression equation is calculated by the given formula in Equation (2):

$$\beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2)$$

(Step 3) TEST the regression equation for goodness-of-fit according to the coefficient of determination (R^2)

(Step 4) IF regression model is adequate according to the ANOVA table **THEN**

DO

IF there is no previously trained ANN for residuals **THEN**

DO

(4.1) DETERMINE the *expected value for flow time* (\hat{Y}_i) by using the fitted regression equation.

(4.2) DETERMINE the *prediction error ratio* (PE_i) for each expected flow time value by the formula given in Equation (3)

$$PE_i = \frac{(Y_i - \hat{Y}_i)}{\hat{Y}_i} = \frac{\varepsilon_i}{\hat{Y}_i} \quad (3)$$

where Y_i is the *observed flow time* and ε_i is the residual that is calculated from the difference between the observed and estimated (fitted) flow time values.

(4.3) TRAIN ANN for PE_i values where \mathbf{X} is the input matrix and \mathbf{PE} is the output matrix.

(4.4) TEST the trained ANN model for goodness-of-fit by using the test data.

IF the test results are acceptable then **RECORD** the ANN topology

AND GO TO Step 5.

ELSE GO TO Step 4.3.

END

(Step 5) USE the regression equation that is fitted at Step 2 and perform flow time prediction (\hat{Y}_i) for the new customer order that will be produced for the first time at the production system (at quotation stage before starting the production).

(Step 6) USE the ANN that is trained at Step 4.3 and perform prediction for the *predicted prediction error rate* $PE(pred)_i$ for the observations presented in Step 5.

(Step 7) USE the predicted $PE(pred)_i$ values at Step 6 to calculate the predicted residuals:

$$\hat{\varepsilon}_i = (PE(pred)_i)(\hat{Y}_i) \quad (4)$$

(Step 8) USE Equation (5) to calculate the adjusted prediction for flow time ($\hat{Y}_i(adj)$) of the order that will be produced for the first time at the production system by using the predicted flow time (\hat{Y}_i) (at Step 5) and the predicted residual ($\hat{\varepsilon}_i$) (at Step 7):

$$\hat{Y}_i(adj) = \hat{Y}_i + \hat{\varepsilon}_i \quad (5)$$

Where the predicted residual ($\hat{\varepsilon}_i$) is calculated by (\hat{Y}_i)(PE_i).

(Step 9) IF there is a new order that its flow time has to be predicted then **RETURN** to Step 5 **ELSE STOP**

The implementation of proposed algorithm is given in the next section.

3. CASE STUDY AND THE EXPERIMENTAL RESULTS

3.1 Production System of the Selected Transformer Producer

In the company that is selected for the application of this manuscript, transformers are produced according to the labor intensive project type production. A transformer is an electrical device which is used to convert one alternative current (AC) voltage to another through electromagnetic induction. In other words, a transformer transfers energy between two or more circuits through electromagnetic induction. By using this device, the electric power can be transmitted to far away from its source [15]. The firm produces oil type and dry type cast resin transformers. In the case company, the flow time varies according to the properties of received orders indicated in the customer order specifications. Except a few similar transformers produced frequently; there are unlimited types of different transformer orders received from the customers. If an order for a transformer that has not been produced in the past is received by the company; the way to calculate the labor time in this firm is to review the flow times of previously produced transformers (which differs from the current order) from the data base and try to estimate the possible flow time and labor cost inferentially for the current order. This causes problems at giving accurate price offers to the customers because of unknown processing times and flow times of unlimited kind of orders those will be produced for the first time.

Although the production process seems unique for a transformer production, the process times of each operation and the flow times of each order vary according to the technical specifications that are demanded by the customer. In this study, the flow time of the oil type power transformers for each order is predicted by conventional regression analysis and *RBFEA* using the technical specifications unlike other studies (which used general job and shop characteristics) reported in the literature. The factors those are indicated in the customer

demand forms (which are prepared by sales department and have an effect on the processing times) are given in Table 1. The minimum and maximum roughly bounds of selected samples are also given in the same table.

Table 1. List of factors those have effect on processing times and their levels [16]

Technical Specifications	Level		
	Unit	Min	Max
Power	kVA	4000	400000
Voltage	kV	10	650
NLTC/OLTC	-	NLTC	OLTC
Weight of core (<i>amount of CRGOSS</i>)	Kg	4000	200000
Weight of windings (<i>amount of Copper</i>)	Kg	1250	85000
Weight of tank (<i>amount of st37-steel</i>)	Kg	1250	85000
Weight of oil	Kg	2000	140000

The processing times for each order that vary according to the customer or local specifications are obtained from Oracle ERP database of the company. For this purpose, a data set that is composed of 20 different specifications are selected. Because of the confidentiality, the original data is not given in this paper. Instead, coded values of specifications and the processing times are used for mathematical modeling and also for training ANN. Input matrix for the training dataset is coded by dividing each column to appropriate constants, and in this way, all the results became smaller than 1 and normalized. The data set used for mathematical modeling with regression and ANN training is given in Table 2.

