PRINTED ISSN: 1308-9080 / OLINE ISSN: 1308-9099

Volume: 15 / Number: 1 / Year: 2020

TURKISH JOURNAL OF SCIENCE & TECHNOLOGY

TURKISH JOURNAL OF SCIENCE AND TECHNOLOGY (TJST)

Year: 2020 Vol: 15 Number: 1

Address:
Fırat Universitesi
Fen Bilimleri Enstitüsü
23119, Elazig - TURKEY

Tel: 0 424 212 27 07
Fax: 0 424 236 99 55
e-mail: fenbilimdergi@firat.edu.tr

New ISSN
Online: 1308-9099
Printed: 1308-9080

Old ISSN
Online: 1306 – 8555
Printed: 1306 – 8547

Refereed journal. Published twice a year

http://dergipark.gov.tr/tjst

TURKISH JOURNAL OF SCIENCE & TECHNOLOGY (TJST)
Published by Firat University

Owner

Prof Dr. Kutbeddin DEMİRDAĞ
Fırat Üniversitesi Rektörü

Responsible Director

Prof. Dr. Soner ÖZGEN
F.Ü. Fen Bilimleri Enstitüsü Müdürü

Editor in Chef
Assoc.Prof. Dr. Erkan TANYILDIZI
Fırat University, Technology Faculty
Department of Software Engineering

Editor
Assis. Prof. Dr. Sencer ÜNAL
Fırat University, Engineering Faculty
Department of Electrical-Electronics Engineering

ADVISORY BOARD

Fırat University Turkish Journal of Science & Technology (TJST)
15-1, 2020

CONTENTS / İÇİNDEKİLER

Turkish Journal of Science & Technology
15(1), 1-12, 2020

An Alternative Approach to Variable Selection Using Hybrid Regression Modeling in Undersized Sample Data

Esra PAMUKÇU*
Department of Statistics, Faculty of Science, Fırat University, Elazığ, Turkey
epamukcu@firat.edu.tr

(Geliş/Received: 11./12/2019; Kabul/Accepted: 09/02/2020)

Abstract: The problems encountered in the analysis of data sets with undersized sample mainly arise from the singular covariance structure. As a solution to this problem, non-singular Hybrid Covariance Estimators (HCEs) have been proposed in the literature. Several multivariate statistical techniques where HCEs are used continue to be developed and introduced. One of these is the Hybrid Regression Model (HRM). Thanks to HCEs, since there is no longer the rank problem in covariance matrix, in HRM analysis the regression coefficients can be estimated as many as the number of variables. However, determining the best predictors in regression model is one of the biggest problems for researchers since the number of variables increases and there is insufficient knowledge about the model. Therefore, some numerical optimization techniques and strategies are required to explain such a wide solution space where the number of alternative subsets of candidate models of predictors can reach millions. In this paper, we introduced a new and alternative approach to variable selection for undersized sample data by using the Genetic Algorithm (GA) and Information Complexity Criteria (ICOMP) as a fitness function in the HRM analysis. To demonstrate the ability of proposed method, we carried out the Monte Carlo simulation study with correlated and undersized data sets. We compared our method with Elastic Net (EN) modeling. According to results, the proposed method can be recommended as an alternative approach to select variable in undersized sample data.

Key words: Genetic algorithm, hybrid regression model, information complexity criteria, variable selection, undersized sample problem

Cılız Örneklem Problemine Sahip Veri Setlerinde Hibrit Regresyon Modellemesi Kullanarak Değişken Seçimine Alternatif Bir Yaklaşım

Öz: Cılız örneklem problemine sahip veri setlerinin analizinde karşılaşılan problemler temel olarak singüler kovaryans matrisinden kaynaklanmaktadır. Bu probleme bir çözüm olarak literatürde Hibrit Kovaryans Tahmin Edicileri (HCE) önerilmiştir. HCE'lerin kullanıldığı bazı çok değişkenli istatistiksel yöntemler geliştirilmeye ve tanıtılmaya devam etmektedir. Bunlardan biri Hibrit Regresyon Modeli'dir (HRM). HCE sayesinde kovaryans matrisinde artık singülerlik problemi olmadığı için, HRM ile değişken sayısı kadar regresyon katsayısı tahmin edilebilir. Bununla birlikte, değişken sayısı çok fazla olduğu ve model hakkında yetersiz bilgiye sahip olunduğu için, regresyon modelindeki en iyi tahmin edicileri belirlemek araştırmacılar için en büyük problemlerden biridir. Bu nedenle tahmin edicilerin alternatif modellerinin alt küme sayısının milyonları bulabildiği böyle geniş bir çözüm uzayını açıklamak için bazı nümerik optimizasyon tekniklerine ve stratejilerine ihtiyaç vardır. Bu çalışmada, ICOMP'ın uygunluk fonksiyonu olarak kullanıldığı bir Genetik Algoritma yapısı ile HRM analizi yapılarak cılız örneklemli veri setleri için değişken seçimine alternatif bir yaklaşım önerilmiştir. Önerilen yaklaşımın kullanılabilirliğini göstermek için korelasyonlu ve cılız örneklemli veri setlerinin kullanıldığı bir Monte Carlo simulasyon çalışması yapılmıştır. Karşılaştırma amacıyla Elastik Net modellemesi kullanılmıştır. Elde edilen sonuçlara göre, önerilen yaklaşımın cılız örneklemli veri setlerinde değişken seçimi için alternatif bir yaklaşım olarak kullanılabileceği söylenebilir.

Anahtar kelimeler: Genetik algoritma, hibrit regresyon modeli, bilgi karmaşıklığı kriterleri, değişken seçimi, cılız örneklem problemi

1. Introduction

According to statistical viewpoint, if inferences about data are required, it is expected that the number of samples should increase exponentially against the number of variables. Nowadays, even if there are many observations, the number of variables may increase radically. In this case, a single observation can have thousands or even millions of dimensions whereas the number of observations that can be reached for the study is expressed as ten or hundred. The traditional techniques in statistics are not capable for analyzing such data sets [1].

Statisticians sometimes say "Big p, Small n" for this problem. Another definition is "the undersized sample problem" or "extremely small sample problem" [2,3].

* Corresponding author: epamukcu@firat.edu.tr ORCID: 0000-0002-5778-9626

The problems encountered in the analysis of data sets with undersized sample mainly arise from the singular covariance structure. As a solution of this problem, for the first time by [4], non-singular Hybrid Covariance Estimators (HCEs) is proposed by hybridizing the Maximum Likelihood Estimator (MLE) with smoothed covariance structures after the stabilization stage of eigenvalues.

Since HCEs overcome the singularity in covariance matrix, the analysis of undersized or high-dimensional data sets with multivariate statistical methods for $\boldsymbol{n} << \boldsymbol{p}$ problem has become possible. Several multivariate statistical techniques in where HCEs are used continue to be developed and introduced [4,5,6]. One of these is the Hybrid Regression Model (HRM) in which HCEs is used as covariance input [7]. When the HRM is performed in case $\boldsymbol{n} << \boldsymbol{p}$, the regression coefficients can be obtained as many as the number of variables even if they are hundreds or thousands. The next stage is to determine the best predictors that have the most effect on the response variable, i.e. the variable selection stage. This is one of the biggest problems for researchers when the number of predictors increases and there is insufficient or no prior knowledge about the model.

Model selection is basically a process of finding the best model from the subset of competing or candidate models by revealing which variables are effective on the response variable. Since the early 1970s, it has been possible to come across many studies on the model selection algorithm and criteria. These include classical methods for the model selection and the methods based on information criteria. The classical methods are generally performed by hypothesis testing. An arbitrary level of confidence is selected in a hypothesis testing process by the researcher to decide whether the variables are included in final model or not. However, most statisticians and other scientists have emphasized that the level of confidence used in the selection of classical models is baseless [8]. Some scientists have stated that the hypothesis testing approach does not have a theoretical accuracy and it is generally insufficiently valid [9]. Although almost all popular statistical packages have many classical model selection procedures based on hypothesis testing, such as Forward Selection, Backward Elimination, and Stepwise techniques, these selection methods do not deal with the dependency structure between variables. Also, since they contain the limitations of the hypothesis testing, these methods are criticized by Boyce et al. (1974, p:16) with the following words "These approaches do not guarantee optimal results". Therefore, the selection of the best predictors involves randomness. He mentioned that also "An exhaustive search would examine $\boldsymbol{2^p}$ possible equations" [10].

The shortcomings in the classical procedures for model selection impose limitations on the selection of the best or near-best model subsets. Although some statisticians and researchers propose to choose from all possible subsets, in many cases this method is not producible in terms of calculation and is also not possible in terms of time and cost [11]. For example, if we have p=20, the number of the subsets of the candidate models are $\boldsymbol{2^{20} = 1048576}$. Therefore, some numerical optimization techniques and strategies are required to explain such a wide solution space.

In general, two components are needed to use numerical techniques in a subset selection problem.

- An algorithm to effectively scan wide solution space
- A measure for comparison of candidate models

In this study, for the HRM, it is constructed a Genetic Algorithm (GA) structure where the variables in the model assigned as '1' and the others variables assigned as '0'. By using ICOMP criterion as the fitness function, it is compared the candidate regression models in population for transferring to the next generation.

The sketch of paper is given by following orders. In Section 2.1, HRM and its background will be introduced. In Section 2.2, the GA structure will be presented in order to explain how to determine the variables exist in the model by using the GA in the HRM. In Section 2.3, we briefly introduce the Elastic Net (EN) modeling. In Section 3, we provide our Monte Carlo simulation study. The last part is divided to conclusion and brief discussion.

2. Material and Methods

2.1. Hybrid regression model (HRM) with information complexity and hybrid covariance estimators

Let us consider the multiple linear regression model in matrix form given by

$$y = X\beta + \varepsilon \tag{1}$$

where y_{nx1} is vector of observaitons on a response variable, X_{nxp} is full rank matrix of variable, β_{px1} is regression coefficient vector and $\varepsilon_{nx1} \sim N(0, \sigma^2)$ is random error vector. The maximum likelihood estimates of β_{px1} are given by

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'y \tag{2}$$

As is seen in equation (2), the Gram matrix $(X'X)$ must be non-singular and invertible in order to obtain the estimate of $\hat{\beta}_{MLE}$. In case of undersized sample, i.e. $n \ll p$, it is clear that the X_{nxp} is not full rank matrix and the Gram matrix is not invertible. As a solution to this problem, in the framework of the regression analysis it has been proposed Hybrid Covariance Estimators (HCEs) as a well-conditioned and non singular covariance estimate instead of using Gram matrix [4]. In the following section, HCEs will be introduced.

2.1.1. Smoothed covariance structures

For a covariance matrix, the Condition Number (CN) defined as the largest eigenvalue λ_{max} divided by the smallest eigenvalue λ_{min} is given by

$$CN = \lambda_{max}/\lambda_{min} \tag{3}$$

The inverse of CN can be used for the definition of singularity of covariance matrix [12] and it is defined as follow,

$$\kappa(\Sigma) = \frac{1}{CN} \tag{4}$$

If $\kappa(\Sigma)$ is near to zero, the covariance matrix is close to the singularity. As a solution to singularity, shrinkage estimators that will shrunk eigenvalues of Σ to a central value have been developed. The main idea of these estimators is to take the convex combination of the maximum likelihood estimation of the sample covariance, i.e. $\hat{\Sigma}_{MLE}$, and a target diagonal matrix $\widehat{D}$ (i.e., taking the weighted average). Then, the shrinkage or smoothed covariance estimator is given by

$$\hat{\Sigma}_S = (1 - \hat{\rho})\hat{\Sigma}_{MLE} + \hat{\rho}\widehat{D} \tag{5}$$

where, $\hat{\rho} \in [0,1]$ is estimate of the optimal shrinkage coefficient ρ, The $\widehat{D}$ matrix is called as shrinkage target matrix and its naive form is as follow,

$$\widehat{D} = \frac{tr(\hat{\Sigma}_{MLE})}{p} I_p = \left(\frac{1}{p}\sum_{j=1}^{p}\lambda_j\right) I_p = \bar{\lambda} I_p \tag{6}$$

where $tr(.)$ represents the trace of matrix, $\lambda_j, j = 1,2,\dots,p$ are eigenvalues of sample covariance matrix and $\bar{\lambda} = \sum_{i=1}^{p}\lambda_i/p$ is arithmetic mean of eigenvalues. By using the weighted average in equation (5), it is placed to a lower weight on extremely large or small eigenvalues. Thus, the effect of these eigenvalues is reduced and a smoothed estimator is obtained.

The smoothed or regularized covariance estimators under linear or quadratic loss functions have been introduced in the literature. Some of them: Empirical Bayes Estimator (EB)[13], Stipulated Ridge Estimator (SRE)[14], Stipulated Diagonal Estimator (SDE)[14], Convex Sum Estimator (CSE)[15,16], Bozdogan's Convex Sum Estimator (BCSE)[17], Oracle Approximation (OAS)[18], Ledoit-Wolf Estimator (LW)[19] are given in Table 1.

Table 1. Smoothed or regularized covariance estimators

$\hat{\Sigma}_S$	**Form***	$\hat{\boldsymbol{\rho}}$
$\hat{\Sigma}_{EB}$	$\hat{\Sigma}_{MLE} + \hat{\rho}_{EB} D_p$	$\frac{p-1}{n(\hat{\Sigma}_{MLE})}$
$\hat{\Sigma}_{SRE}$	$\hat{\Sigma}_{MLE} + \hat{\rho}_{SRE} D_p$	$p(p-1)\left[2ntr(\hat{\Sigma}_{MLE})\right]^{-1}$
$\hat{\Sigma}_{SDE}$	$(1-\hat{\rho}_{SDE})\hat{\Sigma}_{MLE} + \hat{\rho}_{SDE} D_p$	$p(p-1)\left[2ntr\left(\hat{\Sigma}_{MLE}{}^{-1}\right) - p\right]^{-1}$
$\hat{\Sigma}_{CSE}$	$\hat{\rho}_{CSE}\hat{\Sigma}_{MLE} + (1-\hat{\rho}_{CSE}) D_p$	$\frac{n}{n+m}$ where $0 < m < \frac{2[p(1+\beta)-2]}{p-\beta}$, $\beta = \frac{tr(\hat{\Sigma})^2}{tr(\hat{\Sigma}^2)}$ for $p \geq 2$ dimensions.
$\hat{\Sigma}_{BCSE}$	$\hat{\rho}_{BCSE}\hat{\Sigma}_{MLE} + (1-\hat{\rho}_{BCSE}) D_p$	$\frac{1}{\alpha}$, where $\alpha = \frac{1}{n-1}\sum_{j=1}^{p} \sigma_{jj}$
$\hat{\Sigma}_{OAS}$	$(1-\hat{\rho}_{OAS})\hat{\Sigma}_{MLE} + \hat{\rho}_{OAS} D_p$	$min\left(\frac{\left(1-\frac{2}{p}\right) tr(\hat{\Sigma}^2) + (tr\hat{\Sigma})^2}{(n+1-\frac{2}{p})\left[tr(\hat{\Sigma}^2) - \frac{(tr\hat{\Sigma})^2}{p}\right]}, 1\right)$
$\hat{\Sigma}_{LW}$	$(1-\hat{\rho}_{LW})\hat{\Sigma}_{MLE} + \hat{\rho}_{LW} D_p$	$min\left(\frac{\sum_{i=1}^{n} \|x_i{}' x_i - S\|^2{}_F}{n^2\left[tr(\hat{\Sigma}^2) - \frac{(tr\hat{\Sigma})^2}{p}\right]}, 1\right)$

*: n is number of observations, p is number of variables in data, D_p is target matrix

2.1.2. Hybrid covariance estimators (HCEs)

Since the problem of interest of this study is able to perform regression analysis in data sets with undersized sample, it is clear that the maximum likelihood estimates of covariance matrices of these data sets is singular and/or ill conditioned. It is important to obtain a well-conditioned and non-singular covariance structure. From this point, Pamukcu et al. (2015) [4] developed Hybrid Covariance Estimators-HCEs by using following eigenvalue stabilization algorithm defined by Thomaz [20]

- Step-1: Calculate the eigenvalues λ_j and their eigenvectors v_j of $\hat{\Sigma}_{MLE}$, where $j = 1,2,\dots,p$ and p is the number of variables of the data. $\hat{\Sigma}_{MLE}$ represents the maximum likelihood covariance estimator.
- Step-2: Calculate the arithmetic mean of eigenvalues by using: $\bar{\lambda} = \frac{1}{p}\sum_{j=1}^{p} \lambda_j$
- Step-3: Produce the following matrix of eigenvalues based on largest dispersion values:

$$\Lambda^* = \begin{bmatrix} max(\lambda_1, \bar{\lambda}) & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & max(\lambda_p, \bar{\lambda}) \end{bmatrix} \quad (7)$$

- Step-4: The new stabilized covariance matrix is given by:

$$\hat{\Sigma}_{MLE\_STA} = V\Lambda^* V \quad (8)$$

where V is matrix of eigenvectors v_j of $\hat{\Sigma}_{MLE}$. As is seen, the algorithm stabilizes eigenvalues by expanding only the smaller and consequently less reliable eigenvalues of the covariance matrix and by keeping its larger eigenvalues unchanged. Then, one can obtain HCEs by following two stage process:

- Stage-1: Perform the stabilization algorithm above
- Stage-2: Produce the stabilized-smoothed covariance estimators in Table 1.

For example, $\hat{\Sigma}_{HCE} = \hat{\Sigma}_{MLE\_STA\_BCSE}$ is defined as follow

$$\hat{\Sigma}_{MLE\_STA\_BCSE} = (1 - \hat{\rho}_{BCSE})\hat{\Sigma}_{MLE\_STA} + \hat{\rho}_{BCSE} D_p \tag{9}$$

For more on this we refer the readers to [4,21].

The logical and mathematical theme of using Stabilization + Regularization is to achieve positive definition with respect to shrunk and to expand the less appropriate and smaller eigenvalues by using stabilizing. Considering Table 1, there are several forms of HCEs which can be used. It is an important problem which one is optimal covariance for a researcher who wants to use them. For this reason, it is investigated for the choosing optimal covariance structure by [4,5,6,21] especially in the framework of regression analysis and observed that $\hat{\Sigma}_{MLE\_STA\_CSE}$ and $\hat{\Sigma}_{MLE\_STA\_BCSE}$ have superior performance on the others. Therefore, in our simulation study, we use $\hat{\Sigma}_{MLE\_STA\_CSE}$ and $\hat{\Sigma}_{MLE\_STA\_BCSE}$ covariance structures and their model are represented by HRM1 and HRM2, respectively.

2.1.3. Information complexity criteria (ICOMP) for model selection

In the literature, Akaike [22,23] Information Criterion (AIC) is widely used for statistical model selection. This is given by

$$AIC = -2logL(\theta) + 2k \tag{10}$$

where $logL(\theta)$ is log likelihood function of θ parameter in a probability density function, k is the number of parameters. ICOMP (I; Information-COMP; Complexity) are criteria developed by Bozdogan (1988) for the model selection in multivariate linear and nonlinear models [24]. Wheras the AIC is only intended to strike a balance between the lack of fit and the penalty terms, ICOMP aims to establish this balance by taking into account a complexity measure that measures how the parameters in the model relate to each other. Therefore, instead of directly penalizing the number of parameters, it penalizes the covariance complexity of the model introduced by [24]. ICOMP is given by

$$ICOMP = -2logL(\theta) + 2C_1(\Sigma) \tag{11}$$

The second part of equation 11 is called the measure of complexity of model. It is given as follow:

$$C_1(\Sigma) = \frac{p}{2} log\left[\frac{tr(\Sigma)}{p}\right] - \frac{1}{2} log|\Sigma| \tag{12}$$

where $|\Sigma|$ represents determinant of Σ and p is dimension of Σ. As seen, $C_1(\Sigma)$ include two simplest scales of multivariate scattering called determinant and trace in a single function. There are several forms ICOMP criteria, for more about this we refer the reader to [11,25,26,27]. For these criteria, the model which has minimum value of criteria is called as the best model.

2.1.4 Hybrid regression model

In the case of undersized sample, i.e. $n << p$, where the sample variance covariance matrix is singular, we can analyze the data with HRM by using following steps:

- Step-1: $\hat{\Sigma}_{HCEs} = \hat{\Sigma}_{MLE\_STA\_CSE}$ and $\hat{\Sigma}_{HCEs} = \hat{\Sigma}_{MLE\_STA\_BCSE}$ are estimated for the data set
- Step-2: $\hat{\Sigma}_{HCEs}$ are used instead of Gram matrix in multiple regression analysis
- Step-3: The model which has minimum value of information criteria is determined as the best model among the candidate models created with different $\hat{\Sigma}_{HCEs}$.

Indeed, AIC, BIC [28] and Consistent Akaike Information Criterion (CAIC) [29] may be used as information criteria for model selection tool. Specifically, we prefer to use ICOMP in our computations since it is demonstrated

that ICOMP has superior performance on the other criteria. Also, it has been demonstrated that they tend to select more variables that may be related to each other since their penalty terms are not deal with correlation structure between variables [11,17,21,26,27]. The derived form of ICOMP in HRM is defined as follows:

$$ICOMP_{Miss}(HRM) = nlog(2\pi) + nlog(\hat{\sigma}^2) + n + 2C_1\left(Cov(\hat{\beta}_{HRM})_{Missspec}\right) \tag{13}$$

where $Cov(\hat{\beta}_{HRM})_{Missspec} = \hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\hat{\mathcal{F}}^{-1}$is called the "sandwiched covariance" estimator. $\hat{\mathcal{F}}^{-1}$ and $\hat{\mathcal{R}}$ represent the inverse of Fisher information matrix and outer-product form of Fisher information matrix, respectively. These are given by

$$Cov(\hat{\beta}_{HRM})_{Missspec} = \begin{bmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}\begin{bmatrix} \frac{n}{\hat{\sigma}^2} & \frac{nS_k}{2\hat{\sigma}^3} \\ \frac{nS_k}{2\hat{\sigma}^3} & \frac{n(K_t-1)}{4\hat{\sigma}^4} \end{bmatrix}\begin{bmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \tag{14}$$

is called Sandwich covariance matrix and where

$$S_k = Skewness\ coefficient = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^3\right)}{\hat{\sigma}^3} \tag{15}$$

$$K_t = Kurtosis\ coefficient = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4\right)}{\hat{\sigma}^4} \tag{16}$$

where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$and n is the number of observations. When the model is correct we get $\hat{\mathcal{F}}^{-1} = \hat{\mathcal{R}}$ and the formula reduces to the usual inverse Fisher information matrix

2.2. Genetic algorithm for hybrid regression model

After performing HRM to undersized sample data, since the regression coefficients are obtained as many as the number of variables, many of these could be irrelevant or redundant variables. So, the performance of the HRM can be evaluated after they are detected and sorted.

Let p be the number of variables in HRM. If the p increases, the number of subsets of candidate models can reach millions or even billions. Therefore, an effective algorithm is necessary to effectively scan the such a wide solution space. For this purpose, we prefer to use the principles of GA in order to select best predictors in HRM.

The GA is an evolutionary search algorithm that borrows concepts from biological evolution and is a stochastic optimization method inspired by the principles of evolution in nature. The search method is based on the principle of survival of the best. To this end, we begin to work with a community of potential solutions to the problem to be solved. It is called as initial population. Each individual in a population is a potential solution and is coded as chromosomes in accordance with the nature of the problem being studied.

For the variable selection problem in HRM analysis with GA, our implementation basically follows Goldberg's GA (1988) [30]. For a detailed information we refer the readers to [30,31].

Considering in the framework of the regression analysis, the length of the chromosomes is equal to the number of all the variables in data set. Let *k* be the number of the variables. Assuming that there are 10 variables in the data set, i.e. $k = 10$. In this case, a chromosome is a sequence of *k* units. Each unit on the chromosome is called a "gene" and each unit can have a value of '0' or '1'. '0' on the chromosome represents that the corresponding variables are not included in the regression model and '1' represents the included variables in the regression model. A chromosome sample is given below for the regression model in which the variables 1,2,4,5,6,8 and 10 are included [11,27].

Variables	1 2 3 4 5 6 7 8 9 10
Chromosome	1 1 0 1 1 1 0 1 0 1

The general procedure in the GA is straightforward and it is summarized as follow:

1. Generate the initial population of chromosomes
2. Calculate the fitness value for each chromosome in the population: The fitness value of the i[th] member is $f\ (i^{th})$ which is the value of the objective function f at that point [32]. For each chromosome, there is a

numerical fitness value that is proportional to the use or ability of the solution represented by the chromosome. This information guides the selection of more appropriate solutions for each generation. In this study, ICOMP is our fitness function. Firstly, a chromosome will be modeled by using HRM and the performance of the model will be measured by ICOMP. Since the model which has minimum value of ICOMP is defined as probably good model, the chromosomes will be transferred to the next generation.

3. Determine how current population is matched for the next generation: Firstly, it is sorted current population according to the values of fitness function. In our case, the most popular chromosomes which has minimum ICOMP values come to in the beginning of the list. Chromosomes in the list are mated by using a sequential pairs (pair (1,2), pair (3,4)...etc.).
4. Perform the GA operations: By changing the structure of chromosomes with the genetic processes such as mutation and crossover, it is provided to investigate the space of all possible solutions. There are three crossover techniques such as single, multiple and uniform crossover. However, these are related to the processing of the GA and no further details will be given.
5. Pass on offspring to new generation.
6. Loop back to step 2 until stopping criteria met: These steps given above are performed only for one generation. After we reach to a certain number of generations or to optimal value of fitness function, the process is terminated. In our case, there is no an optimal value for the fitness function, i.e. ICOMP, therefore the algorithm continue until the number of generations is reached.

Note that, each iteration in the GA is called as generation. The chromosome with the highest fitness in the final generation is proposed as a solution to the problem. It is expected that this proposed solution is optimal or near to the optimal solution.

2.3 Elastic net modeling

Let us recall the multiple regression model in matrix form given equation (1). The objective function of the vanilla matrix representation of EN is defined by

$$L(\beta, \lambda_1, \lambda_2) = \|y - X\beta\|^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \tag{17}$$

where $\|\beta\|_1 = \sum|\beta|$ is the L_1 norm penalty (the LASSO penalty) and $\|\beta\|_2^2 = \sum\beta^2$ is the L_2 norm penalty (the ridge penalty) To avoid the double shrinkage in equation (17), one can undo the penalty by scaling up the estimates from the vanilla matrix representation of EN. Then an improved estimator is

$$\hat{\beta} = \sqrt{1+\lambda_2}\tilde{\beta} \tag{18}$$

Further, one can show that the improved estimates in matrix form are given by [33]

$$\hat{\beta}_{EN} = arg\min_{\beta}\left[\beta'\left(\frac{X'X+\lambda_2 I}{1+\lambda_2}\right)\beta - 2y'X\beta + \lambda_1\|\beta\|_1\right] \tag{19}$$

It is clear that

$$\left(\frac{X'X+\lambda_2 I}{1+\lambda_2}\right) = \frac{1}{1+\lambda_2}X'X + \frac{\lambda_2}{1+\lambda_2}I \tag{20}$$

We can write the equation (20) equivalently as

$$\left(\frac{X'X+\lambda_2 I}{1+\lambda_2}\right) = (1-\hat{\rho}_{EN})X'X + \hat{\rho}_{EN}I \tag{21}$$

where $\hat{\rho}_{EN} = \frac{\lambda_2}{1+\lambda_2}\ \epsilon\ (0,1)$ is shrinkage coefficient and $X'X$ is the Gram matrix.

3. Results

3.1. Monte Carlo simulation study

In this part, we present a Monte Carlo simulation study on correlated data sets with undersized sample. According to explanations about HCEs performance in Section 2.1.2, we study two different HCEs in HRM to able to select the best predictors by using GA and ICOMP. For comparison, we study EN modeling with different three λ_2 tuning parameters as grid values. These are given in Table 2.

Table 2. Compared Models in Simulation Study

Models	With
HRM1	$\hat{\Sigma}_{MLE\_STA\_CSE}$
HRM2	$\hat{\Sigma}_{MLE\_STA\_BCSE}$
EN1	$\lambda_2 = 0.001$
EN2	$\lambda_2 = 0.01$
EN3	$\lambda_2 = 0.1$

The simulation protocol for generating correlated data sets with undersized sample is below:

- p : Number of variables=30, 40, 50, 60
- n : Number of observations=10, 20, 30
- $X_{(nxp)}$: The data set
- r : Correlation between the correlated variables = fixed as 0.5
- σ : Error variance = fixed as 0.01
- ε : Error

Note that the data sets are derived from the multivariate normal distribution with zero mean and Σ covariance matrix, i.e. $MVN(0, \Sigma)$.For $r = 0.5$ and $p = 5$, an example to calculate the variance-covariance matrix Σ is as follow:

$$\Sigma = \begin{bmatrix} 1 & r & \frac{r}{2} & \frac{r}{4} & \frac{r}{8} \\ r & 1 & r & \frac{r}{2} & \frac{r}{4} \\ \frac{r}{2} & r & 1 & r & \frac{r}{2} \\ \frac{r}{4} & \frac{r}{2} & r & 1 & r \\ \frac{r}{8} & \frac{r}{4} & \frac{r}{2} & r & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.500 & 0.250 & 0.125 & 0.062 \\ 0.500 & 1 & 0.500 & 0.250 & 0.125 \\ 0.250 & 0.500 & 1 & 0.500 & 0.250 \\ 0.125 & 0.250 & 0.500 & 1 & 0.500 \\ 0.062 & 0.125 & 0.250 & 0.500 & 1 \end{bmatrix} \quad (22)$$

As is seen, the covariance structure in equation (22) provides correlation between sequential variables. We generate the data sets with different samples and variables size. After having the data sets, we use following steps in order to show the performance of detection of true variables of proposed method and EN modeling, for comparison.

- We generate the response variable by using $y = 5 + 3X_1 - 2X_2 + 6X_3 + \sigma\varepsilon$. So, the variables $\{X_0, X_1, X_2, X_3\}$ are important variables which have effect on response variable.
- *'elasticnet'* module in MATLABR2015a © is used to perform EN modeling. It can be easily found in https://in.mathworks.com/matlabcentral/mlcdownloads/downloads/submissions/58182/versions/2/previews/elasticnet.m/index.html
- To perform GA with HRM and ICOMP, a graphical user interface GUI is written by the author following the steps in Section 2.2.
- In order to compare the performance of detection of true variables of models, we use the average value of True Positive Rate (TPR) [34] given by

$$TPR = \frac{1}{r}\sum_{q=1}^{r} TPR^{q} \tag{23}$$

where $TPR^{q} = \frac{\#truly\ selected\ variables}{\#total\ variables}$ for the models in q^{th} run and r is replication number.
We note that the TPR value is equal to 1, if the list of total selected variables is $\{X_0, X_1, X_2, X_3\}$, otherwise $TPR\epsilon[0,1)$. If it is added an unimportant variable to model, then the TPR value is equal to 4/5=0.8. Additionally, we use following GA parameters in Table 3 in our simulation study. This process for all data sets with different size is repeated 100 times. The simulation results are given in Table 4:

Table 3. GA parameters

Parameter	**Value**
Generation number	10
Population size	50
Generation seeding	Sorted
Probability of crossover	0.75
Type of crossover	Two-points
Probability of mutation	0.1
Objective functions	ICOMP

Table 4. The Simulation results according to TPR values

Data size	**HRM1**	**HRM2**	**EN1**	**EN2**	**EN3**	**CPU time (sec)**
10x30	0.1723	**0.2643**	0.1664	0.1697	0.1814	42.1
10x40	0.1116	**0.1582**	0.0994	0.1017	0.1373	44.5
10x50	0.0862	0.1283	0.1003	0.0933	**0.1323**	**47.9**
20x30	**0.2748**	0.2054	0.1067	0.1040	0.1575	42.9
20x40	**0.2423**	0.1776	0.1165	0.1151	0.1242	47.1
20x50	0.1333	**0.1478**	0.0895	0.0911	0.0946	48.2
30x40	**0.2078**	0.1712	0.0837	0.0848	0.0875	45.4
30x50	0.1396	**0.1475**	0.0894	0.0861	0.0921	52.5
30x60	0.0330	0.0541	0.0727	**0.0742**	0.0740	52.1

We note that the highest TPR value the models are indicated as bold in the Table 4. According to results, we observe that our proposed method has superior performance against EN modeling. It is clear that the performance of EN modeling depend on the λ_2 tuning parameter. How to choose the λ_2 tuning is an important problem in EN. In practice, it is difficult to assign which value of the λ_2 tuning parameter is appropriate. Often the λ_2 tuning parameter is fixed arbitrarily at the beginning. In the literature cross-validation (CV) (5-fold, or 10-fold) has been used which is a time consuming operation. To avoid time consuming operation, we used some grid values for λ_2 tuning parameter. Otherhand, the computation of the HCEs covariance matrix and to perform HRM analysis with GA is fast in terms of the CPU time and it is not heavy. We should emphasize the values of CPU time in Table 4 are belong to whole simulation study. Compared to EN modeling, which is current and frequently used method, the HRM yields better results and it can be suggested as an alternative approach in variable selection for undersized sample data.

4. Conclusion and Brief Discussion

The variable selection is important as much as modeling. In the literature, GA has been proposed in different ways for variable selection in regression analysis [33-37]. But none of them is related to the solution of the

undersized sample problem, i.e. $n << p$. As mentioned before, if the number of the variable is equal to p, the number of subset of candidate model is $2^p - 1$. For example, if we have $p = 20$, the number of the subset of candidate models are equal to $2^{20} = 1048576$. Therefore it is difficult to determine which model is the best among these models without an effective search tool such as GA.

In this paper it was proposed a new andalternative approach to variable selection for undersized data by using HRM with HCEs and GA. We introduced the GA structure where $ICOMP_{Miss}$ used as a fitness function in order to select the best model from such a wide solution space. To demonsrate the effectiveness and utility of the proposed method, the Monte Carlo simulation study was performed on correlated and undersized data sets. The results were compared with Elastic Net (EN) modeling.

EN is a hybrid regression model between Ridge Regression (RR) and Least Absolute Shrinkage and Selection Operator (LASSO) [38]. It is claimed that it can be used for simultaneous modeling and variable selection in high dimensional or undersized data in [38]. However, in our simulation study ,we observed that EN modeling has poor performance to detect true variables when compared with our method. This has demonstrated the success of HCEs covariance matrices used in HRM analysis.

Nowadays, data sets can have radically increasing dimensions. In order to cope with these data sets, it is necessary dimension reduction or feature selection or to increase the number of samples. It may not always be possible to reach the number of samples required to be able to perform classical statistical methods. Therefore the methods that work well are required even if the sample size is too small. According to the results, it can be recommended that the proposed approach based on HRM and GA can be used in the analysis of these data sets.

Acknowledgments: I would like to thank to referees and the editor for their thoughtful and constructive suggestions that greatly improved the paper. I wish to thank to my supervisor Prof. Dr. Hamparsum Bozdogan from University of Tennessee since he introduced and guided to me in this interesting problem area and continued support.

Abbreviations: The following abbreviations are used throughout the text:

AIC	Akaike Information Criterion
BCSE	Bozdogan's Convex Sum Estimator
BIC	Bayesian Information Criterion
CAIC	Consistent Akaike Information Criterion
CN	Condition Number
CSE	Convex Sum Estimator
EB	Empirical Bayes Estimator
EN	Elastic Net
GA	Genetic Algorithm
HCEs	Hybrid Covariance Estimators
HRM	Hybrid Regression Model
HRM1	HRM model with MLE_STA_CSE covariance matrix
HRM2	HRM model with MLE_STA_BCSE covariance matrix
ICOMP	Information Complexity Criteria
ICOMPMiss	Information Complexity Criterion under misspecification
LASSO	Least Absolute Shrinkage and Selection Operator
LW	Ledoit-Wolf Estimator
MLE	Maximum Likelihood Estimator
MLE_STA_BCSE	Maximum Likelihood Stabilized Bozdogan's Convex Sum Estimator
MLE_STA_CSE	Maximum Likelihood Stabilized Convex Sum Estimator
OAS	Oracle Approximation Estimator
RR	Ridge Regression
SRE	Stipulated Ridge Estimator
SDE	Stipulated Diagonal Estimator
TPR	True Positive Rate

References

[1] Donoho, DL. Available online: statweb.stanford.edu/~donoho/Lectures/AMS2000/Curses.pdf. 2 (accessed on 20 April 2019)

[2] Cunningham P. Dimension Reduction. In: Machine Learning Techniques for Multimedia. Cord M, Cunningham P. Eds, Cognitive Technologies. Springer, Berlin, Heidelberg, 2008.