Table 2. Coded values for sample data set [16]

Sample Number	Power (kVa)	Max. Voltage	NLTC/OLTC	CRGOSS (Kg)	Copper (Kg)	ST 37 – Steel (Kg)	Oil (Kg)	Observed Flow Time (Hours)
i	$X1i$	$X2i$	$X3i$	$X4i$	$X5i$	$X6i$	$X7i$	Y_i
1	0.3865	0.648	0.1	0.38580	0.321652	0.321744	0.5240	0.3122906
2	0.1000	0.220	0.1	0.04480	0.075080	0.132796	0.2032	0.1607784
3	0.1000	0.220	0.1	0.04480	0.075244	0.131572	0.2032	0.1427294
4	0.2300	0.380	0.1	0.19470	0.109040	0.192804	0.2450	0.1315236
5	0.1300	0.300	0.1	0.13874	0.106316	0.142404	0.1828	0.1241224
6	0.2000	0.420	0.2	0.19991	0.078528	0.202468	0.2528	0.1648076
7	0.0762	0.161	0.2	0.06628	0.039656	0.062800	0.0812	0.0611054
8	0.0762	0.161	0.2	0.06628	0.039632	0.062528	0.0812	0.0623226
9	0.0742	0.161	0.2	0.06736	0.039636	0.062280	0.0804	0.0607624
10	0.0742	0.161	0.2	0.06736	0.039688	0.062064	0.0804	0.0592614
11	0.0090	0.032	0.2	0.01130	0.007804	0.011848	0.0144	0.0160044
12	0.0090	0.032	0.2	0.01130	0.007812	0.011836	0.0144	0.0146910
13	0.0135	0.033	0.2	0.01532	0.016752	0.015196	0.0200	0.0185196
14	0.0135	0.033	0.2	0.01532	0.016548	0.015196	0.0200	0.0174830
15	0.0135	0.033	0.2	0.01532	0.016784	0.015196	0.0200	0.0176256
16	0.0160	0.154	0.1	0.02125	0.015140	0.030600	0.0400	0.0350266
17	0.0350	0.154	0.1	0.03346	0.027456	0.039308	0.0636	0.0417936
18	0.0200	0.033	0.1	0.01856	0.017780	0.028224	0.0336	0.0280730
19	0.0200	0.033	0.1	0.01856	0.017804	0.028224	0.0336	0.0255004
20	0.04	0.11	0.1	0.0304	0.043896	0.063424	0.0060	0.0557000

The following section gives the results of regression analysis, *RBFEA* and performance comparisons, respectively.

3.2 Flow Time Estimation of Transformers Using the Regression Equation

The mathematical relationship between the factors (power (X_1), voltage (X_2), NLTC/OLTC (X_3), weight of core (X_4), weight of windings (X_5), weight of tank (X_6), weight of oil (X_7)) and the response (flow time (Y)) is calculated by the least square estimation. Regression coefficients for the regression model can be calculated by Equation (1)-(2) [17-20].

The X matrix is composed of the observed factor values given in Table 2 and the mathematical equation given in Equation (6) is calculated by using Minitab 16 which is a statistical analysis program. When the quadratic terms or interactions are considered in the mathematical model, the performance of the model decreases dramatically for the confirmation tests. Therefore, only the linear terms are selected. The R^2 (coefficient of determination) value for the model given in Equation (6) is calculated as 99.5% and R^2 (predicted) is calculated as 92.36% using the factor values given in Table 3.

$$\begin{aligned} \text{Flow Time} = & -0.0201215 - 0.371798(\text{Power}) + 0.0984925(\text{Voltage}) + 0.125379(\text{NLTC/OLTC}) \\ & - 0.366960(\text{Weight of Core}) + 0.480010(\text{Weight of windings}) + 0.829958(\text{Weight of tank}) + 0.223555(\text{Weight} \\ & \text{of oil}) \end{aligned} \quad (6)$$

According to the results of analysis of variance (ANOVA) for the 0.95% confidence level (where $\alpha = 0.05$) P-value is calculated as $0.000 < \alpha$ which means that the given mathematical model is significant. The performance of the mathematical equation given in Equation (6) is presented in Table 3. Also, the prediction error ratio which is calculated by using Equation (3) and the accuracy of the prediction are given in the same table.

Table 3. Performance of regression analysis for the samples given in Table 2

Sample Number	Observed value	Expected (fitted) value	Accuracy (%)	Prediction Error Ratio
i	Y_i	\hat{Y}_i		PE_i
1	0.312291	0.309539	99.12	0.00889
2	0.160778	0.152146	94.63	0.05674
3	0.142729	0.151209	94.06	-0.05608
4	0.131524	0.140013	93.55	-0.06063
5	0.124122	0.132806	93.00	-0.06539
6	0.164808	0.160851	97.60	0.02460
7	0.061105	0.057468	94.05	0.06329
8	0.062323	0.05723	91.83	0.08898
9	0.060762	0.057195	94.13	0.06237
10	0.059261	0.057041	96.25	0.03893
11	0.016004	0.017412	91.21	-0.08084
12	0.014691	0.017406	81.52	-0.15598
13	0.01852	0.022688	77.49	-0.18373
14	0.017483	0.02259	70.79	-0.22607
15	0.017626	0.022703	71.19	-0.22364
16	0.035027	0.035444	98.81	-0.01178
17	0.041794	0.042314	98.75	-0.01230
18	0.028073	0.020891	74.42	0.34378
19	0.025500	0.020902	81.97	0.22000
20	0.055700	0.052274	93.85	0.06554

Confirmations are also performed for the given regression equation. The data used for confirmation, the results obtained and the performance comparisons are given in Section 3.4. As proposed in this study, the technical specifications of the orders affect some part of the

prediction error. Therefore, the correlation between the technical specifications and prediction error may be investigated by the researchers. For this purpose, correlation analysis is performed for the absolute values of PE_i , which are given in Table 3. The Pearson correlations are calculated between the technical specifications and the prediction error using Minitab. The correlation between the power and the PE_i is calculated as -0.491 while correlations between the PE_i and voltage, the weight of core, the weight of tank and the weight of oil are calculated as -0.621, -0.455, -0.520, and -0.486, respectively. The significance values (P-values) are calculated as less than 0.05 (5%) which means that the correlation coefficients are significant. However, the correlations for NLTC/OLTC and weight of windings are not significant. These results indicate that there is a medium-sized correlation between each technical specification and prediction error ratio. This relationship indicates that the prediction performance of regression equation can be increased by using ANN.

3.3 Flow Time Estimation of Transformers Using the Proposed Residual Based Flow Time Estimation Algorithm (RBFEA)

The selected case company exports its products to 75 countries and has a wide range of products including the rectifier, current rectifier, wind turbine transformers and medium power transformers changing level under load, and these products vary according to the customer and local specifications. The designs of the same transformers with equal power and voltage values may be completely different because of technical specification variations. These differences require special engineering operations and these operations need specific designs related to the specifications. As a result of these design differences, both materials and labor time are far away from being stabilized.

In this section, to cope with this problem and perform accurate prediction of flow times (mostly depending on processing times) and labor costs of orders; a feed forward back propagation artificial neural network (FF-BPN) was employed by using Matlab program to estimate the prediction error rate (PE_i). From Step 1 to Step 4.2 of RBFEA are completed by the calculations given in Section 3.2 and will not be repeated here. The PE_i values given in Table 3 are used as output matrix and the factor values of samples given in Table 2 are used as input matrix of FF-BPN. Then predicted PE_i values ($PE(pred)_i$) are used to calculate the predicted residuals $\hat{\epsilon}_i$, and finally $\hat{\epsilon}_i$ is used to adjust the flow time estimation of the regression equation. Preliminary investigations are conducted to choose a suitable network topology and training algorithm for the network. The design parameter for hidden layers fixed to 3 hidden layers and given transfer (activation) functions at Table 4 after numerous preliminary trials. The ANN topology used in this study is presented in Figure 1 [21, 22].

According to the trials for investigating the appropriate network topology of the given network in Figure 1; the number of neurons at each hidden layer (from 4 to 16), type of activation function (purelin, tansig, logsig), learning rates (lr) (0.01 to 0.09 with different combinations for each experimental trial) and momentum constants (mc) (from 0.75 to 0.95 with different combinations for each experimental trial) are used and summarized in Table 4.