[3] Fiebig DG. On the maximum entropy approach to undersized samples. Appl Math Comput 1984; 14:301-312

[4] Pamukçu E, Bozdogan H, Çalık S. A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification. Comput Math Method M 2015; Article ID 370640: 1-14

[5] Bozdogan H, Pamukçu E. Novel Dimension Reduction Techniques for High-Dimensional Data Using Information Complexity. In Optimization Challenges in Complex, Networked and Risky Systems INFORMS 2016. p:140-170

[6] Mohebbi S, Pamukcu E, Bozdogan H. A new data adaptive elastic net predictive model using hybridized smoothed covariance estimators with information complexity. J Stat Comput Sim 2019: 89(6); 1060-1089.

[7] Pamukcu E. A new hybrid regression model for undersized sample problem. Celal Bayar University Journal of Science 2017; 13(3): 803-813

[8] Linhart H, Zucchini W. Finite sample selection criteria for multinomial models. Statistiche Hefte 1986; 27: 173-178

[9] Burnham KP, Anderson DR. Kullback-Leibler information as a basis for strong inference in ecological studies. Wildlife Res. 2001; 28: 111-119

[10] Boyce DE, Faire A, Weischedel R. Optimal subset selection: multiple regression, interdependence, and optimal network algorithms. Springer-Verlag, 1974. p:16.

[11] Bozdogan H. Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithm. In: Statistical Data Mining and Knowledge Discovery. H. Bozdogan (ed). Chapman and Hall/CRC. Florida, 2004.

[12] Bozdogan H. Information Complexity and Multivariate Learning in High Dimensions with Applications in Data Mining. Forthcoming book. 2020

[13] Haff LR. Emprical bayes estimation of the multivariate normal covariance matrix. The Annals of Statistics. 1980; 8(3):586-597

[14] Shurygin A. The linear combination of the simplest discriminator and Fisher's one. In Applied Statistics. Nauka (ed). Moscow. Rusia. 1983.

[15] Press S. Estimation of a normal covariance matrix. Technical Report. University of British Columbia. 1975.

[16] Chen MCF. Estimation of covariance matrices under a quadratic loss function. Research Report S-46, Department of Mathematics, SUNY at Albany (Island of Capri, Italy). 1976; p:1–33.

[17] Bozdogan H. A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation. Invited paper. In Istanbul University Journal of the School Business Administration 2010; 39(2): 370-398

[18] Chen Y, Wiesel A, Eldar YC, Hero AO. Shrinkage algorithms for mmse covariance estimation. IEEE Trans. On Signal Processing 2010; 58 (10): 5016–5029.

[19] Ledoit O, Wolf M. A well conditioned estimator for large dimensional covariance matrices. J Multivariate Anal 2004; 88: 365-411

[20] Thomaz CE. Maximum Entropy Covariance Estimate for Statistical Pattern Recognization. PhD Dissertation, Department of Computing Imperial College. University of London. UK, 2004.

[21] Pamukçu E. Choosing the optimal hybrid covariance estimators in adaptive elastic net regression models using information complexity. J Stat Comput Sim 2019; 89(16): 2983-2996.

[22] Akaike H. Information theory and extension of the maximum likelihood principle. 2nd International Symposium on Information Theory. Budapest: Academiai Kiado 1973; p:267-281,

[23] Akaike H. A new look at the statistical model identification. IEEE Transaction and Automatic Control 1974; AC-19:719-723

[24] Bozdogan H. ICOMP: A new model-selection criterion. In: 1. Conference of the international federation of classification societies 1987; p. 599-608.

[25] Bozdogan H. On the information based measure of covariance complexity and its application to the evaluation of multivariate linear models. Commun Stat Theor M 1990; 1: 221-278

[26] Bozdogan H, Haughton DMA. Information complexity criteria for regression models. Comput Stat Data An 1998; 28: 51-76

[27] Bozdogan H, Howe JA. Misspecified multivariate regression models using the genetic algorithm and information complexity as the fitness function. European Journal of Pure and Applied Mathematics 2012; 5(2), 211-249

[28] Schwarz G. Estimating the dimension of model. Ann Stat 1978; 6: 461-464

[29] Bozdogan H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. Psychometrika 1987; 52(3): 345-370

[30] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA. 1989

[31] Michalewicz Z. Genetic algorithms+ data structures= evolution programs. Springer Science & Business Media, 2013.

[32] Jang JSR. Derivative-Free Optimization. In Neuro-Fuzzy and Soft Computing: A Computational Approach To Learning and Machine Intelligence. Prentice-Hall, USA. 1997; p: 173-196
[33] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc B 2005; 67(2): 301-320.
[34] Shahriari S, Faria S, Gonçalves AM. Variable selection methods in high-dimensional regression—A simulation study. Commun Stat-Simul C 2015; 44(10): 2548-2561.
[35] Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. J Chemometr 1992; 6(5): 267-281
[36] Chatterjee S, Laudato M, Lynch LA. Genetic algorithms and their statistical applications: an introduction. Comput Stat Data An 1996; 22(6): 633-651
[37] Minerva T, Paterlini S. Evolutionary approaches for statistical modelling. Published in: Proceedings of the 2002 Congress on Evolutionary Computation CEC'02. Honolulu, HI, USA 2002; Cat. No.02TH8600
[38] Tolvi J. Genetic algorithms for outlier detection and variable selection in linear regression models. Soft Comput 2004; 8(8); 527-533
[39] Paterlini S, Minerva T. Regression Model Selection Using Genetic Algorithms. Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing. Proceedings of the 11th WSEAS. 2010.

Turkish Journal of Science & Technology
15(1), 13-18, 2020

A Statistical Randomness Generation Algorithm Based on Nonlinear Behavior of Discrete Time Chaotic Systems

Erkan TANYILDIZI[1], Fatih ÖZKAYNAK[2*]
[1,2] Department of Software Engineering, Technology Faculty, Fırat University, Elazığ, Turkey
[1] etanyildizi@firat.edu.tr, [2]ozkaynak@firat.edu.tr

(Geliş/Received: 10/11/2019; Kabul/Accepted: 19/02/2020)

Abstract: Statistical randomness is a critical requirement for many applications. Generally, it is common to use a generator algorithm for statistical randomness. In this study, a generator algorithm proposed benefiting from chaotic systems. This proposed approach is based on chaotic maps with a simpler mathematical model compared to other chaotic system classes. So the generator has high practical applicability. In addition, optimization algorithms to guarantee statistical properties of generator.

Key words: Randomness, chaos, optimization.

Ayrık Zamanlı Kaotik Sistemlerin Doğrusal Olmayan Davranışını Temel Alan İstatistiksel Rasgelelik Üreteç Algoritması

Öz: İstatistiksel rasgelelik birçok uygulama için kritik bir gereksinimdir. Genellikle istatistiksel rasgelelik için bir üreteç algoritması kullanılması yaygın bir yaklaşımdır. Bu çalışmada kaotik sistemlerden yararlanılarak bir üreteç algoritması önerilmiştir. Önerilen bu yaklaşım diğer kaotik sistem sınıflarına göre daha basit matematiksel model sahip kaotik haritaları temel almaktadır. Bu yüzden üretecin pratik uygulanabilirliğinin yüksektir. Ek olarak optimizasyon algoritmaları üretecin istatistiksel özelliklerini garanti etmektedir.

Anahtar kelimeler: Rasgelelik, kaos, optimizasyon.

1. Introduction

Statistical randomness based on chaos theory is a hot topic [1-3]. A raw literature review shows that over the last decade, the number of studies related to chaos-based random number generators has been over 15,000. At this point, a question comes to mind. Have chaotic systems been used with the most appropriate approach? The aim of this study is to search for answers to this question and to determine the chaotic system parameters which can produce the deterministic random numbers required by various applications with the using of optimization algorithms.

The rest of the study has been organized as follows. In the second section, the general characteristics of chaotic systems are shortly explained. In the third section, randomness and deterministic random number generators are briefly introduced. In the fourth section, the details of the proposed method which show how to select of the chaotic system parameters to be used to provide the best randomness requirements by using optimization algorithms well-known in the literature are presented. The obtained results are discussed and the study is summarized in the last section.

2. Chaotic Systems

The relationship between the causes and consequences of real-world events is very complex. Chaotic systems is precisely based on this complex relationship. Irregular and unpredictable behavior of nonlinear systems is called chaos [1]. A small change in the initial conditions and control parameters of chaotic systems leads to very different outputs. This phenomenon is known as the butterfly effect and indicates that a very small change in a butterfly's flapping can trigger a hurricane. It is stated that the estimation of outputs in chaotic systems is impossible. In other words, a chaotic system is similar to a probabilistic system. But the source of the disorder is not the unpredictable external influences but the real dynamics of the system.

There are various classes of chaotic systems such as discrete-time, continuous-time, time-delayed, spatial and hyper-chaotic. These chaotic system classes are ranked from simple to complex considering the mathematical

* Corresponding author: ozkaynak@firat.edu.tr. ORCID Number of authors: [1] 0000-0003-2973-9389, [2] 0000-0003-1292-8490

models they possess. This study has been based on discrete-time chaotic systems. Because these systems have a simpler structure than others. The most important reason for this preference is to ensure effectiveness by keeping the complexity of the deterministic-random number generator as low as possible. Two different discrete-time chaotic systems have been used in the study. The mathematical models of these chaotic systems are given in Table 1 [1].

Table 1. Properties of some discrete time chaotic systems

Chaotic Map	Mathematical Model	Initial Condition	Control Parameters
Logistic Map	$x_{n+1} = a * x_n * (1 - x_n)$	$x_n \in [0,1]$	$a \in [3.5, 4]$
Circle Map	$x_{n+1} = x_n + a - \frac{b}{2\pi}\sin(2\pi x_n) \bmod 1$	$x_n \in [0,1]$	$a \in [0, 1]$, $b \in [0, 4\pi]$

3. Random Number Generators

Randomness is a fundamental characteristic required in many applications such as statistics, game theory, simulation, numerical analysis, entertainment and cryptology. Real world Random Number Generators (RNG) are divided into two basic classes. These classes are Deterministic RNG (also known as pseudo random number generators) and True RNG [3, 4].

Figure 1 shows the general design architecture of the Deterministic RNG (DRNG). In this general design architecture, $r_1, r_2, \dots, r_n \in R$ shows random numbers, and $s_n \in S$, shows the internal state. Here the S and R finite sets are called the state space and output space of the DRNG, respectively. The $\psi: S \to R$ output function generates the r_n random number from the current s_n internal state. The s_n status is then updated to $s_{n+1} := \varphi(s_n)$ using the φ state transition function. The initial internal state value s_1 is updated using the seed value s_0 in $s_1 := \varphi(s_0)$ format or by using more complex designs. It is clear from s_0 seed value that all $s_1, s_2, \dots, s_n$ inner states and $r_1, r_2, \dots, r_n$ random numbers can be estimated. Therefore, the seed value must be selected randomly in safety-priority applications and the state transition function and the output function must be sufficiently complex. The disadvantage of DRNG is that the output values can be fully determined by the seed value and that future random numbers are only dependent on the current internal state. Therefore, the internal state must be protected even if the device is not activated. The advantages of DRNG are that they are cheaper since they do not need any dedicated hardware [3, 4].

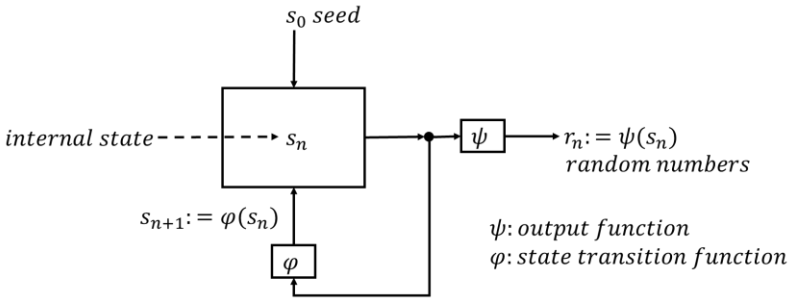

Figure 1. General design architecture of DRNG

One of the important sources for design and analysis of DRNG is Knuth's classic work called "The Art of Programming" [5]. Since the publication of this study, many RNG algorithms have been proposed and many researchers have continued to work on this hot topic. Ripley addressed the problems faced by personal computer users in order to produce random numbers in his study published in 1983 [6]. Ripley was developed effective methods of generating exponential, normal and Poisson distribution arrays. In 1990, L'Ecuyer solved the problem of generating uniform random variables for a user with a moderate computer use knowledge [7]. In a 1990 study by James, pseudo-random number generators for Monte Carlo calculations were addressed [8]. In another study conducted in 1990, Lagarias identified pseudo-random number generators based on number theory, one way functions and secret-key encryption systems [9]. Lagarias also summarizes the results of these generators on cryptanalysis. A comprehensive evaluation study on random number generators was performed by Ritter [10].

When the current literature is examined, it is seen that another important topic used in RNG design is chaos based randomness [11-18]. The close similarities between the two disciplines led the researchers to focus their attention on this subject. Because the complex relationship between the causes and consequences of real world events is one of the basic requirements expected from a RNG and is based on this complex relationship in chaotic systems.

4. Proposed Method

In the literature, a study was carried out to determine the optimal starting conditions and control parameters for the Logistic map using the brute force approach [19]. The pseudo-code of the algorithm is given in Figure 2. In the study proposed by Ozkaynak, it was very difficult to determine the most suitable values in infinite space and only an analysis study has been carried out using the 3 values after the comma. The complexity of the algorithm is $O(n^3)$, as can be easily seen from the pseudo-code. In other words, the problem is computationally difficult. Therefore, optimization algorithms are needed to find an approximate solution.

```
BitSequence[1000000]
for (a in 0, 1, 0.001)
    for (Xn in 3.5, 4, 0.001)
        for (i in 1, 1000000, 1)
            Xn+1=a*Xn*(1-Xn)
            if(Xn+1>0.5)
                BitSequence[i]=0
            else
                BitSequence(i)=1
            Xn=Xn+1
        end
    end
end
```

Figure 2. The pseudo-code of the brute force algorithm in Ref [19]

The optimization process tries to obtain the best solution among all solutions under the given conditions when a problem is solved. The variables that affect the performance of optimization and have values under our control are called decision variables. The objective function is created by analytically demonstrating the effects of decision variables on the objective. In most cases, only certain values of decision variables should be used. These limitations on the values of decision variables are called constraints. In other words, the goal of optimization is to find the best combination of objective functions. providing all constraints given among all possible combinations of decision variables.

The unique aspect of the study is to realize RNG designs by determining the initial conditions and control parameters of chaotic systems using optimization algorithm. Objective function of optimization algorithm is randomness requirements. As a result of the literature review, it was determined that there was no study aiming at reaching this goal before. Although there are studies in the literature that use optimization algorithms for chaos control or chaotic parameter estimation, these studies are very different from the purpose of the proposed method. In the study, the initial conditions and control parameters of two different chaotic systems have been selected with seven well-known meta-heuristic optimization algorithms. The optimization algorithms [20. 21]:

- Optimization algorithms based on Biological: Differential Evolution (DE), Particle Swarm Optimization (PSO), Symbiosis Organisms Search (SOS) Algorithm,
- Optimization algorithms based on Physics and Chemistry: Gravitational Search Algorithm (GSA), Harmony Search Algorithm (HS),
- Optimization algorithms based on Mathematics: Golden Sine Algorithm II (GoldSA-II).

The parameters used in the comparison algorithms are given as follow:

- PSO: Inertia Weight Damping Ratio = 0.99, Inertia Weight = 1, Global Learning Coefficient = 2.0, Personal Learning Coefficient = 1.5.
- GSA: Rnorm = 2, Rpower = 1, Elitist Check = 1.

- HS: Harmony Memory Consideration Rate= 0.9, Fret Width Damp Ratio= 0.995, Fret Width = 0.02*(Upper Bound - Lower Bound), Pitch Adjustment Rate=0.1.
- DE: Upper Bound of Scaling Factor=0.8, Lower Bound of Scaling Factor =0.2, Crossover Probability=0.2.
- SOS: Benefit factor (BF): random number either 1 or 2.
- ACO: Sample Size:40, Intensification Factor (Selection Pressure)= 0.5, Deviation-Distance Ratio=1.
- GoldSA-II : Gold section constant=[-pi*rand, pi*rand], Golden ratio (τ)= 0.618033

Analysis has been realized for 1,000,000 bits have been produced by using two discrete chaotic map outputs. After the generation of random number sequences, properties of statistical randomness have been checked using two different tests. First statistical randomness test is known as the monobit (frequency) test [22]. Definition of this test is that "Monobit test measures whether the number of 0s and 1s produced by the generator are approximately the same as would be expected for a truly random sequence." Monobit test results are given in Table 2. Chaotic maps in Table 1 shows as F1 and F2 symbols, respectively.

Table 2. Monobit test results for 1,000,000 random bit

	PSO	**ACO**	**DE**	**GSA**	**HS**	**SOS**	**GoldSA-II**
F1	258	**68**	238	1.484	592	416	212
F2	**88**	132	624	14.854	9.220	15.604	670

Another statistical randomness test is the chi-square test. Definition of this test is that "A chi-square statistic compares these substring proportions to the ideal 1/2. The statistic is referred to a chi-squared distribution with the degrees of freedom equal to the number of substrings. The chi-squared distribution is used to compare the goodness-of-fit of the observed frequencies of a sample measure to the corresponding expected" [22, 23]. In this test, the produced 1,000,000 bit has been divided into 4-bits lengths to produce 250,000 decimal numbers ranging from 0 to 15. In this case, according to the chi-square test, the expected frequency values are 15,625. Observed values for seven different optimization algorithms are given in Table 3.