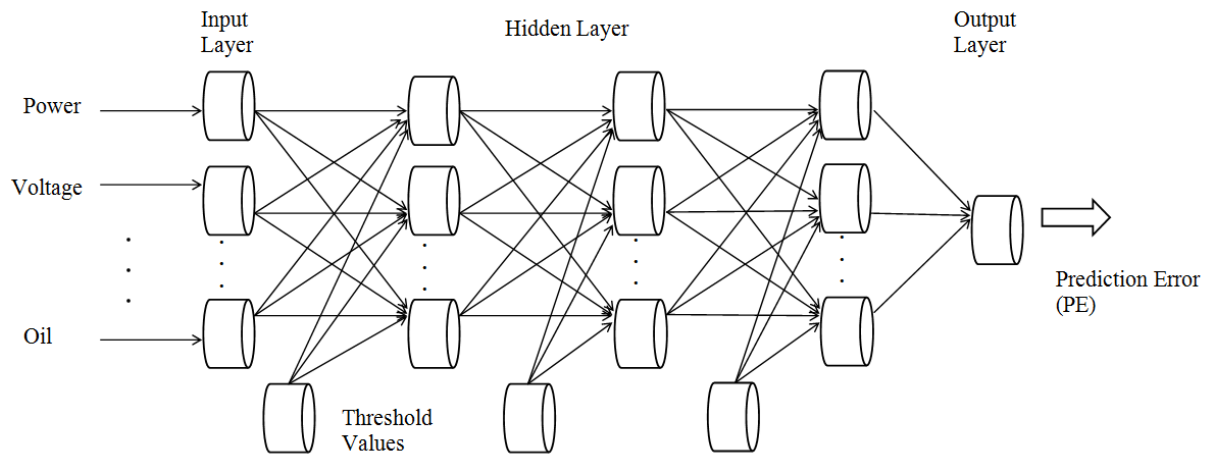


Figure 1. ANN topology for the given problem

The performance measurement is determined as mean square error (MSE) and 0.001 is the target MSE value. The MSE results of each parameter combination are recorded to a matrix that is defined in Matlab 10.0. Experimental runs are carried out to determine the MSE for each combination of parameters serially. The parameter combination that gives the minimum MSE is selected from the records as optimum training parameters. Totally 19 different topologies are constructed with various network parameter combinations of mc and lr (totally 324 combinations) and minimum MSE is reached at the given topology in Figure 1. In this topology, 5 neurons are used for each hidden layer. Purelin activation function is used for the neurons of input and output layers while logsig activation function is used for the neurons of hidden layers. The optimum training parameters of this network are found as 0.01 for lr and 0.95 for mc . It is clearly observed from the results that the memorization capability of the net is well enough for the training data set. By this way, Step 4.3 and Step 4.4 of *RBFEA* are completed. The final optimum network topology and its parameters are given in Table 4. The training performance for ANN are given in Figure 2.

Table 4. The optimum network topology and its parameters

Network parameters	Network parameter levels that are used in trials	Optimum values for network parameters
Type of ANN	FF-BPN	FF-BPN
Number of hidden layers	3	3
Number of neurons at each hidden layer respectively	4 – 16 (12 levels)	5, 5, 5
Number of neurons at input layer	7	7
Number of neurons at output layer	1	1
Transfer functions of hidden layers respectively	Combinations of Purelin, Tansig, Logsig	Logsig, Logsig, Logsig
Transfer function of input layer / output layer	Combinations of Purelin, Tansig, Logsig	Purelin, Purelin
Momentum constant / learning rate	0.75, 0.85, 0.95 0.01- 0.09 (9 levels)	0.95 0.01

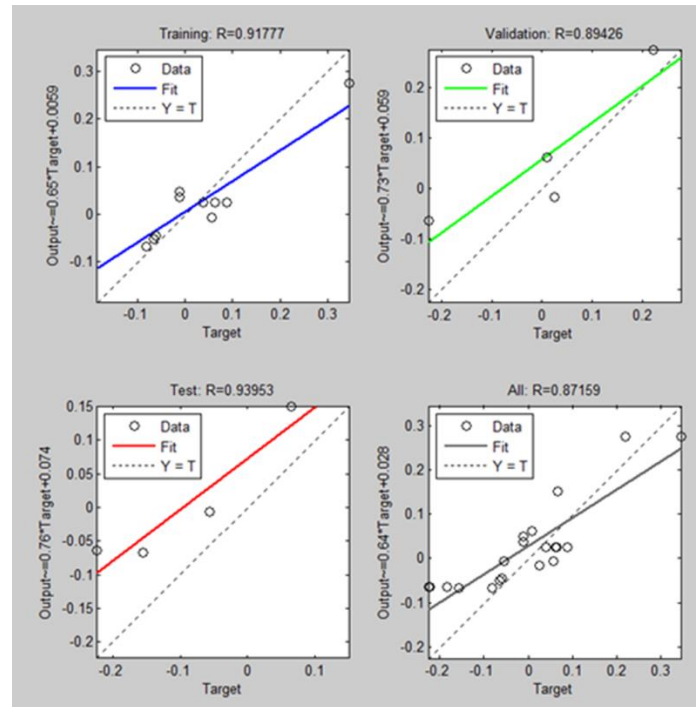


Figure 2. The training performance for ANN

By performing the Step 5 - Step 8 of *RBFEA*, the $PE(pred)_i$, $\hat{\epsilon}_i$, $\hat{Y}_i(adj)$ values are calculated. The formulation of $\hat{Y}_i(adj)$ for the case problem is given in Equation (7). Also, the surface plot for Equation (6) and Equation (7) are given in Figure 3 and Figure 4, respectively. The difference between the modeling approaches of conventional regression and *RBFEA* can be clearly observed from these figures.

$$\hat{Y}_i(adj) = \hat{Y}_i + \hat{\epsilon}_i = [-0.0201215 - 0.371798(Power) + 0.0984925(Voltage) + 0.125379(NLTC/OLTC) - 0.366960(Weight\ of\ Core) + 0.480010(Weight\ of\ windings) + 0.829958(Weight\ of\ tank) + 0.223555(Weight\ of\ oil)] + \hat{\epsilon}_i \tag{7}$$