Table 3. Observed values for produce 250,000 decimal numbers ranging from 0 to 15

		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
PSO	**F1**	15576	15653	15521	15667	15613	15591	15674	15594	15724	15612	15584	15720	15631	15603	15599	15638
	F2	15710	15486	15621	15893	15540	15480	15900	15442	15414	16006	15362	15419	15856	15468	15610	15793
ACO	**F1**	15616	15566	15623	15747	15594	15674	15685	15605	15659	15584	15552	15637	15623	15667	15605	15563
	F2	15613	15352	15710	15790	15758	15469	15787	15442	15504	15768	15719	15639	15553	15788	15469	15639
DE	**F1**	15706	15656	15555	15657	15592	15671	15568	15619	15566	15637	15595	15693	15538	15682	15624	15641
	F2	15720	15540	15698	15679	15838	15466	15793	15537	15482	15727	15415	15523	15673	15643	15549	15717
GSA	**F1**	14983	15631	15952	15416	15862	15661	15659	15664	15720	15477	15617	15589	15685	15689	15859	15536
	F2	14647	17172	16044	15831	17128	14118	16436	15185	17396	15205	15216	15376	15148	16189	14961	13948
HS	**F1**	15615	15553	15607	15681	15524	15648	15730	15712	15541	15615	15678	15642	15615	15703	15578	15558
	F2	16483	16125	16536	15322	16163	15417	13897	16233	16105	15603	14937	14923	15587	14756	16453	15460
SOS	**F1**	15671	15606	15597	15592	15576	15628	15622	15618	15785	15665	15563	15643	15607	15624	15593	15610
	F2	17135	16118	16168	15578	16398	14913	15510	15085	16209	15254	14847	15484	15593	15312	15312	15084
GoldSA-II	**F1**	15671	15660	15624	15577	15554	15653	15623	15684	15643	15582	15561	15610	15565	15687	15664	15642
	F2	15707	15627	15699	15567	15649	15659	15850	15263	15391	15704	15639	15746	15632	15829	15469	15569

In the case study 1, random numbers have been generated between 0 and 15. Therefore, degree of freedom (DF) of chi-square test is 16. All confidence values are given in Table 4 for the degree of freedom 16.

Table 4. Confidence values for degree of freedom 16

	Confidence Values										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252

In order to be able to say that the random numbers produced are statistically random, the calculated chi-square value should be smaller than the values in the Table 4. The calculated chi- square values for the seven different optimization algorithms are given in Table 5. The best, worst, average and standard deviation values are given in

the Table 5. The most appropriate chi square values are shown in bold font for the best and mean values, in the Table 5. Among the optimization algorithms, GoldSA-II algorithm has been found to have better results than other algorithms. However, it should not be overlooked that other algorithms meet the randomness requirements for many confidence values.

Table 5. The chi- square test results for 1.000.000 random bit

		PSO	ACO	DE	GSA	HS	SOS	GoldSA-II
F1	**Best**	2.7180	2.5686	2.5702	46.4393	3.9171	2.5446	**1.9651**
	Mean	3.9958	2.8057	3.5108	1.7950e+03	4.4168	3.7470	**2.6258**
	Worst	4.7456	3.1441	4.8901	6.4223e+03	4.7971	4.5102	**3.0024**
	Std	0.9422	0.2278	0.9242	2.6923e+03	0.3338	0.7433	0.4048
F2	**Best**	41.1837	19.9224	**14.9336**	1.0426e+03	528.8125	372.6726	22.0122
	Mean	251.3002	241.4392	205.3796	1.6048e+03	872.4748	482.8232	**139.4131**
	Worst	551.7382	507.1160	354.7210	2.4460e+03	1.5448e+03	664.1293	**334.8357**
	Std	188.6843	227.7812	150.2168	756.5389	419.9311	108.5574	122.3634

In the above analysis, all optimizers have been run independently 5 times in equal population and iterations and the results have been recorded. The number of populations for chaotic system functions is 10 and the maximum number of function evaluations is 1.000. Table 6 shows the determined initial conditions and control parameters using seven optimization algorithm. Numerical deterioration is an important problem for chaotic systems [24, 25]. In order not to be affected by this problem, the highest precision provided by the computer where the analyzes have been performed has been used. Therefore, in order to show the numerical values in the best way, Table 6 is given in two parts.

Table 6. Most Suitable System Parameters for for 1,000,000 random bit

	Logistic Map (F1)	
	Xn	**a**
PSO	0.583587285455663	4
ACO	0.747882339784262	4
DE	0.312955856922984	4
GSA	0.438200825697660	3.99647848943457
HS	0.603238510606358	4
SOS	0.269877194226175	4
GoldSA-II	0.805314780517318	4

	Circle Map (F4)		
	Xn	**a**	**b**
PSO	0.765111672245320	1	11.502628728405833
ACO	0.717488331494669	0.999999826893032	11.502276826049323
DE	0.251024684144349	1	11.5022593721011
GSA	0.589147901804019	0.707607013522381	10.313736784726066
HS	0.492349236734390	0.978890082799524	11.718949046051932
SOS	0.181617966853290	0.732578163414611	10.1026142685088
GoldSA-II	0.100001102537958	1	11.5022370468944

5. Conclusion and Discussion

The purpose of this study is to determine the optimal initial conditions and control parameters of the chaotic systems to be used in the chaos based DRNG designs. Optimization algorithms have been used to achieve this purpose. The optimal initial conditions and control parameters for discrete-time chaotic systems have been obtained. All these results draw attention to the problems of doing various studies on the theoretical similarities between chaos and randomness. Therefore, in chaos-based RNG designs, the initial conditions and control parameters of the chaotic system should be selected in the most appropriate way. RNG design architecture and application specific randomness requirements must be taken into account when making this selection. In this study, a method has been proposed to make these choices in the most appropriate way. The parameters used in the analysis are given. Researchers who wish to work in this area can adapt the proposed method according to their

requirements. Therefore, after successful implementation of the proposed method, the outcomes are expected to contribute significantly to the literature of chaos-based randomness.

References

[1] Sprott J. Elegant Chaos Algebraically Simple Chaotic Flows. World Scientific, 2010.

[2] Özkaynak F. Brief review on application of nonlinear dynamics in image encryption. Nonlinear Dynam 2018; 92(2): 305-313. https://doi.org/10.1007/s11071-018-4056-x

[3] Stipčević M, Koç ÇK. True Random Number Generators. In: Koç ÇK (eds) Open Problems in Mathematics and Computational Science. Cham: Springer, 2014.

[4] Schindler W. Random Number Generators for Cryptographic Applications. Koç ÇK (ed.): Cryptographic Engineering. Signals and Communication Theory, Berlin: Springer, 2009.

[5] Knuth D. The Art of Computer Programming, Vol. 2, Seminumerical Algorithms. 2nd ed. Reading, Massachusetts: Addison-Wesley,1981.

[6] Ripley B. Computer Generation of Random Variables: A Tutorial. Int Stat Rev 1983; 51: 301-319.

[7] L'Ecuyer P. Random Numbers for Simulation. Commun ACM 1990; 33(1): 85-97.

[8] James F. A review of pseudorandom number generators. Compu Phys Commun 1990. 60: 329-344.

[9] Lagarias JC. Pseudorandom Number Generators in Cryptography and Number Theory. Proc Symp Appl Math 1990; 42: 115–143.

[10] Ritter T. The Efficient Generation of Cryptographic Confusion Sequences. Cryptologia 1991; 15(2): 81-139.

[11] Aydın Y, Özkaynak F. A Provable Secure Image Encryption Schema Based on Fractional Order Chaotic Systems. The 23rd International Conference ELECTRONICS 2019; 17-19 June 2019; Palanga, Lithuania.

[12] Tapiero CS, Vallois P. Randomness and fractional stable distributions. Physica A: Statistical Mechanics and its Applications 2018; 511: 54-60

[13] Bürhan Y, Artuğer F, Özkaynak F. A Novel Hybrid Image Encryption Algorithm Based on Data Compression and Chaotic Key Planning Algorithms. IEEE 7th International Symposium on Digital Forensic and Security; June 10-12 2019; Barcelos, Portugal.

[14] Dastgheib MA & Farhang M. A digital pseudo-random number generator based on sawtooth chaotic map with a guaranteed enhanced period. Nonlinear Dynam 2017; 89: 2957-2966 https://doi.org/10.1007/s11071-017-3638-3

[15] Murillo-Escobar MA, Cruz-Hernández C, Cardoza-Avendaño L, et al. A novel pseudorandom number generator based on pseudorandomly enhanced logistic map. Nonlinear Dynam 2017 87: 407-425. https://doi.org/10.1007/s11071-016-3051-3

[16] Lv X, Liao X & Yang B. A novel pseudo-random number generator from coupled map lattice with time-varying delay. Nonlinear Dynam 2018; 94: 325-341. https://doi.org/10.1007/s11071-018-4361-4

[17] Özkaynak F. Cryptographically secure random number generator with chaotic additional input. Nonlinear Dynam 2014; 78: 2015-2020. https://doi.org/10.1007/s11071-014-1591-y

[18] Sahari ML & Boukemara I. A pseudo-random numbers generator based on a novel 3D chaotic map with an application to color image encryption. Nonlinear Dynam 2018; 94: 723-744. https://doi.org/10.1007/s11071-018-4390-z

[19] Özkaynak F. The Effects on Performance of Using Chaotic Systems in Entropy Source of Deterministic Random Number Generators. 11th CHAOS 2018 International Conference; 5 - 8 June 2018; Sapienza University of Rome, Italy. pp.415-420

[20] Özer AB. CIDE: Chaotically Initialized Differential Evolution. Expert Syst Appl 2010; 37(6): 4632-4641

[21] Tanyıldızı E, Demir G. Golden Sine Algorithm: A Novel Math-Inspired Algorithm. Advances in Electrical and Comput Eng 2017; 17(2): 71-78.

[22] Rukhin A, Soto J, Nechvatal J, Smid M, Barker E, Leigh S, Levenson M, Vangel M, Banks D, Heckert A, Dray J, Vo S. A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST Special Publication 800–22rev1a, 2010.

[23] Masoom MA, Umbach D and Saleh AKMDE. Estimating Life Functions of Chi Distribution Using Selected Order Statistics. IIE Transactions 1992; 24(5): 88-98.

[24] Özkaynak F. A novel method to improve the performance of chaos based evolutionary algorithms. Optik 2015; 126(24): 5434-5438. https://doi.org/10.1016/j.ijleo.2015.09.098

[25] Persohn KJ, Povinelli RJ. Analyzing logistic map pseudorandom number gen- erators for periodicity induced by finite precision floating-point representation. Chaos Solit Fract 2012; 45: 238–245.

Turkish Journal of Science & Technology
15(1), 19-28, 2020

Sustainable Bicycle Path Planning for Medium-Sized Cities by Using GIS-Based Multicriteria Decision-Making Analysis: A Case Study From Turkey

Rifat OLGUN[1*]
[1] Department of Park and Horticulture, Serik G-S Sural Vocational School of Higher Education, Akdeniz University, Antalya, Turkey
[*1] rifatolgun@akdeniz.edu.tr

(Geliş/Received: 28/01/2020; Kabul/Accepted: 14/02/2020)

Abstract: The world's population is increasing rapidly and consequently, urban areas are also expanding and developing rapidly. This growth and development in cities leads to the differentiation of urban transportation systems and, in particular, the increase in the number of motor vehicles. This increase in the number of motor vehicles in order to provide transportation leads to many problems (air pollution, traffic congestion, carbon gas emission etc.). In order to prevent these problems, many researchers and decision makers state that sustainable transportation systems such as public transportation, walking and cycling should be used. The objective of the research is to develop a sustainable bike path proposal for cities through GIS-based multicriteria decision-making analysis, taking into account physical, environmental and visual factors of medium-sized cities. In this context, the physical, environmental and visual factors of the city of Nigde, Turkey, which is considered to be a medium-sized city, were evaluated with GIS-based multicriteria decision-making analysis, and a sustainable bike path proposal was developed for Nigde according to the zoning plans.

Key words: Transportation system, bicycle, gis, multicriteria decision-making, ahp.

Gis Tabanlı Çok Kriterli Karar Verme Analizi Kullanarak Orta Ölçekli Kentlere Yönelik Sürdürülebilir Bisiklet Yolu Planlaması: Türkiye Örneği

Öz: Dünya nüfusu hızlı bir şekilde artmakta ve buna bağlı olarak kentsel alanlar da hızlı bir şekilde büyümekte ve gelişmektedir. Kentlerde yaşanan bu büyüme ve gelişme kent içi ulaşım sistemlerinin farklılaşmasına ve özelliklede motorlu taşıtlarda artış yaşanmasına neden olmaktadır. Ulaşımın sağlanması amacıyla motorlu taşıtlarda yaşanan bu artış beraberinde birçok sorunun (hava kirliliği, trafik sıkışıklığı, karbon gazı emisyonu vb.) ortaya çıkmasına sebep olmaktadır. Bu sorunları engellemek amacıyla çok sayıda araştırmacı ve karar verici toplu taşıma aracı kullanma, yürüyüş yapma, bisiklet kullanımı gibi sürdürülebilir ulaşım sistemlerinin kullanılması gerektiğini ifade etmektedir. Araştırmanın amacı, orta ölçekli kentlere ait fiziksel, çevresel ve görsel faktörler dikkate alınarak CBS tabanlı çok kriterli karar verme analizi ile bu kentlere yönelik sürdürülebilir bisiklet yolu önerisi geliştirmektir. Bu kapsamda orta ölçekli kent olarak değerlendirilen Niğde/Türkiye kentinin sahip olduğu fiziksel, çevresel ve görsel faktörler CBS tabanlı çok kriterli karar verme analizi ile değerlendirilmiş ve imar planlarına göre Niğde kentine yönelik sürdürülebilir bisiklet yolu önerisi geliştirilmiştir.

Anahtar kelimeler: Ulaşım sistemi, bisiklet, gis, çok kriterli karar verme, ahp.

1. Introduction

There has been a significant increase in the urban population in recent years due to an increase in the world population and migration from rural areas to city centers. The global urbanization rate was calculated as 29% in 1950, 37.4% in 1975 and 47.1% in 2000 [1,2]. Today, more than half of the world's population living in societies that are increasingly globalized and interacting with each other lives in urban areas. In this context, it is estimated that this increase in the urban population will continue steadily, and this rate is expected to be approximately 60% by 2030 [3,4]. In light of this worldwide development, the urban population in Turkey, particularly with post-1950 migration from rural areas to cities, has increased significantly. According to 2019 data from the Turkish Statistical Institute [5], 92.3% of the country's population lives in cities, while 7.7% lives in villages.

With the increase in the population, urban settlements have expanded, and consequently vehicle use has increased [6]. This increase in the number of motor vehicles to provide transportation has caused many problems. These problems arising from the existing transportation system necessitate the planning of different transportation systems. For example, the increase in global problems, such as climate change, the increase in oil production and use, traffic congestion and road safety, encourages the use of sustainable transportation systems, such as walking,

* Corresponding author: rifatolgun@akdeniz.edu.tr. ORCID Number of authors: [1] 0000-0002-5396-057X

cycling and public transportation [7,8]. In this context, these alternatives should also be considered in spatial planning studies.

Sustainable transport, defined by Black [9] as "meeting the current transportation and mobility needs without compromising the ability to meet the transportation and mobility needs of future generations," has recently become an important target in transport planning and research [10]. Sustainable bike path planning requires a systematic approach that takes into account the planning criteria of the zoning plans that are effective in the development of the city, and it necessitates the use of Geographic Information Systems (GIS) in conjunction with Multicriteria Decision-Making Methods (MCDM). In this context, the integration of GIS and Analytic Hierarchy Process (AHP) greatly simplifies the decision-making process [11,12,13,14]. It is possible to evaluate multicriteria decision-making analysis based on GIS, in two different sections. The first one is geographic information systems, and the other is a multicriteria decision-making analysis method.

GIS are used in many different fields (agriculture, geology, urban planning, etc.). One of these areas is research conducted for the planning of bike paths. For example, analyses such as determining the most suitable routes for bikes, analyzing the most cost-effective route, and estimating the demand for bike paths can be evaluated in this context [15,10].

Decision making is defined as the process of determining the best choice among all possible alternatives. However, there is no single method used by decision makers facing many different problems and alternatives to determine the best option [16]. In this context, the Multicriteria Decision-Making Method (MCDM) is one of the most important decision-making methods and serves to determine the best option among all existing solutions [17,18]. Despite debates in the academic field, the Analytic Hierarchy Process (AHP) method is one of the most preferred multicriteria decision making methods. Originally introduced by Saaty in the 1970s, this method has been updated, with changes made by different researchers to date [19].

In the field of transportation systems, the AHP method is used in fields such as planning traffic, planning transportation, prioritizing urban transportation options, planning the most appropriate rail network, and selecting light rail corridors and routes [20,21,22,23]. In addition, this method is also preferred for tasks such as evaluating the public transport fare system, classifying the sustainability of transportation investments, performing quality analysis of public transport service, and prioritizing public transport companies [24,25,23].

The objective of the research is to develop a sustainable bike path proposal for cities through GIS-based multicriteria decision-making analysis, taking into account physical, environmental and visual factors of medium-sized cities. In this context, the physical, environmental and visual factors of the city of Nigde, Turkey, which is considered to be a medium-sized city, were evaluated with GIS-based multicriteria decision-making analysis, and a sustainable bike path proposal was developed for Nigde according to the zoning plans.

2. Materials and Methods

2.1. Study area

Nigde, which is a medium-sized city in the Central Anatolia region of Turkey, was selected as the research area. Nigde is adjacent to the Adana, Kayseri, Konya and Mersin provinces, which are important major cities of Turkey. Located between 37°25′-38°58′ north latitude and 33°10′-35°25′ east longitude, Nigde Province has a surface area of 7.795,22 km² and an altitude of 1.229 meters [26,27]. In the city, which has a continental climate, agricultural and animal husbandry activities are carried out at a high rate (Figure 1).

According to the data obtained from the zoning plan reports, the city of Nigde was generally established on flat and near-flat areas around the hill where Nigde Castle is located and to its west. Since the first settlements were located on the hill and around the castle, today, the densest settlement area is also located in this area. Moreover, due to the influence of the municipality and government offices, which are located in the immediate vicinity of the hill, public institutions and organizations are located in this region. Thus, this region, where education, trade, administrative, health and cultural buildings coexist, forms the city center of Nigde [28].

The research area, which connects Bor Plain and Misli Plain, located between Melendiz Mountain and the Aladağlar and Bolkar Mountains, is currently being developed in a linear manner along a southwest-northeast line [14]. According to the zoning plans in force, dense residential areas were planned north of the city center. With the urban transformation project carried out in this region, the population density has shifted to this region.