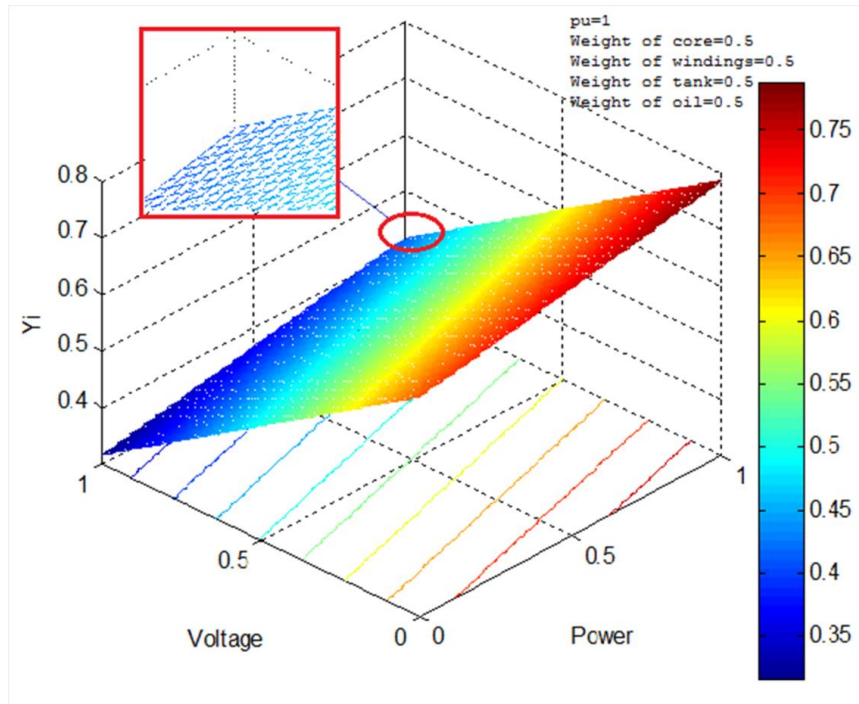


Figure 3. Surface plot for conventional regression (Equation (6))

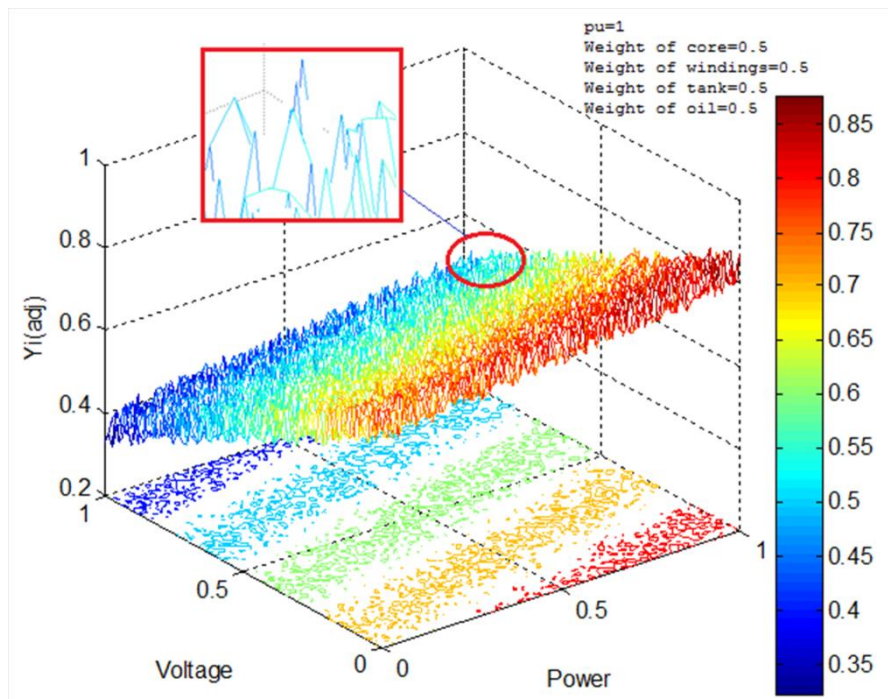


Figure 4. Surface plot for RBFEA (Equation (7))

Validation results for the training data (the samples that were also used for regression analysis) are given in Table 5. According to the results presented in Table 5, the overall accuracy of RBFEA estimation is better than the estimation of the conventional regression model. For the samples 1, 16, 17 and 20, the regression analysis gives better results in comparison with the proposed method (marked with bold). However, the rest of the samples are better modeled and predicted by RBFEA. The validation results and the performance comparison indicate that the generalization capability of RBFEA is better than the conventional regression analysis. That is why the effect of unknown factors those causes high residuals can be added to the regression equations by using RBFEA.