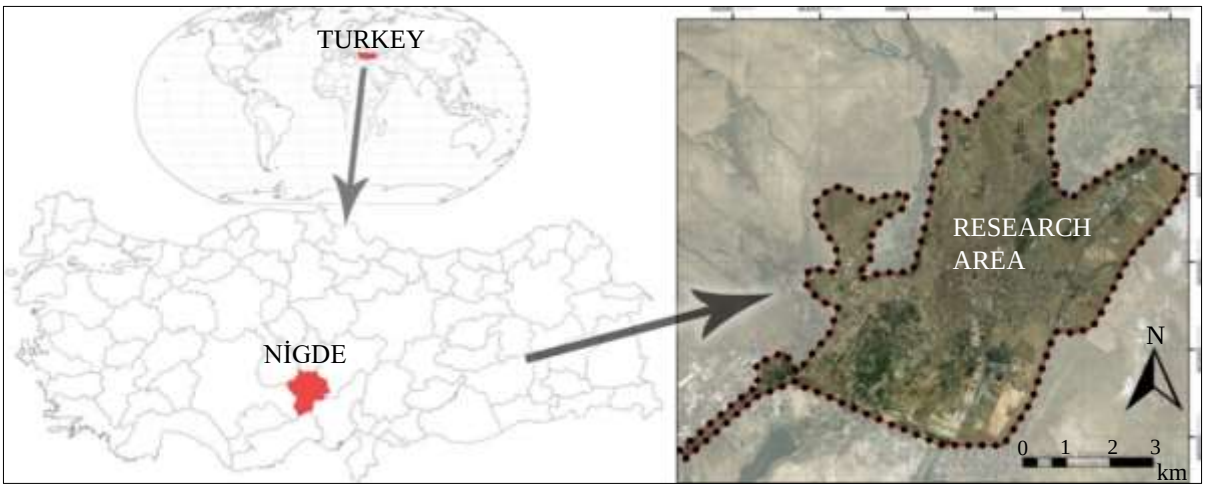

Figure 1. Location of the research area

2.2. Methods

This research was carried out in five stages in order to plan sustainable bike paths for medium-sized cities.

The first stage of the research consisted of obtaining the data related to the subject and field of the research. In this context, the social, cultural and physical data that will be used for the planning of sustainable bike paths for the research area were obtained from field studies, interviews with related institutions and organizations, and the extant literature. The existing plans for the research area were provided by the Municipality of Nigde, the Special Provincial Administration of Nigde and the Provincial Directorate of Environment and Urbanization.

The second stage of the research consisted of determining factors, subfactors and suitability values for sustainable bike path planning. A model for determining the route of a bicycle path was created in the Inner City Bike Paths Guide, prepared by the Turkish Republic Ministry of Environment and Urbanization. In this context, the criteria included in the determination of the bicycle path route are grouped under three main headings: physical (slope, roadway width, physical condition), environmental (land use type, user density, integration to transportation system, traffic density) and visual (building height, proximity to green spaces, building status) criteria. Physical criteria determine the effect of the physical conditions of the road on the model; the environmental criteria determine the effect of the number of users, usage type differences, the connection of the route with the transportation system to be determined, the status of ownership, and the traffic density in the model; and visual criteria determine the effect of the users' driving quality, feeling of safety, and the effect of the relationship with the items around him or her in the model.

Suitability values for the evaluation subfactors were determined from the most important to the least important using a 4-point likert scale. In the event that the evaluation subfactors affected the potential use of the settlement equally, both subfactors were given the same value as the slope and land use type factors included in the evaluation factors.

The third stage of the research consisted of calculating binary comparison matrices. To determine the suitability coefficients of the factors for the planning of a sustainable bike path by taking physical, environmental and visual criteria into account by expert groups, a questionnaire including binary comparison questions was prepared. The questionnaire was applied to 16 experts in the disciplines of architecture (3), urban and regional planning (4), landscape architecture (6) and civil engineering (3) on a voluntary basis. Binary comparison matrices were created according to comparisons made by experts. In this context, a relative importance scale (1-9) developed by Saaty (1983) was used to assign values to binary comparison matrices [14,28,29]. If one factor is more important than another, it receives values such as 1, 3, 5, and 9 (Table 1). If it is unimportant, it receives the values that are the opposite of these values (1/3, 1/5 et al.) [30,31].

Table 1. Description of the fundamental scale of values [32,33]

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgement slightly favour one activity over another
5	Essential or strong importance	Experience and judgement strongly favour one activity over another
7	Demonstrated importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Absolute importance	The evidence favouring one activity over another is of the highest possible order or affirmation
2,4,6,8	Intermediate values between adjacent judgments the two	
Reciprocals of above non-zero	If activity (i) has one of the above non-zero numbers assigned to it when compared with activity (j), then (j) has the reciprocal value when compared with (i)	A reasonable assumption
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining n numerical values to span the matrix

Once the binary comparison matrix is prepared, the matrix must be normalized. To normalize the matrix, the sum is taken for each column in the matrix, and the elements of the matrix are divided into the sum of the columns with which it is related. Then, the sum of the rows created for each alternative or criterion is taken in the normalized matrix. The resulting values are the priority values for the criteria or alternatives. The matrix formed by the priority values is defined as the "priority vector matrix."

For each criterion/alternative in the priority vector matrix, the priority value is multiplied by all of the elements in the column in the binary comparison matrix of that criterion/alternative, and the weighted total matrix is obtained. The row total values in the weighted total matrix are divided into priority vector matrix row values and the eigenvalue (λmax) is calculated by taking the arithmetic mean of the values in the last matrix created.

The following formulas were used to calculate the consistency ratio and consistency index of the generated matrix. In this context, the matrix is considered to be consistent if the consistency ratio is 10% or less [34,35]. In this research, the consistency ratio was calculated as 0.03.

$$CI= (\lambda max-n) / (n–1) \qquad CR= CI / RI \tag{1}$$

CR: Consistency Ratio, CI: Consistency Index, RI: Random Index, λmax: The largest eigenvalue of the matrix, n: Number of elements in the matrix.

Table 2. Random Index (36)

n	1	2	3	4	5	6	7	8	9	10
RI	0,00	0,00	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

The fourth stage of the research consisted of the preparation of the data infrastructure for the physical (slope, roadway width, physical condition), environmental (land use type, user density, integration to transportation system, traffic density) and visual (building height, proximity to green spaces, building status) criteria included in the process of determining the sustainable bicycle path route. In this context, data obtained from Nigde Municipality, Nigde Special Provincial Administration, the Provincial Directorate of Environment and Urbanization and field studies were digitized in shapefile format in the ArcGIS environment in accordance with the evaluation factors determined for sustainable bike path planning. This format was then converted to a raster data format after preliminary operations.

In the fifth stage of the research, specified evaluation factors and evaluation subfactors were grouped within themselves, and their suitability values were determined according to their relative importance values. Suitability values given to evaluation subfactors were multiplied by coefficients determined as a result of expert opinions, and the total weight of each factor was determined. Data were obtained by using the 'multi-layer weighted overlay method,' a spatial analysis method in the ArcGIS spatial analyst tools module. As such, routes for the sustainable bike path proposal were obtained (Figure 2).

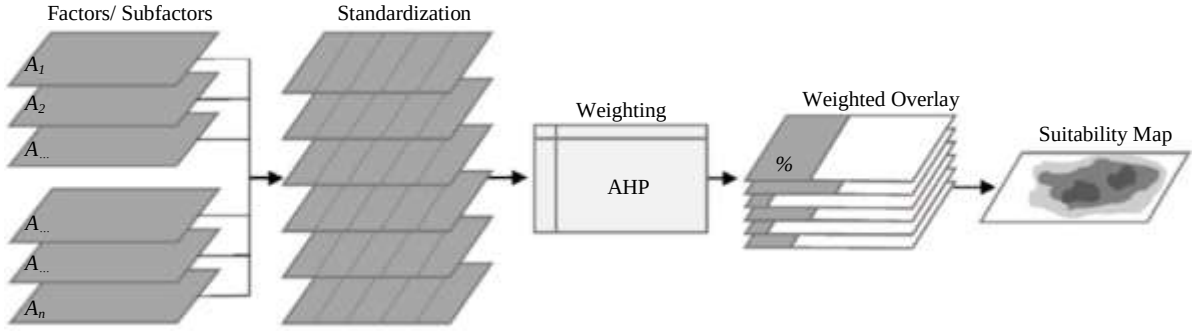

Figure 2. The structure of GIS-based multicriteria decision-making [37]

3. Results and Discussion

Research on the use of bicycles as a means of transport has shown that there are many environmental factors affecting the use of bicycles. In studies conducted by different researchers, it has been determined that factors such as the width of roads, type of land use (commercial, residential, public space etc.), population density, traffic density, and presence of open and green space affect individuals' bicycle use [38,39,40].

In addition, adverse conditions in topography/terrain conditions (steep slopes), road surface (uneven surfaces), traffic conditions (presence of heavy vehicles), and problems with the connection of roads are factors that negatively affect the use of bicycles. These factors need to be taken into consideration in order to increase the use of bicycles and to ensure more comfortable and safe travel for individuals who ride bicycles [41,42].

In the study conducted for sustainable bike path planning in the city of Nigde, when the weight values of the revealed groups and factors were calculated, it could be seen that physical criteria had the highest weight value (0.751). This was followed by environmental factors (0.178) and visual factors (0.070). When the weight degrees of the factors in these groups were calculated, it could be seen that slope had the highest weight level (0.731) among physical factors, traffic density (0.551) had the highest weight among environmental criteria, and closeness to green areas (0.691) had the highest weight among visual criteria. Similarly, in a study conducted by Habibian et al. (2017) in Shiraz, Iran, experts stated that the slope factor was the most important factor in the planning of bike paths (Table 3).

The slope criterion, road width criterion and physical condition of the roads criterion, which were included in the physical criteria for the research area, were examined under 5, 4, and 3 subcriteria, respectively.

The city has developed along a southwest/northeast axis due to the mountainous areas to the north and southeast of the research area. In this context, the slope is flat and close to flat in the central areas of the city. However, it is seen that the slope increases as we progress towards the north and southeast of the city. Especially in the north of the research area, the areas planned as residential areas in the zoning plans are the areas with the greatest slope.

The roads around the residential areas in the zoning plans have a width of over 30 m. The road widths are narrow at many points of the city due to the historical areas within the existing residential area of the city center and the unplanned development in previous periods. However, the width of the main arteries connecting the existing residential areas in the city center with the planned residential areas in the north and south of the city varies between 20 and 30 m. It can also be seen that there is no continuity of road widths at some points.

The physical condition of the existing roads was examined in the field studies carried out within the scope of the research. The roads in the research area are generally in good physical condition. However, deterioration in the coating of the roads in the south of the city has reduced the physical quality of the road. Since new roads will be built within the project determined in the zoning plans, their physical condition has been evaluated as good for this reseach (Figure 3). Land use type criterion, user density criterion, integration into transportation system criterion and traffic density criterion, which were included in the environmental criteria for the research area, were examined under 5, 4, 3, and 3 subcriteria, respectively. When the research area was evaluated in terms of land use type, there was not much forest area around and within the area.

Table 3. The evaluation factors used in the determination of sustainable bicycle path

Groups	Group priority	Main Criteria	Subcriteria	Score	Local Weight	Percent (%)	Consistency ratio (CR)	
							Local	Global
Physical Criteria	0.751	Slope (%)	0-2	4	0.731	73.1	0.06	0.03
			2.1-5	3				
			5.1-8	2				
			8.1-11	1				
			11<	1				
		Roadway Width	0-10 m	1	0.188	18.8		
			10.1-20 m	2				
			20.1-30 m	3				
			30 <… m	4				
		Physical Condition	Good	4	0.081	8.1		
			Middle	3				
			Bad	1				
Environmental Criteria	0.178	Land Use Type	Housing Zone	1	0.131	13.1	0.09	
			UrbanWork-Commerce Space	2				
			Public Space	4				
			Forest-Agricultural Space	3				
			Green Space	4				
		User Density (person/ha)	0-40	1	0.274	27.4		
			41-80	2				
			81-150	3				
			151<…	4				
		Integration to Transportation System	0-300 m	4	0.044	4.4		
			300-1000 m	3				
			1000<… m	2				
		Traffic Density	Intensive 120	4	0.551	55.1		
			Medium Density 40-120	3				
			Not Intense 0-40	2				
Visual Criteria	0.070	Building Height	0-10.5 m	4	0.091	9.1	0.05	
			10.5-30.5 m	3				
			30.5-50.5 m	2				
			Free Height	1				
		Proximity to Green Spaces	0-200 m	4	0.691	69.1		
			201-500 m	3				
			501-800	2				
			800<… m	1				
		Building Status	Detached Building	4	0.218	21.8		
			Attached Buildings	2				
			Semi-Detached Buildings	3				

As the research area is located between the Bor Plain and the Misli Plain, there are fertile agricultural lands in the northeast and southwest of the area. There are many active (park, sports field, children playground) and passive green spaces (cemetery, refuge, traffic island etc.) in the research area. In the study carried out by Olgun [28] for the city of Nigde, it was determined that there were 177 active green spaces with an area of 589,276.01 m^2 in the current state and 4.09 m^2 of active green spaces per capita. In the application zoning plan prepared in accordance with the 2035 projection, 647 active green spaces with an area of 2,717,293.15 m^2 were identified. Urban working areas and commercial areas are densely located to the east of the city center. The residential areas are densely located on flat and nearly flat areas to the west and around the hill where Nigde Castle, the first settlement of the city, is located. In the zoning plans, the areas located to the north of the city center were planned for residential areas. Therefore, it was observed that the population density in the areas to the north of the city center is high in the zoning plans. In addition, there is an intense population increase in the western region, where there has been rapid construction activity recently. In areas where the population density is high, it was observed

that the traffic density is also high. Traffic along the main artery in particular is more intense than it is on other roads.

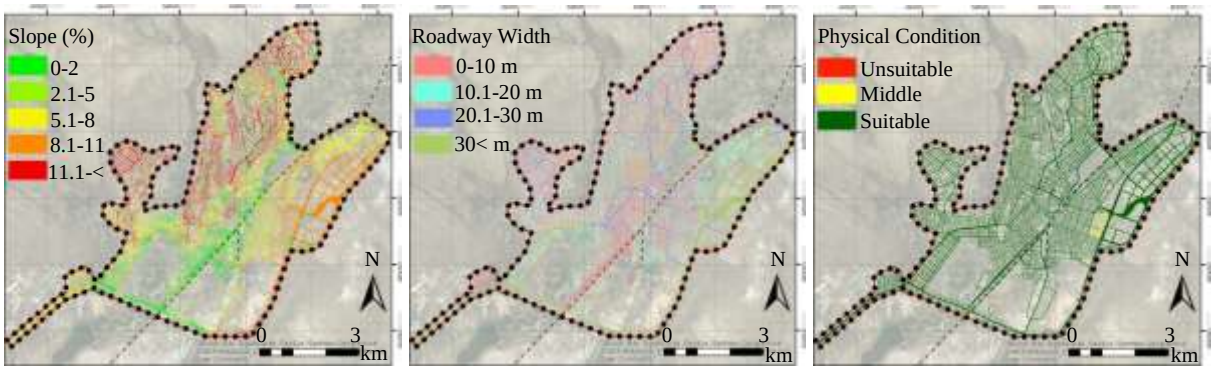

Figure 3. Analysis of physical criteria for sustainable bicycle path proposal

Research conducted in Turkey and in many countries around the world emphasize that bike routes should be integrated with the public transportation system. However, it is not possible to ensure safe and appropriate bike-public transport system integration in many cities in Turkey [43]. In terms of its transportation system, there is an intercity bus terminal to the south of the city center and a train station in the city center. There are also public transportation stops, which provide transportation within the city. Within this scope, the areas where the public transportation vehicles are densely located in the city have been taken into consideration in the integration of bike paths into the transportation system (Figure 4).

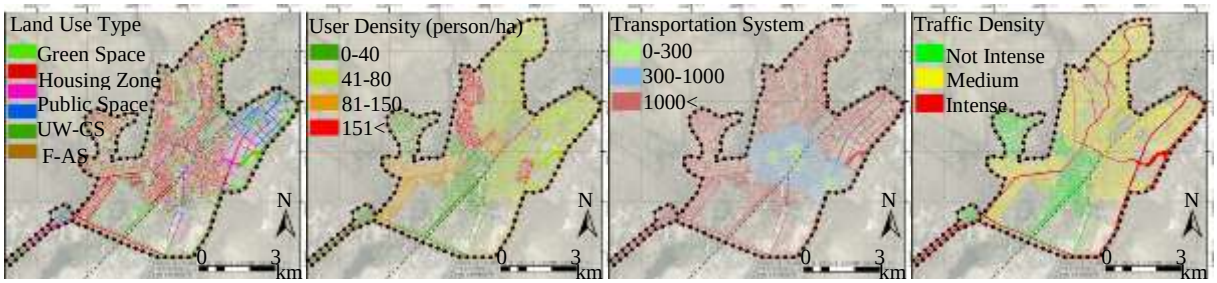

Figure 4. Analysis of environmental criteria for sustainable bicycle path proposal

Building height criterion, proximity to green spaces criterion, and building regulation status criterion, which were included in the visual criteria for the research area, were examined under 4, 4, and 3 subcriteria, respectively. According to the zoning plans, there are buildings with different height values within the research area. However, it can be seen that there are a large number of buildings with a height of 50.50 m in the whole research area.

According to the zoning plans, there are many green spaces in different regions within the research area. Some of these green spaces are at the level of a neighborhood park, while others are the size of a city park. When the influence area of the green spaces planned in the zoning plans was examined, the majority of the research area remained within 200 m of the influence area of the green spaces.

When the research area was examined in terms of the building regulation status, it can be seen that the overall research area was planned in the separate building order. However, it can also be seen that there is an adjacent and block building system in the city center and in the regions to the east of the city center (Figure 5).