Table 5. Performance comparison for conventional regression analysis and RBFEA for the samples given in Table 2

Sample Number	Observed Value	Expected (fitted) Value by Regression Equation	Predicted Prediction Error Rate (ANN)	Predicted Residuals	Expected (fitted) Value by RBFEA	Accuracy For RBFEA Estimation (%)	Accuracy For Regression Estimation (From Table 5) (%)
i	Y_i	\hat{Y}_i	$PE(pred)_i$	$\hat{\epsilon}_i = (PE(pred)_i)(\hat{Y}_i)$	$\hat{Y}_i(adj) = \hat{Y}_i + \hat{\epsilon}_i$		
1	0.312291	0.309539	0.06090	0.018851	0.32839	95.10	99.12
2	0.160778	0.152146	0.0071	0.00108	0.153226	95.07	94.63
3	0.142729	0.151209	-0.0068	-0.00103	0.150181	95.04	94.06
4	0.131524	0.140013	-0.0452	-0.00633	0.133684	98.38	93.55
5	0.124122	0.132806	-0.0508	-0.00675	0.126059	98.46	93.00
6	0.164808	0.160851	0.0162	0.002606	0.163457	99.17	97.60
7	0.061105	0.057468	0.0242	0.001391	0.058859	96.18	94.05
8	0.062323	0.05723	0.0242	0.001385	0.058615	93.67	91.83
9	0.060762	0.057195	0.0255	0.001458	0.058653	96.40	94.13
10	0.059261	0.057041	0.0255	0.001455	0.058496	98.69	96.25
11	0.016004	0.017412	-0.0669	-0.00116	0.016247	98.51	91.21
12	0.014691	0.017406	-0.0669	-0.00116	0.016242	90.45	81.52
13	0.01852	0.022688	-0.0643	-0.00146	0.021229	87.24	77.49
14	0.017483	0.02259	-0.0643	-0.00145	0.021137	82.71	70.79
15	0.017626	0.022703	-0.0643	-0.00146	0.021243	82.97	71.19
16	0.035027	0.035444	-0.037	-0.00131	0.034133	97.38	98.81
17	0.041794	0.042314	-0.0482	-0.00204	0.040274	96.23	98.75
18	0.028073	0.020891	0.2752	0.005749	0.02664	94.62	74.42
19	0.025500	0.020902	0.2752	0.005752	0.026654	95.67	81.97
20	0.055700	0.052274	0.1503	0.007857	0.060131	92.63	93.85

Confirmation tests are also performed for the given regression equation and the proposed RBFEA. The data used for confirmation, confirmation results and the performance comparisons are given in Section 3.4.

3.4 Confirmation Tests and Performance Comparison

In this section confirmation tests are performed for both the conventional regression analysis and the RBFEA. For this purpose, the test data presented in Table 6 are used. These samples were not previously used for regression analysis, ANN training or RBFEA predictions.

Table 6. List of test data for confirmation

Sample Number	Power (kVa)	Max. Voltage	NLTC/ OLTC	CRGOSS (Kg)	Copper (Kg)	St 37 – Steel (Kg)	Oil (Kg)	Observed value
i	$X1i$	$X2i$	$X3i$	$X4i$	$X5i$	$X6i$	$X7i$	Y_i
21	0.01	0.132	0.1	0.01399	0.007628	0.028648	0.0388	0.03524
22	0.08	0.08	0.1	0.07684	0.073264	0.103600	0.1288	0.11720
23	0.125	0.38	0.1	0.12896	0.084288	0.148068	0.1876	0.11863
24	0.266	0.482	0.1	0.20545	0.112548	0.182487	0.2787	0.13647
25	0.196	0.380	0.2	0.19981	0.076541	0.196354	0.2438	0.15998

By using the input factor values given in Table 6 and the regression equation given in Equation (6), the expected values of regression analysis are fitted and given in the third column of Table 7. To calculate the flow time estimation for *RBFEA*, the $PE(pred)_i$, $\hat{\epsilon}_i$, \hat{Y}_i (adj) values are calculated and by using these values the flow time estimation for *RBFEA* is performed (given in the sixth column). Also, the accuracy of predictions for two approaches is given in the last two columns.

Table 7. Confirmation tests

Sample Number	Observed Value	Expected (fitted) Value by Regression Equation	Predicted Prediction Error Rate (ANN)	Predicted Residuals	Expected (fitted) Value by <i>RBFEA</i>	Accuracy For <i>RBFEA</i> Estimation (%)	Accuracy For Regression Estimation (%)
i	Y_i	\hat{Y}_i	$PE(pred)_i$	$\hat{\epsilon}_i = (PE(pred)_i)(\hat{Y}_i)$	$\hat{Y}_i(adj) = \hat{Y}_i + \hat{\epsilon}_i$		
21	0.03524	0.0327	0.0698	0.002282	0.034982	99.26	92.23
22	0.11720	0.0923	0.1649	0.01522	0.10752	90.99	73.02
23	0.11863	0.1413	-0.0550	-0.00777	0.133529	88.84	83.96
24	0.13647	0.1334	0.0350	0.00467	0.138069	98.84	97.69
25	0.15998	0.1504	0.0587	0.00883	0.159228	99.53	93.63

The results of confirmation tests indicate that the predictions performed by *RBFEA* ($\hat{Y}_i(adj)$) are closer to the real observations when it is compared with the results of conventional regression analysis. According to these results, it can be concluded that the *RBFEA* increases the prediction performance of conventional regression analysis and can be used effectively for these types of problems. It is well known that for large samples ANN gives more accurate predictions. So, the performance of *RBFEA* can be increased by increasing the number of samples.