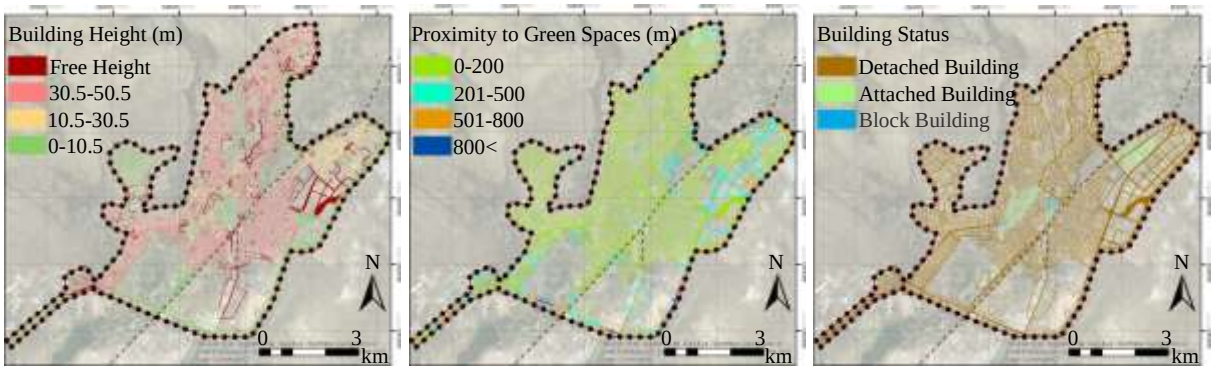

Figure 5. Analysis of visual criteria for sustainable bicycle path proposal

According to the results of the analyses, the land structure of the city, which is currently developed along a southwest-northeast axis in a linear manner stemming from the city center, does not have a high degree of slope. An integrated, sustainable bike path was determined in line with the city's axis of development, with the analyses taking into account the effect of slope and other factors/subfactors. However, due to the high slope of the planned settlement areas, particularly in the north of the city, inadequate road widths, user density and the effects of other factors in accordance with expert opinions, a bike path could not be proposed in this area in the current zoning plans (Figure 6).

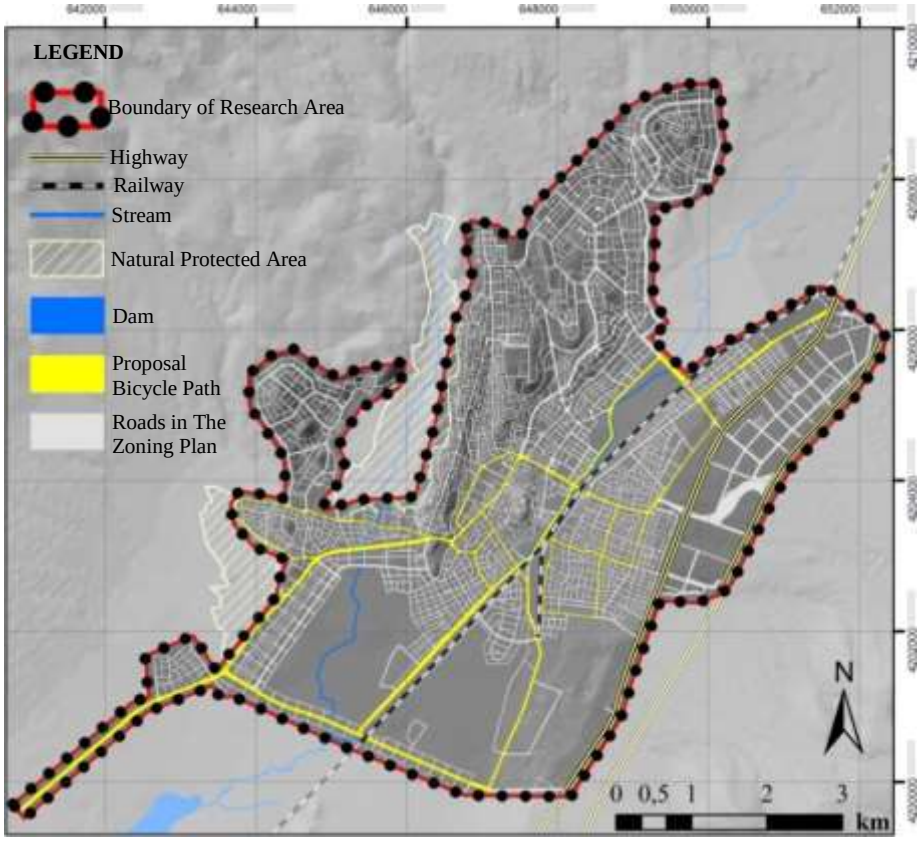

Figure 6. Sustainable bicycle path proposal map for the research area

Urban growth and an increase in housing and population, which emerge with the implementation of zoning plans prepared according to a specific projection, affect the transportation system of cities, necessitating the creation of new alternatives. Examining the relationship between urban growth and motor vehicle traffic density

in their study, Polat et al. [44] reported that an increase in the number of houses affected traffic density. In this context, the urbanization movements that gained momentum in the city of Nigde, especially in recent times, have caused the urban environment to change rapidly. Buildings constructed with developing technology, differentiated and proliferating transportation lines, an increasing number of vehicles, and decreasing green areas constitute the new face of the city of Nigde. Furthermore, the inability of the city center to handle the increasing population has led to the establishment of new residential areas outside of the city [28,45]. This necessitates the creation of new roads and transportation systems. Scientific research conducted by different researchers in various cities around the world has concluded that the presence of bike paths in cities increases the use of bicycles [46]. In this context, the implementation of sustainable bike paths that will be planned in the zoning plans of the research area will lead to an increase in the bicycle usage rate among individuals living in the city. Also developing the necessary infrastructure, raising the awareness of the local administrations and public will increase the use of bicycles [47].

4. Conclusions

Problems related to the transportation system are growing constantly due to an increasing population and urbanization. In this context, researchers are developing different suggestions for solving these problems. In this research, a sustainable bike path was proposed for the city of Nigde, a medium-sized city. The proposed bike path, which was designed by using AHP together with Geographic Information Systems and Multi-Criteria Decision-Making analysis methods, in line with the determined factors and subfactors, will have a sustainable structure, as it was created taking into account the natural and cultural values of the city. Due to the bicycle paths that will be integrated with the transportation system of the city, the use of bicycles will increase, and this will provide many benefits, both to individuals (fuel costs, health, road safety, social activities, etc.) and to the city (air pollution, natural and cultural assets, etc.). In addition, this study will guide planners and decision makers in sustainable bike path planning for medium-sized cities.

References

[1] United Nations. World Urbanization Prospects: The 2003 Revision. New York, USA: United Nations Publication, 2004.
[2] Demir K, Çabuk S. The population growht of metropolitan cities in Turkey. ErU J Soc Sci Inst 2010; 28(1): 193-215.
[3] United Nations. World Urbanization Prospects: The 2014 Revision. New York, USA: United Nations Publication, 2014.
[4] Gupta K, Roy A, Luthra K, Maithani S, Mahavir. Gis based analysis for assessing the accessibility at hierarchical levels of urban green spaces. Urban for Urban Gree 2016; 18: 198-211.
[5] Turkish Statistical Institute. Address based population registration system results, 2019. http://www.turkstat.gov.tr Accessed 12 December 2019.
[6] Pedroso G, Bermann G, Sanches-Pereira A. Combining the functional unit concept and the analytic hierarchy process method for performance assessment of public transport options. Case Stud Transp Policy 2018; 6: 722-736.
[7] Wei F, Lovegrove G. An empirical tool to evaluate the safety of cyclists: Community based, macro level collision prediction models using negative binomial regression. Accid Anal Prev 2013; 61: 129-137.
[8] Saplıoğlu M, Yüzer Günay E. Investigating the effective parameters of safe bicycle route by using a survey study. Sigma J Eng & Nat Sci 2016; 7(1): 89-96.
[9] Black WR. Sustainable transportation: A us perspective. Journal of Transport Geography 1996; 4(3): 151-159.
[10] Rybarczyk G, Wu C. Bicycle facility planning using gis and multi-criteria decision analysis. Appl Geogr 2010; 30: 282-293.
[11] Mohit MA, Ali MM. Integrating gis and ahp for land suitability analysis for urban development in a secondary city of Bangladesh. Jurnal Alam Bina 2006; 8(1): 1-20.
[12] Stojkovic S, Đurdic S, Andelkovic G. Application of multi-criteria analysis and gis in ecotourism development (case study: Serbian Danube region). Glasnik Srpskog Geografskog Društva 2015; 95(1): 51-66.
[13] Serbu R, Marza B, Borza S. A spatial analytic hierarchy process for ıdentification of water pollution with gis software in an eco-economy environment. Sustainability 2016; 8(11): 1-25.
[14] Olgun R, Yılmaz T. Determination of ecologically suitable settlement areas by using gis based multi-criteria decision making analysis: The case of Nigde province. Fresen Environ Bull 2019; 28(5): 3768-3777.
[15] Snyder SA, Whitmore JH, Schneider IE, Becker DR. Ecological criteria, participant preferences and location models: A gis approach toward atv trail planning. Appl Geogr 2008; 28(4): 248-258.
[16] Angelis DI, Lee CY. Strategic investment analysis using activity based costing concepts and analytical hierarchy process techniques. Int J Prod Res 1996; 34(5): 1331-1345.
[17] Işıklar G, Büyüközkan G. Using a multi-criteria decision making approach to evaluate mobile phone alternatives. Comput Stand Inter 2007; 29(2): 265-274.
[18] Huang TCK, Chen YL, Chang TH. A novel summarization technique for the support of resolving multicriteria decision making problems. Decis Support Syst 2015; 79: 109-124.

[19] Goepel KD. Comparison of judgment scales of the analytical hierarchy process-a new approach. Int J Inf Technol Decis Mak 2017; 18(2): 445-463.
[20] Saaty TL. Transport planning with multiple criteria: The analytic hierarchy process applications and progress review. J Adv Transport 1995; 29(1): 81-126.
[21] Ludin A, Latip S. Using multi-criteria analysis to identify suitable light rail transit route. Jurnal Alam Bina 2007; 9(1): 131-142.
[22] Pogarcic I, Francic M, Davidovic V. Application of ahp method in traffic planning. 16th International Symposium on Electronics in Transport; 9-10 October 2008; Ljubljana, Slovenia. pp. 1-8.
[23] Orman A, Düzkaya H, Ulvi H, Akdemir F. Multi-criteria evaluation by means of using the analytic hierarchy process in transportation master plans: Scenario selection in the transportation master plan of Ankara. GU J Sci 2018; 31(2): 381-397.
[24] Duleba SZ, Mishina T, Shimazaki J. An analysis of the public transport's supply quality by using the analytic hierarchy process. JOMSA Conference Papers, Osaka, Japan, 2010.
[25] Nassi CD, Costa FC. Use of the analytic hierarchy process to evaluate transit fare system. Research in Transportation Economics, 2012; 36(1): 50-62.
[26] Taşkın B. Wind power plants installed in Nigde physical and economic analysis. MSc, Nigde University, Nigde, Turkey, 2013.
[27] Altay, Ö. Investigation of heavy metal content of soil from Nigde city. MSc, Erciyes University, Kayseri, Turkey, 2016.
[28] Olgun, R. Determination of strategic objectives and development of planning strategies for open and green spaces of Nigde city. Ph.d. Thesis, Akdeniz University, Antalya, Turkey, 2018.
[29] Vaidya OS, Kumar S. Analytic hierarchy process: An overview of applications. Eur J Oper Res 2006; 169: 1-29.
[30] Çol Yılmaz D, Gerçek H. Prioritization of integrated bicycle network clusters in Istanbul using analytic hierarchy process. PA J Eng Sci 2014; 20(6): 215-224.
[31] Görener Ö. A swot-ahp approach for assessment of medical tourism sector in Turkey. Alphanumeric Journal 2016; 4(2): 159-170.
[32] Saaty TL. Decision making with the analytic hierarchy process. Int J Serv Sci 2008; 1(1): 83-98.
[33] Aydi A, Abichou T, Nasr IH, Louati M, Zairi M. Assessment of land suitability for olive mill wastewater disposal site selection by integrating fuzzy logic, ahp, and wlc in a gis. Environ Monit Assess 2016; 188: 59.
[34] Wind Y, Saaty TL. Marketing applications of the analytic hierarchy process. Manage Sci 1980; 26: 641-658.
[35] Özdemir Işık B, Demir S. Integrated multi-criteria decision-making methods for the sustainability of historical-cultural structures on the Trabzon coastline. Sustainability 2017; 9(11): 2114.
[36] Zare K, Mehri-Tekmeh J, Karimi S. A swot framework for analyzing the electricity supply chain using an integrated ahp methodology combined with fuzzy-topsis. Int Strateg Manag Rev 2015; 3(1-2): 66-80.
[37] Qashqo B. Gis-based multi-criteria decision analysis for identifying water distribution points: A case study in Lapilang & Suspa regions, Nepal. MSc, University of Salzburg, Vienna, 2018.
[38] Buehler R, Pucher J. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. Transportation 2012; 39: 409-432.
[39] Forsyth A, Krizek K. Urban design: Is there a distinctive view from the bicycle? J Urban Des 2011; 16(4): 531-549.
[40] Willis DP, Manaugh K, El-Geneidy A. Uniquely satisfied: Exploring cyclist satisfaction. Transport Res 2013; 18: 136-147.
[41] Fraser SD, Lock K. Cycling for transport and public health: a systematic review of the effect of the environment on cycling. Eur J Public Health 2011; 21(6): 738-43.
[42] Iseki H, Tingstrom M. A gis bikeability/bikeshed analysis incorporating topography, street network and street connectivity. The 13th World Conference on Transport Research; 15-18 July 2013; Rio de Janeiro, Brazil. pp. 1-20.
[43] Saplıoğlu M, Aydın MM. Choosing safe and suitable bicycle routes to integrate cycling and public transport systems. J Transp Health 2018; 10: 236-252.
[44] Polat ZA, Memduhoğlu A, Hacar M, Duman H. Determining The relationship between urban growth and motor vehicle traffic: Istanbul case study. Omer Halisdemir Univ J of Eng Sci 2017; 6(2): 442-451.
[45] Çetinkaya Karafakı F. The effects of Nigde city center's active green areas to the quality of urban life. Turkjans 2016; 3(3): 184-191.
[46] Lin JJ, Liao RY. Sustainability sı: Bikeway network design model for recreational bicycling in scenic areas. Netw Spat Econ 2016; 16(1): 9-31.
[47] Erdogan R, Olgun R, Tulek B, Zaimoglu Z. Investigating the provisions of bicycle use and inclinations of riders: Konyaalti (Antalya) case study. Int J Agric Environ Res 2016; 2(5): 1368-1380.

Turkish Journal of Science & Technology
15(1), 29-35, 2020

The Shortest Road Algorithm Approach In Determining The Route Of Solid Waste Collection Vehicles: The Case Of Manisa 75. Yıl Neighborhood

Ali Osman GÖKCAN[1*], Resul ÇÖTELİ[2], Erkan TANYILDIZI[3]
[1] MCBU Turgutlu Vocational School, Electronics and Automation Technologies Department, Manisa, Turkey
[2] Fırat University, Faculty of Technology, Department of Energy Systems Engineering, Elazığ, Turkey
[3] Fırat University, Faculty of Technology, Department of Software Engineering, Elazığ, Turkey
[*1] ali.osman@outlook.com, [2] rcoteli@gmail.com, etanyildizi@firat.edu.tr[3]

(Geliş/Received: 15/10/2019; Kabul/Accepted: 20/02/2020)

Abstract: Many studies have been conducted in recent years to solve the Shortest Route Problem, and the solutions produced as a result of these studies are used in all aspects of everyday life. Package distribution of cargo companies, school services leaving students home, bank branches collecting money and forwarding it to the central bank, municipalities collecting wastes, and bread companies distributing bread to bakery markets, and similar transportation services in which vehicle routing is needed make use of these solutions. Some algorithms related to the Shortest Route Problem were examined in this article, and among these, the Dijkstra Algorithm was applied to a real example problem. In this example, 75. Yıl Neighborhood of Yunus Emre District of Manisa Province was modeled with the Graph Data Model, and the best road destination for solid waste collection vehicles was determined. Based on the results obtained, it was determined that the Dijkstra Algorithm identified the shortest route successfully.

Key words: Shortest road, vehicle routing, shortest path algorithm, graph model.

Katı Atık Toplama Araçlarının Güzergâh Tespitinde En Kısa Yol Algoritması Yaklaşımı: Manisa 75.Yıl Mahallesi Örneği

Öz: Son yıllarda en kısa yol probleminin çözümüne yönelik birçok çalışma yapılmakta ve bu çalışmalar neticesinde üretilen çözümler hayatın her alanında kullanılmaktadır. Kargo şirketlerinin paket dağıtımı, okul servislerinin öğrencileri evine bırakması, banka şubelerden paraların toplanıp merkez bankaya iletilmesi, belediyelerin atıkları toplaması ve ekmek fırının marketlere ekmek dağıtımı gibi araç rotalamanın gerekli olduğu her türlü taşımacılık hizmeti bu çözümlerden faydalanmaktadır. Bu makalede en kısa yol problemine ilişkin bazı algoritmalar incelenmiş ve incelenen algoritmalardan Dijkstra algoritması gerçek bir örnek problem üzerinde gerçekleştirilmiştir. Ele alınan örnekte Manisa ili Yunus Emre İlçesi 75. Yıl Mahallesinin belirlenen bir kesiti graf veri modeli ile modellenmiş ve katı atık toplama araçları için en uygun yol güzergâhı tespit edilmiştir. Elde edilen sonuçlardan Dijkstra algoritmasının en kısa yolu oldukça başarılı bir şekilde tespit ettiği görülmüştür.

Anahtar kelimeler: En kısa yol, araç rotalama, en kısa yol algoritması, graf modeli.

1. Giriş

Parallel to the developing technology, the solutions that are produced as a result of the Shortest Route Problem are widely used in all areas of everyday life. Package distribution of cargo companies, school services leaving students home, bank branches collecting money and forwarding it to the central bank, municipalities collecting wastes, bread companies distributing bread to bakery markets, and similar transportation services in which Navigation and GPS are needed make use of these solutions. There might be many ways to travel from a source point to a specific destination. Shortcut discovery algorithms target to find the shortest road to navigate between these two points [1].

The logic of shortcut discovery algorithms is based on reaching the target from different nodes with the shortest route, and this algorithm constitutes the basis of vehicle routing problems. Because, in Shortest Route Problems, the purpose is to calculate the shortest route from a specific node or to all other nodes [2].

Vehicle Routing Problem (VRP) has been applied in many fields. Vehicles need to be routed especially in the transportation business. Some of the transportation works in which VRP can be used are listed as follows [3]:

1- Routing student or worker employee vehicles,

* Corresponding author: ali.osman@outlook.com. ORCID Number of authors: [1] 0000-0001-8477-1520, [2] 0000-0002-7365-4318, [2] 0000-0003-2973-9389

2- Collection solid wastes,
3- Package collection and distribution of cargo companies,
4- Distribution of the food and beverage products of companies to grocery stores or dealers,
5- Collection of the money from bank branches at the end of the day,
6- Food distribution of food companies to institutions like workplaces and schools,
7- Distribution of the milk produced in dairies.