4. RESULTS AND DISCUSSIONS

The accurate prediction capability of *RBFEA* is high and the prediction errors for the test data set are acceptable according to the results presented in Table 7. The accuracy of predictions performed by *RBFEA* for the validation data ranges between 88.84 and 99.53%, which is given in Table 7. The variations are resulted from the labor-intensive work, material specifications, moving and waiting times for each operation.

When the literature is reviewed, the studies on flow time estimation, due date estimation, determining the completion time etc. are used as the general job and job shop characteristics. In these presented studies, the processing times or their probability distributions are thought to be known before the production starts. But in this study, both processing times and the flow times are generally not known before the production and flow times were able to be predicted accurately by discussing technical specifications demanded by the customer, before starting

production. Such a study that uses technical specifications to predict the flow times of a product that was not produced previously does not exist in the literature.

In the present paper, a new flow time estimation algorithm is proposed. It is observed that the proposed algorithm produces acceptable results for the test data set which is not used in the modeling or training steps previously. By using proposed algorithm called *RBFEA*, the decrease in the flow time prediction error and the increase in the competitive power are provided. Furthermore, it is observed that the generalization capability of *RBFEA* is impressive when the performance of *RBFEA* is compared with the regression model calculated by least square estimation. This is because the effect of unknown factors those cause high residuals can be added to the regression equations by using *RBFEA*.

5. CONCLUSIONS

Technical improvements, innovations and R&D (research and development) activities may provide firms to compete in the market. However, as in the presented case sector, calculating the production times and labor costs accurately are mandatory for competition and to survive. In this study, a new residual based flow time estimation algorithm - that uses technical specifications demanded by the customer - is proposed to predict the flow time (and to predict the labor cost by this way). Using the technical specifications directly to predict the flow time is novel according to the literature review presented in this study. Using the technical specifications of customer orders directly provides accurate flow time prediction for the regression equations. Although the technical specifications that are not considered in the regression equation increase the lack-of-fit, it is impossible to consider all of the technical specifications. For example, in the case study of this paper, according to the power and voltage combinations of transformers; the length, weight and many other specifications of the product vary and this causes differences between the expected and observed flow times. It is impossible to consider all these details before the production. Therefore, a new algorithm is proposed to adjust the predictions of the fitted model by discussing the technical specifications of customer orders. The proposed algorithm uses the regression equation to predict the flow time and adopts feed forward back propagation artificial neural network (FF-BPN) to calculate the residual error sourced from uncontrollable production parameters. It combines the results of the regression equation and estimated residual term to decrease the lack-of-fit and provide the accurate estimation.

The performance evaluation is performed by comparing the performances of conventional regression analysis and *RBFEA*. The results showed that the generalizing and accurate prediction capability of *RBFEA* is better for this type of problems when it is compared with the conventional mathematical modeling. By using *RBFEA*, the selected case company was able to give accurate price offer by considering the customer demands. This brings the decrease in the lost sales sourced from the high price offers, which also ensures the companies' profitableness avoiding working at a loss. The author aims to investigate the performance of the proposed *RBFEA* method by discussing different regression models in the future study.

ACKNOWLEDGEMENT

The author would gratefully like to thank, the R&D Department of Best Transformers Company for its great support and also to Mr. Omur Karademir and Dr. Ibrahim Kucukkoc whose valuable supports lead to revealing this paper. The author also acknowledges the valuable suggestions of the editor and the anonymous referees which lead to an improved organization of this paper.

REFERENCES

- [1] Ragatz GL, Mabert VA. A Simulation Analysis of Due Date Assignment Rules, *Journal of Operations Management*, Cilt. 5, 1984, s. 27–39.
- [2] Vig MM, Dooley KJ. Mixing Static and Dynamic Flow Time Estimates for Due-Date Assignment, *Journal of Operations Management*, Cilt. 11, 1993, s. 67–79.
- [3] Arizono I, Yamamoto A, Ohta H. Scheduling for Minimizing Total Actual Flow Time by Neural Networks, *International Journal of Production Research*, Cilt. 30, 1992, s. 503–511.
- [4] Enns ST. A Dynamic Forecasting Model for Job Shop Flowtime Prediction and Tardiness Control, *International Journal of Production Research*, Cilt. 33, No. 5, 1995, s. 1295–1312.
- [5] Harris CR. Modelling the Impact of Design, Tactical, and Operational Factors on Manufacturing System Performance, *International Journal of Production Research*, Cilt. 35, No. 2, 1997, s. 479-499.
- [6] Veral EA. Computer Simulation of Due-Date Setting in Multi-Machine Job Shops, *Computers & Industrial Engineering*, Cilt. 41, No. 1, 2001, s. 77-94.
- [7] Govind N, Roeder TM. Estimating Expected Completion Times with Probabilistic Job Routing, *Proceedings of the 2006 Winter Simulation Conference*, Cilt. 1, No. 5, Winter Simulation Conference, Monterey, CA, Dec 03-06, 2006, s. 1804-1810.
- [8] Li S, Li Y, Liu Y et al. A GA-Based NN Approach for Makespan Estimation, *Applied Mathematics and Computation*, Cilt. 185, No. 2, 2007, s. 1003-1014.
- [9] Alenezi A., Moses SA, Trafalis TB. Real-Time Prediction of Order Flow Times Using Support Vector Regression, *Computers & Operations Research*, Cilt. 35, No. 11, 2008, s. 3489-3503.
- [10] Asadzadeh SM, Azadeh A, Ziaefar AA. Neuro-Fuzzy-Regression Algorithm for Improved Prediction of Manufacturing Lead Time with Machine Breakdowns, *Concurrent Engineering-Research and Applications*, Cilt. 19, No. 4, 2011, s. 269-281.
- [11] Chen T, and Wang YC. An Iterative Procedure for Optimizing the Performance of the Fuzzy-Neural Job Cycle Time Estimation Approach in a Wafer Fabrication Factory, *Mathematical Problems in Engineering*, Article 2013, Article Number: 740478.
- [12] Kumru M, Kumru PY. Using Artificial Neural Networks to Forecast Operation Times in Metal Industry, *International Journal of Computer Integrated Manufacturing*, Cilt. 27, No. 1, 2014, s. 48-59.
- [13] Pan Q, and Dong Y. An improved migrating birds optimisation for a hybrid flowshop scheduling with total flowtime minimisation, *Information Sciences*, Cilt. 277, 2014, s. 643-655.
- [14] Ribas I, Companys R, Tort-Martorell X. An efficient Discrete Artificial Bee Colony algorithm for the blocking flow shop problem with total flowtime minimization, *Expert Systems with Applications*, Cilt. 42, No. 15-16, 2015, s. 6155-6167

- [15] Harlow JH. *Electric Power Transformer Engineering*, 3rd Edn. CRS Press, Taylor & Francis Group, USA, 2012.
- [16] Karaoglan AD, Karademir O. Flow time and product cost estimation by using an artificial neural network (ANN): A case study for transformer orders, *The Engineering Economist*, 2016, (INPRESS). (DOI:10.1080/0013791X.2016.1185808)
- [17] Montgomery DC. *Design and Analysis of Experiments*, 5th Edn. John Wiley & Sons, Inc., New York, 2001.
- [18] Giesbrecht FG, Gumpertz ML. *Planning, Construction, and Statistical Analysis of Comparative Experiments*. John Wiley & Sons, Inc., New Jersey, 2004.
- [19] Mason RL, Gunst RF, Hess JL. *Statistical Design and Analysis of Experiments*, 2nd Edn. John Wiley & Sons, Inc., New Jersey, 2003.
- [20] Castillo ED. *Process Optimization - A Statistical Approach*. Springer, New York, 2007.
- [21] Ham FM, Kostanic I. *Principles of Neurocomputing for Science and Engineering*, 1st Edn. McGraw-Hill, New York, 2001.
- [22] Karaoglan AD. An Integrated Neural Network Structure for Recognizing Autocorrelated and Trending Processes, *Mathematical & Computational Applications*, Cilt. 16, No. 2, 2011, s.514-523.

CV / ÖZGEÇMİŞ

Aslan Deniz KARAOGLAN; Assistant professor (Yrd.Doç.Dr.)

He got his bachelors' degree in the Industrial Engineering Department at Gazi University, Ankara/Turkey in 2001; his master degree in the Industrial Engineering Department at Balıkesir University, Balıkesir/Turkey in 2006; PhD degree in the Industrial Engineering Department at Dokuz Eylül University, Izmir/Turkey in 2010. He is an academic member of the Industrial Engineering Department at Balıkesir University. His major areas of interest are: Design of Experiments, Statistical Process Control, Artificial Neural Networks, Optimization, and System Analysis.

Lisans derecesini 2001'de Gazi Üniversitesi Endüstri Mühendisliği Bölümünden, Ankara/Türkiye; yüksek lisans derecesini 2006'da Balıkesir Üniversitesi Endüstri Mühendisliği Bölümünden, Balıkesir/Türkiye; Doktora derecesini 2010'da Dokuz Eylül Üniversitesi Endüstri Mühendisliği Bölümünden aldı. Halen Balıkesir Üniversitesi Endüstri Mühendisliği Bölümünde öğretim üyesi olarak görev yapmaktadır. Ana ilgi alanları: Deney Tasarımı, İstatistiksel Süreç Kontrol, Yapay Sinir Ağları, Optimizasyon, ve Sistem Analizidir.