Solid waste is defined as the production of unwanted items littered after use. In addition, it can also be defined as unwanted solid items after any activity discarded by the community [4]. In general, solid wastes can be collected in three groups. The first group consists of materials that are now unwanted like scrap metal, glass, cans, paper, plastic, wood and similar materials, which can be useful objects that may be recycled or accumulated [5]. The second group consists of wastes that can be separated for destruction or burned for energy recovery [6]. The third group consists of abandoned and naturally-discarded waste-like materials that are unwanted [4].

The fast increase in solid wastes is a critical problem not only in our country but also in most countries all around the world because this poses serious risks for human health. For this reason, effective solid waste management must be planned; and for this, it is necessary to design and develop appropriate collection route plans [7]. For municipalities, the collection and transportation of solid wastes have a high share among many other costs [8].

In cities, solid wastes are usually produced in residential areas and commercial and industrial facilities. For this reason, the collection of solid wastes is requested because of the density of the population and human activities. As the distribution and quantity of waste production increases, collection logistics becomes more complex. Although these problems always exist, 50-70% of the total amount of the money spent to collect, transport and dispose of solid wastes is used in the collection step [9]. For this reason, the collection process must be improved as it may affect the total costs and durations of solid waste management. One of the ways of improving waste collection is to guide garbage collectors for choosing the shortest routes during collection [7]. For this purpose, the routes of solid waste collection vehicles must be examined, and collection and transport routes must be created with the lowest cost and minimum time and distance. In our present day, many municipalities and waste collection companies perform waste collection operations with intuitive behaviors of their employees completely at their initiatives considering relevant studies [8].

The shortest route problem within the vehicle routing problem was dealt with in the present study, and the Dijkstra Algorithm, which is among the algorithms used to solve this problem, was explained with a real example. The shortest road was found among the solid waste containers located in a pre-defined section of 75. Yıl Neighborhood of Yunus Emre Municipality of Manisa Province.

The continuation of the present study is organized as follows. Similar studies in the literature were mentioned in the second part of the study. The area where the study was carried out, the data used, the analysis of the data used in the graph model, and the process of determining the solid waste collection routes are explained in the third part. Finally, the results of the study were evaluated, and what can be done in future studies are mentioned in the fourth part.

2. Relevant Studıes

Many studies were conducted in the past to model and solve the Shortest Route Problem. Some algorithms were developed by Dijkstra [10], Bellman-Ford [11], Floyd-Warshall [12], Johnson [13], Martin [14], Bhandari [15], and Dreyfus [16]. On the other hand, Ramakrishna and Wook proposed a genetic algorithm to solve this problem [17]. Singh, Rathi and Haris used the Dijkstra algorithm to carry out solid waste management. The method proposed in this study was applied to Kanpur Province of India, and the shortest path algorithm was performed by using the C programming language for the collection of solid wastes [18]. In the study conducted by Bayzan, he performed the simulation of the vehicle routing problem in a geographic area adapted to the graph data model with C# programming language. With the help of this application, the effect of the waiting times of the vehicles on the cost of solid waste management was examined [3]. In their study, Büke and Erturaç conducted the shortest route analysis, service analysis and optimal location analysis for Sakarya University Esentepe Campus by using Dijkstra Algorithm, and mapped it. They also prepared a web-based page to enable users to easily access the prepared database [19]. Selim determined a trekking route by conducting the most appropriate route analysis based on Dijkstra Algorithm by working with the logic of graph theory. The trekking route of Güllük Mountain National Park, which hosts the Termessos Ancient City, was determined to find the shortest route and the lowest slope [20].

Gene and Lin made use of a genetic algorithm with random key-based encryption to find a solution to the shortest route problem. In addition, they also developed a new algorithm by using arithmetic crossover, mutation and transition operator [21]. In the study conducted by Tomis, Martinovic, Slaninova, Rapant and Vondrak, they planned a time-based route for the highways of Czech Republic. They calculated the dynamic travel route and duration with the algorithm they used. Alternative routes were developed against possible traffic problems [22]. In the study that was conducted by Norhafezah, Nurfadzliana and Megawati, they simulated solid waste collection paths with a mathematical model of Dijkstra Algorithm. With the algorithm that was applied to Shah Alam, which is the capital of Selangor Province of Malaysia, they provided route optimization by shortening the distance traveled by solid waste collection vehicles [7]. In their study, Cooke and Halsey brought a solution to the shortest route problem by creating a modified form of Bellman's Iteration Scheme [23]. Cowen, Jensen, Hendrix, Hodgson and Schill worked on an econometric route detection model, which explored potential railway routes. They showed that the results with approximate values from Geographic Information Systems could be balanced with the data that had high resolution and accuracy by using the high-accuracy numerical height model [24].

3. MATERIAL AND METHOD

With the fast increase in the population of the world, urbanization also increased, and new settlements came into being. With this development, transportation networks among settlements and transportation alternatives that can be used in the transportation network also increased. The target was not only to reach another point from one point but also to reach it as soon as possible and with the lowest cost. Several methods were made use of and modeling was made to find the distance among settlements and to determine the shortest route between them. One of these models is Graph Theory [3].

Many sciences and social sciences problems can be solved with Graph Theory. Problems like mapping, traveler vendor, widest independent cluster, distribution, central placement, shortest route, and network flow problem are some of the problems, which might be solved with the Graph Theory [25]. Especially some problems can be modeled by adapting them to graphs. Transportation networks among settlements can be solved easily when modeled with Local Area Networks (LAN), Wide Area Networks (WAN) and similar networks [3].

Modeling a problem with edge and nodes and showing this model in the form of a diagram refers to the basic use of the Graph Theory. In real life, settlements, intersections, warehouses, shops, and schools show the node points on the graph; and the roads connecting these to each other show the edges of the graph. One of the first problems to which the Graph Theory was applied was the Königsberg Bridge problem that was solved by Leonhard Euler in 1736. The Königsberg Bridge and the Graf Model are given in Figure 1.

Figure 1. Königsberg Bridge and Graph Model

Yunus Emre District was established as a result of the separation of the provincial center to two districts when Manisa Province has the metropolitan status [26]. The Dijkstra Algorithm, which is used to solve the vehicle routing problem, or in other words, to detect solid waste collection route, was applied to 75. Yıl Neighborhood among the 87 neighborhoods of this district. The necessary data on the solid waste collection system of the neighborhood was obtained from the Sanitation Works Directorate of Yunus Emre Municipality. This neighborhood has 3 sub-districts. The solid waste collection sub-districts of the 75. Yıl Neighborhood are shown in Figure 2. The first district was the Sanayi Sub-District. Here, solid wastes are collected here every weekday

except Sunday. The second sub-district is the area where the main artery is located. The third sub-district is the residential area where household solid wastes are found, and solid wastes are collected every weekday. Although the route designated for the collection of solid wastes in this area was developed and improved every period, it is generally determined according to the initiative of the driver of the waste collection vehicle and other staff.

Figure 2. 75. Yıl Neighborhood Sub-Districts (Manisa/Yunus Emre)

The locations of the containers and the digital data of the street are given in Figure 3. Here, the containers form the nodes of the graph structure; and are named in letters like A, B, C, and D. The roads between are the edges of the graph. In the present study, the data on the distances of the streets were obtained from Google Maps. The raw data that contained the distance was converted into a graph data type. The graph data were applied as a weighted neighboring matrix with a traffic direction among the nodes.

Figure 3. Solid Waste Container Locations and Digital Data

When the nodes given in Figure 3 are evaluated, it is seen that there might be many alternatives to reach the CA node from BM. It is possible to denote some of these as follows:

Preference: BM > BE > CA,
Preference: BM > AU > AN > CA,
Preference: BM > AN > CA,
Preference: BM > AU > AT > AM > CA,
Preference: BM > BL > BO > BE > CA

A person who chooses No. 1 will reach his/her target in the shortest way with the lowest cost; however, a person who chooses routes 4 or 5 will reach his/her target in the longest way. Another example may be given in the section in Figure 4. As shown in Figure 4, a vehicle can follow different routes to reach AO from AL. However, with the help of the algorithm used, the vehicle will use the shortest way by selecting the fifth option.

Figure 4. Route Choices between AL - AO Nodes

Figure 5. The Shortest Route Determined

The destination obtained by using the Dijkstra Algorithm to collect domestic solid wastes is given in Figure 5. According to this result, if a vehicle starting from A Point follows a route by visiting all knots as follows A > B

> E > C > D > U > AA >> CB > CC > AC > AB > AD > AG > AF > AE > AH > AI > AJ > AK > AL > AO > AS > AY > AV > AZ > BG > BD > BF > BH > BI > BR > BP > BL > BK > BJ > AU > AT > AM > AN > BM > BE > CA > BZ > BY > Y > O > N > M > L > V > BV >BU > BS > BO > BN > BT > Z > J > K > H> I > F > G > P > S > T, it will have followed the shortest route.

4. Results

The Dijkstra Algorithm, which is among the algorithms used to solve the shortest route problem, was applied to a real example in this article. In the study, the purpose was to find the shortest route among the solid waste containers located in a predefined section of 75. Yıl Neighborhood of Yunus Emre Municipality. The study was conducted with the help of real data from the residential area defined in the study. Firstly, the identified settlement was adapted to the graph model. Then the Dijkstra Algorithm was applied to solve the shortest route problem on a 66-node graph. In addition, the data like slope, traffic density, instant road status, and vehicle capacity were neglected in this study.

In case the solid waste collection and transportation operations were carried out by using the routs that were at the initiatives of vehicle drivers without depending on a specific route, the total road traveled was 3348 km; however, this value was reduced to 2016 km as a result of the optimization study. In this case, approximately 40% of fuel and time were saved. As a result, it is seen that if municipalities use route determination algorithms in solid waste collection and transport operations, the time spent on collecting solid wastes, workforce, and expenses will decrease. In addition, it is also considered that the savings will increase with route optimization, less vehicle use, reduced usage-related costs like maintenance, repair, spare parts and labor costs. This study was conducted for a limited area with a single algorithm and with some omissions. In future studies, routing for an entire district or city will be done by using different shortest route algorithms and considering the omissions. In addition, interactive route applications may be developed by using Geographic Information Systems and instant road data.

References

[1] Rohila K, Gouthami P, Priya M. Dijkstra's shortest path algorithm. International Journal of Innovative Research in Computerand Communication Engineering 2014; 2(10): 6139-6144.
[2] Kalkancı Ç. The organization of emergency response teams fighting against winter conditions of natural disaster management. Istanbul University Institute of Science, PhD Thesis, Istanbul, 2014.
[3] Bayzan Ş. Determination of vehicle routes by using the shortest road algorithms and simulation in the .net environment. Pamukkale University Institute of science, Master Thesis, Bursa, 2005.
[4] Desa A, Abd Kadir N, Yusooff F. Environmental awareness and education: A key approach to solid waste management (SWM) – A case study of a University in Malaysia. Waste Management - An Integrated Vision 2012; pp.101-111.
[5] Fazeli A, Bakhtar F, Jahanshaloo L, Che Sidik N, Esfandyari B A. Malaysia's stand on municipal solid waste conversion to energy: A review. Renewable and Sustainable Energy Reviews 2016; 58: 1007-1016
[6] Ma J, Hipel K. Exploring social dimensions of municipal solid waste management around the globe - A systematic literature review. Waste Management 2016; pp.1-10
[7] Norhafezah K, Nurfadzliana A. H, Megawati O. Simulation of municipal solid waste route optimization by Dijkstra's algorithm. Journal of Fundamental and Applied Sciences 2017; ISSN 1112-9867.
[8] Çakır ME, Yetiş A, Yeşilçınar M, Ulukavak M. Determination of the optimum routes and performing of the spatial distribution maps for solid waste using GIS: A case of study from Suruç (Şanlıurfa), BEU Journal of Science 2019; 8 (2), 595-603.
[9] Billa L, Pradhan B, Yakuup A. GIS routing and modelling of residential waste collection for operational management and cost optimization. Pertanika Journal Science & Technology 2014; 22(1): 193-211.
[10] Dijkstra EW, A Note on two problems in connection with graphs. Numerische Mathematik 1959; 1(1): 269–271.
[11] Bellman E. On a routing problem. Appl Math 1958; Vol 16: 87–90
[12] Floyd RW. Algorithm 97: Shortest Path. Communications of the ACM 1962; 5(6): p.345
[13] Johnson DB. Efficient algorithms for shortest paths in sparse networks. Journal of the ACM 1977; 24(1)
[14] Martins EDQV, Dos Santos JLE. The labelling algorithm for the multiobjective shortest path problem. Technical Report, Departamento de Matematica, Universidade de Coimbra, Portugal 1999; 1-24
[15] Bhandari R. Optimal physical diversity algorithms and survivable networks, In Proceedings Second IEEE Symposium on Computer and Communications, 1997; pp. 433–441.
[16] Dreyfus S. An appraisal of some shortest path algorithms", Operations Research 1969; 17(3): 395–412.
[17] Wook C, Ramakrishna RS. A Genetic algorithm for shortest path routing problem and the sizing of populations, IEEE Transactions on Evolutionary Computation 2002; 6(6): 566-579.

[18] Singh G, Singh B, Rathi S, Haris S. Solid waste management using shortest path algorithm. International Journal of Engineering Science Invention Research & Development 2014; 1(2), e-ISSN: 2349-6185.
[19] Büke CO, Erturaç MK. Network analysis method investigation of Sakarya University Campus Esentepe and web-based report. Nature Sciences (NWSANS) 2016; 11(4): 14-25.
[20] Selim S. Modeling of hiking routes using the least cost path algorithm. Süleyman Demirel University Journal of Natural and Applied Sciences 2019; 23(2): 617-624.
[21] Gen M, Lin L. A New approach for shortest path routing problem by random key- based GA. Genetic and Evolutionary Computation Conference 2006; Vol 1 and 2 , 1411–1412.
[22] Tomis R, Martinovič J, Slaninová K, Rapant L, Vondrák I. Time-Dependent route planning for the highways in the Czech Republic. 14th Computer Information Systems and Industrial Management (CISIM), Sep 2015, Warsaw, Poland. pp.145-153.
[23] Cooke KL, Halsey E. The shortest route through a network with time-dependent ınternodal transit times. Journal of Mathematical Analysis and Applications 1966; 14(3): 493–498.
[24] Cowen DJ, Jensen JR, Hendrix C, Hodgson M, Schill SR. A GIS assisted rail construction econometric model that incorporates LIDAR data. Photogrammetric Engineering and Remote Sensing 2000; 66(11): 1323-1328.
[25] Christofides N. Graph theory an algorithmic approach. Acadeınic Press, London, 1986
[26] http://www.manisa.gov.tr/yunusemre (16.01.2020)

Turkish Journal of Science & Technology
15(1), 37-41, 2020

The Estimation of Total Lengths in the Earlier Ages of Mirror Carp (*Cyprinus carpio* L., 1758) Using Back-Calculation Methods

Can Kaan TEKSAR[1], Metin ÇALTA[2*]
[1] İzmir Food Control Laboratory Directorate, İzmir, Turkey
[2] Fisheries Faculty, Firat University, Elazig, Turkey
mcalta@firat.edu.tr

(Geliş/Received: 13/02/2020; Kabul/Accepted: 20/02/2020)

Abstract: In this study, the total lengths of the smaller mirror carp, that could not be caught, were estimated with the back-calculation methods using the scale and otolith lengths of larger fish. For this purpose, 50 fish samples were captured from Keban Dam Lake during October-November 2017 and October-December 2018. They were 4-9 in ages, 31.2-47.2 cm in total length and 442.5-1497.8 g in weight. The age of fish and the lengths of each age ring on scale and otolith were determined by using LAS V4.8 image analysis program attached to Leica S8APO microscope. The total lengths of fishes were calculated using Fraser-Lee and Dahl-Lea back-calculation equations for scale and otolith respectively. The mean total lengths calculated for all age groups from scale radius lengths and otolith lengths were L_1:16.1, L_2:21.3, L_3:27.3, L_4:30.5, L_5:35.7, L_6:36.7, L_7:42.9, L_8:45.3 ve L_9: 45.9 cm and L_1:14.2, L_2:17.8, L_3:22.8, L_4:25.5, L_5:33.1, L_6:36.0, L_7:41.0, L_8:41.4 ve L_9:44.0 cm respectively. In all age groups, total lengths of fish calculated from otolith lengths were found lower than total lengths of fish calculated from scale lengths.

Key words: Mirror carp, *Cyprinus carpio*, back calculation, fish length, scale, otolith

Geri Hesaplama Yöntemleri Kullanılarak Aynalı Sazanın (*Cyprinus carpio* L., 1758) Daha Önceki Yaşlardaki Toplam Boylarının Tahmini

Öz: Bu çalışmada, büyük boy aynalı sazanların (*Cyprinus carpio* Linnaeus, 1758) pul ve otolit ölçümlerinden geri hesaplama yöntemleriyle yakalanamayan küçük balıkların toplam boyları hesaplandı. Bu amaçla, yaşları 4-9, toplam boyları 31.2-47.2 cm ve ağırlıkları 442.5-1497.8 g arasında olan 50 adet balık örneği (20 adet Ekim-Kasım 2017 ve 30 adet Ekim-Aralık 2018 aylarında) Keban Baraj Gölünden yakalandı. Balıkların yaşı ile her bir yaş halkasına ait pul ışın boyları ve otolit boyları bilgisayar destekli Leica S8APO mikroskop yardımıyla LAS V4.8 imaj analiz programı kullanılarak belirlendi. Balık boyları pul için Fraser-Lee, otolit için ise Dahl-Lea geri hesaplama yöntemleri kullanılarak hesaplandı. Tüm yaş grupları için pul ışın boyları kullanılarak hesaplanan toplam balık boyu ortalama değerleri L_1:16,1; L_2:21,3; L_3:27,3; L_4:30,5; L_5:35,7; L_6:36,7; L_7:42,9; L_8:45,3 ve L_9: 45,9 cm olarak, otolit boyları kullanılarak hesaplanan toplam balık boyu ortalama değerleri ise L_1:14,2; L_2:17,8; L_3:22,8; L_4:25,5; L_5:33,1; L_6:36,0; L_7:41,0; L_8:41,4 ve L_9:44,0 cm olarak hesaplandı. Tüm yaş gruplarında otolit boylarından hesaplanan toplam balık boyları, pul ışını boylarından hesaplananlara göre daha düşük bulundu.

Anahtar Kelimeler: Aynalı sazan, *Cyprinus carpio*, geri hesaplama, balık boyu, pul, otolit

1. Introduction

There are many resources on the determination of age and growth in fish [1-9]. In fishery studies, it is important to have a sufficient number of fish samples at different sizes for accurate estimation of age and growth. However, sometimes it is not always possible to catch smaller fishes depending on the fishing method used and preferring a different habitat by the smaller fishes. Furthermore, fisheries legislation prohibits the capture of smaller fishes in order to protect the fish stocks. These situations make it difficult to obtain information about the age and growth of smaller fishes in a population. For this reason, the age and growth of smaller fishes, that cannot be sampled, are scientifically estimated from the measurements of some bony structures of the larger size individuals of the same species by using a suitable back-calculation method [10-15]. The critical comparison of back-calculation methods used for estimation of fish growth have been done [16-20].

Mirror carp is the most common species of the carp family in the inland waters of Turkey. The various aspects of mirror carp population in Keban Dam Lake were investigated by researchers [21-24]. Some studies have been done using the back-calculation methods to estimate fish lengths at previous ages on *Capoeta trutta* [7], *Barbus rajanorum mystaceus* [25], *Acanthobrama marmid* [26] and *Luciobarbus mystaceus* [27] from Keban Dam Lake

* Corresponding author: mcalta@firat.edu.tr, ORCID Number of authors: [1] 0000-0003-0554-2725, [2] 0000-0002-1652-8972

and on *Capoeta umbla* from Hazar Lake [28]. However, there is no any research on the estimation of fish lengths at previous ages the back-calculation methods. Therefore, the main aim of this study estimates the total lengths of mirror carp *(Cyprinus carpio)* species from Keban Dam Lake at previous ages by using two different back calculation methods for scale and otolith measurements. In addition, it will be determined that which bony structure is more reliable for back-calculation in the mirror carp.

2. Materials and Methods

In this study, 50 bigger mirror carp species (20 in the October-November 2017 and 30 in the October-December 2018) from Keban Dam Lake were caught from the area close to the dam embankment (Figure 1) by a commercial fisherman.

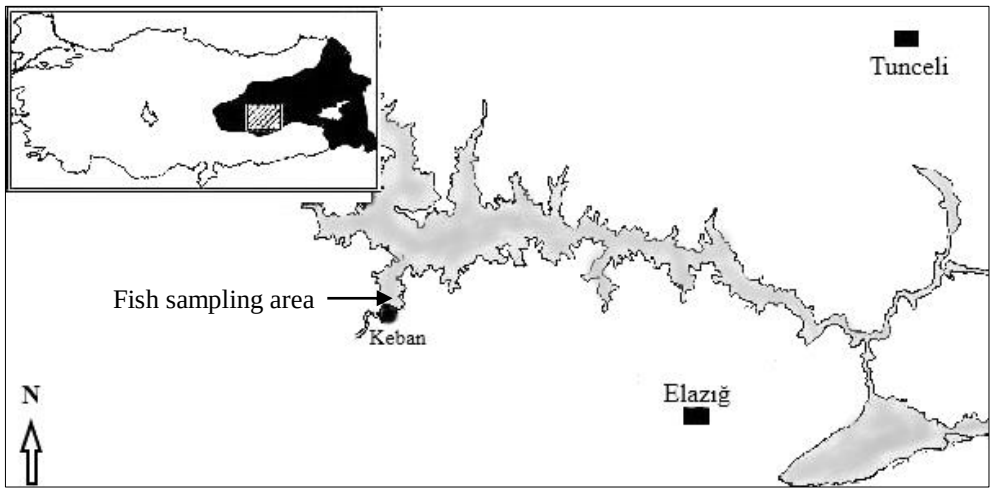

Figure 1. Fish sampling area in Keban Dam Lake

Fish samples were brought to the fish systematics laboratory of Fisheries Faculty of Firat University and the following procedures were applied.

- Total lengths and body weights were determined nearest ± 1 mm and 0.1 g precision respectively.
- Sagittal otoliths and some scales from just under the first dorsal fin were removed, cleaned and kept in 96% alcohol for microscopic examination.
- The ages of fishes were determined from both sagittal otoliths and scales.
- The lengths of each age ring (annulus) were measured nearest 0.001 mm precision by using image analysis software (LAS V4.8) connected to Leica S8APO microscope.
- The total lengths of fish at previous ages were calculated using Fraser-Lee back-calculation equation [29, 30] for scale and Dahl-Lea back-calculation equation [31, 32] for otoliths given below.

Fraser-Lee back-calculation equation: $L_n = a + (L-a) \times (S_n/S)$

Dahl-Lea back-calculation equation: $L_n = L \times (S_n/S)$

Where;

L_n= back-calculated length of the fish at age "n"

L= total length of fish at the time of capture

S_n= length of scale radius or otolith at age "n"

S= length of scale radius or otolith of fish at the time of capture

a= intercept from the regression of body length on mean scale length

The results were statistically examined by means of SPSS Ver.22 programme (IBM Cooperation).

3. Results

In this study, 50 mirror carp species from Keban Dam Lake were used. They were 4-9 in age group, 31.2-47.2 cm in total length and 442.5–1497.8 g in body weight. The "a" value in Fraser-Lee back-calculation equation was found as 22.7 mm by using the regression analysis between total length and scale radius length of fish. This

value indicates the first formation of scale started while the fish were 22.7 mm in total length. The total fish lengths of the previous ages estimated by using the back calculation methods from the scales and otolith measurements are given in Table 1 and Figure 2. The total fish lengths at previous ages calculated from scale and otolith measurements were found smaller for all age groups compared with the observed total lengths (Figure 3).

Table 1. The distribution of 50 fish samples according to age groups, measured total lengths (TL) and total lengths (L_1-L_9) calculated using back-calculation methods from scale and otolith measurements.

Age groups	N	TL (cm)	Total lengths calculated from scale									Total lengths calculated from otolith								
			L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
IV	5	32.1	16.7	19.0	24.7	26.9						12.1	15.5	19.3	24.7					
V	9	37.9	16.9	21.9	25.2	27.8	35.9					13.2	16.3	20.4	25.1	32.6				
VI	17	41.6	16.5	22.1	27.7	30.2	35.8	37.0				14.5	18.0	24.1	26.0	32.4	35.0			
VII	10	43.2	15.7	21.1	28.4	32.1	34.6	36.2	43.6			13.6	17.4	22.9	24.7	33.4	35.6	40.6		
VIII	6	45.8	15.4	22.3	29.1	32.8	35.8	37.5	42.8	44.9		15.2	20.0	24.1	26.6	32.8	36.0	41.5	41.6	
IX	3	47.2	15.2	21.5	28.9	33.1	36.2	35.9	42.2	45.6	45.9	16.5	19.8	26.2	26.1	34.5	37.4	40.9	41.2	44.0
Mean			16.1	21.3	27.3	30.5	35.7	36.7	42.9	45.3	45.9	14.2	17.8	22.8	25.5	33.1	36.0	41.0	41.4	44.0
SD			0.7	1.2	1.9	2.6	0.6	0.7	0.7	0.5	-	1.6	1.8	2.6	0.8	0.8	1.0	0.5	0.3	-

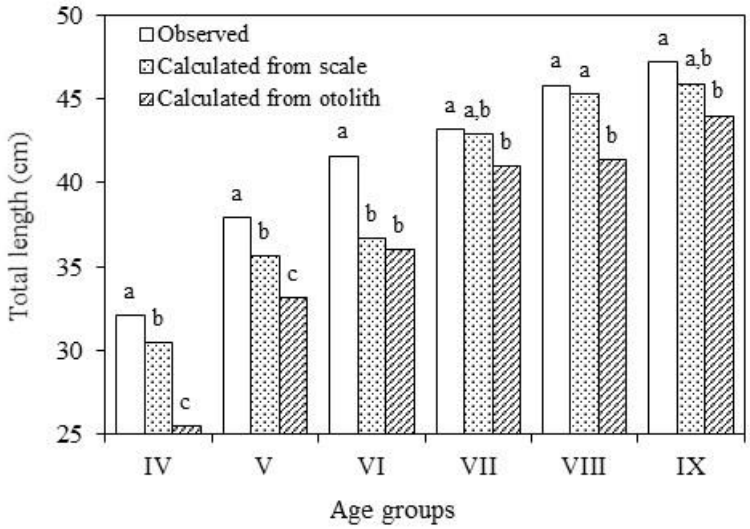

Figure 2. Observed and back-calculated total lengths of mirror carp from Keban Dam Lake. Different letters above the columns indicate significant differences ($P < 0.05$; ANOVA Duncan's Multiple Range Test).

As can be seen in Figure 3, while the gap between the measured values line and the calculated values lines is wider in the first early ages, this gap decreases in the older ages (Figure 3).

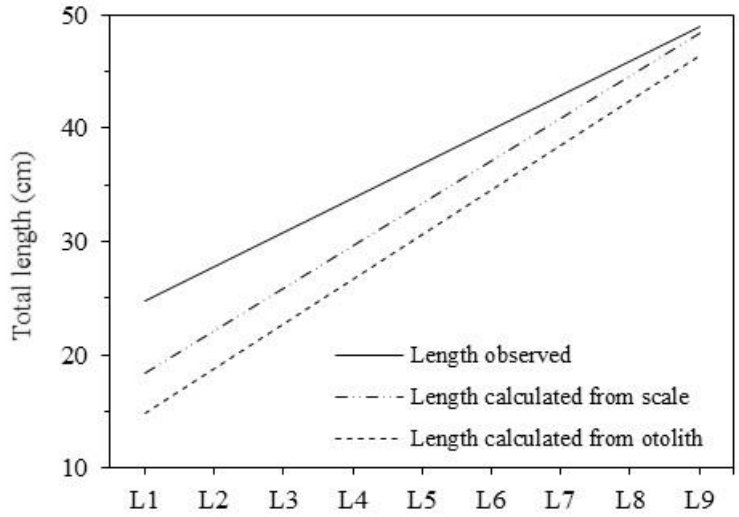

Figure 3. The comparison of total lengths (L_1-L_9) observed and calculated from scale and otolith measurements in mirror carp from Keban Dam Lake.

4. Discussion

In this study, 50 samples belong to mirror carp with 4-9 in age, 31.2-47.2 cm in length and 442.5-1497.8 g in weight were examined to estimate the total lengths of fish at earlier ages by using back-calculation methods from scale and otolith measurements. The total lengths of mirror carp at earlier ages calculated from scale and otolith measurements were found to be very close to each other. However, the measured total lengths of mirror carp for all age groups were always determined to be higher than the total lengths of mirror carp calculated from both scale and otolith measurements. Some studies have also been showed that measured total lengths were always higher than calculated lengths using back-calculation methods [7, 25, 26, 28]. In addition, they were showed that the gap between calculated and measured lengths increased when earlier annuli measurements of scales and otoliths obtained from older fishes were used for back-calculation. The reason for this is that the bony structures measured do not show the same growth rate as the fish during the life of fish. In addition, there is a difficulty in the age readings and measurements of annuli lengths of bony structures taken from the older individuals.

In conclusion, back-calculation methods provide an important advantage to estimate the length of smaller fishes that cannot normally be captured.

References

[1] Lagler KF. Freshwater Fishery Biology, WM.C. Brown Co., Publ., Dubuque, Iowa, 1956.
[2] Chugunova NI. Age and Growth Studies in Fish, Israel Prog. Sci. Transl. No: 610, Natl. Sci. Found Washington, D.C., 1963.
[3] Brothers EB, 1987. Methodological approaches to the examination of otoliths in aging studies. In: Age and Growth of Fish, Summertfelt, R.C. and G.E. Hall (Eds.), Iowa State University Press, 319-330.
[4] Çelikkale MS. Balık Biyolojisi, Karadeniz Teknik Üniversitesi Sürmene Deniz Bilimleri ve Teknolojisi Yüksekokulu Yayınları, No: 101, Trabzon, 1991.
[5] Erkoyuncu İ. Balıkçılık Biyolojisi ve Populasyon Dinamiği, Ondokuz Mayıs Üniversitesi. Yayınları, No: 95, Sinop, 1995.
[6] Şen D, Aydın R, Çalta M. Relationships between fish length and otolith length in the population *Capoeta umbla* (Heckel, 1843) of inhabiting Hazar Lake, Elazığ, Turkey. Arch. Polish Fish., 2001; 9: 267-272.
[7] Aydın R, Çalta M, Şen D. Age and growth of *Capoeta trutta* (Pisces: Cyprinidae) from Keban Dam Lake, Turkey. Arch. Polish Fish., 2003; 11(2): 237-243.
[8] Eroğlu M, Şen D. Otolith size-total length relationship in spiny eel, *Mastacembelus mastacembalus* (Banks & Solander, 1794) inhabiting in Karakaya Dam Lake. Journal of FisheriesSciences.com, 2009; 3: 342-351.
[9] Geldiay R, Balık S. Türkiye Tatlısu Balıkları, Ege Üniversitesi Fen Fakültesi Kitaplar Serisi, No: 46, Bornova/İzmir, 2009.

[10] Miller GL, Nelson R. Goldeye, *Hiodon alosoides*, in Lake Oahe: Abundance, Age, Growth, Maturity, Food and the Fishery, 1963-69. Technical Papers of the U. S. Fish and Wildlife Service. United States Department of the Interior Fish of Fish and Wildlife Service, Washington, pp: 13, 1974.
[11] Nelson WR. Age, Growth, and Maturity of Thirteen Species of Fish from Lake Oahe during the Early Years of Impoundment, 1963-68. Technical Papers of the U. S. Fish and Wildlife Service. United States Department of the Interior Fish and Wildlife Service, Washington, pp: 30, 1974.
[12] Starostka VJ, Nelson WR. Age, Growth, Sexual Maturity, and Food of Channel Cat-fish in Central Lake Oahe, 1968-69. Technical Papers of the U. S. Fish and Wildlife Service. United States Department of The Interior Fish and wildlife Service Washington, 1974.
[13] Tanyolaç J. Age and Growth of Carp, *Cyprinus carpio* L. in Lake Eymir, Ankara. Communications de la Faculte des Sciences de I' Universite d'Ankara, Tome 23, Serie, C3: zoologie, pp: 12, 1979.
[14] Duncan KW. On the back-calculation of fish lengths; modifications and extensions to the Frasier-Lee equation. J. Fish Biol., 1980; 16: 725-730.
[15] Bartlett JR, Randerson PF, Williams R, Ellis DM. The use of analysis of covariance in the back-calculation of growth in fish. J. Fish Biol., 1984; 24: 201-213.
[16] Francis RICC. Back-calculation of fish length: a critical review. J. Fish Biol., 1990; 36: 883-902.
[17] Zivkov M. Critique of proportional hypotheses and methods for back-calculation of fish growth. Environ. Biol. Fish., 1996; 46: 309-320.
[18] Vigliola L, Meekan MG. The back-calculation of fish growth from otoliths. In tropical fish otoliths: information for assessment, management and ecology. Edited by B.S. Green, B.D. Mapstone, G. Carlos, and G.A. Begg. Springer, Heidelberg, Germany, 2009, pp. 174–211.
[19] Vigliola L, Harmelin-Vivien M., Meekan MG. Comparison of techniques of back-calculation of growth and settlement marks from the otoliths of three species of Diplodus from the Mediterranean Sea. Canadian Journal of Fisheries and Aquatic Sciences, 2000; 57(6): 1291-1299.
[20] Wilson JA, Vigliola L, Meekan MG. The back-calculation of size from otoliths: validation and comparison of models at an individual level. J. Exp. Mar. Biol. Ecol., 2009; 368: 9-21.
[21] Çalta M, Ural MŞ. Acute toxicity of the synthetic pyrethroid deltamethrin to young mirror carp, *Cyprinus carpio*. Fresen. Environ. Bull., 2004; 13: 1179-1183.
[22] Ural MŞ, Çalta M. Acute toxicity of dichlorvos (DDVP) to fingerling mirror carp, *Cyprinus carpio* L. B. Environ. Contam. Tox., 2005; 75: 368-373.
[23] Güç G. Keban Baraj Gölü (Elazığ)'nde yaşayan aynalı sazan (*Cyprinus carpio* Linnaeus, 1758)'ın üreme biyolojisi. Fırat Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 59 s., Elazığ, 2006.
[24] Sayın B, Çalta M. Keban Baraj Gölü'nde yaşayan aynalı sazan (*Cyprinus carpio* Linnaeus, 1758)'ın otolit biyometrisi. Fırat Üniv. Fen Bilimleri Dergisi, 2017; 29: 27-32.
[25] Şen D, Yılayaz Ö. Keban Baraj Gölünde yaşayan *Barbus rajanorum mystaceus* (Heckel, 1843)' ün geri hesaplama yöntemiyle uzunluklarının belirlenmesi. G.Ü. Gazi Eğitim Fakültesi Dergisi, 2001; 21(1): 53-57.
[26] Şen D, Aydın R. Keban Baraj Gölü'nde yaşayan *Acanthobrama marmid* Heckel, 1843'ün geri hesaplama metodu ile uzunluklarının belirlenmesi. G.Ü. Gazi Eğitim Fakültesi Dergisi, 2001; 21(1): 47-51.
[27] Teksar CK, Çalta M. Back-Calculation of total lengths of *Luciobarbus mystaceus* (Pallas, 1814) from scale and otolith measurements. Turkish Journal of Science and Technology, 2019; 14(1): 33-36.
[28] Şen D, Aydın R, Çalta M. Back-calculation of fork lengths of *Capoeta umbla* (Pisces: Cyprinidae) from otolith lengths. Pak. J. Biol. Sci., 2002; 5(4): 506-508.
[29] Fraser CM. Growth of spring salmon. Transactions of the Pacific Fisheries Society, 1916; 29-39. Fraser CM. Growth of spring salmon. Transactions of the Pacific Fisheries Society, 1916; 29-39.
[30] Lee RM. A review of the methods of age and growth determination in fishes by means of scales. Fisheries Investigations, Series II, Marine Fisheries, Great Britain Ministry of Agriculture, Fisheries and Food, 1920; 4(2): 1-35.
[31] Dahl K. The scales of the herring as a means of determining age, growth and migration. Reports on Norwegian Fishery and Marine Investigations, 1907; 2(6): 1-39.
[32] Lea E. On the methods used in the herring investigations. Publications de Circonstance, Conseil Permanent International pour I'Exploration de la Mer. 1910; 53: 7-25.

Turkish Journal of Science & Technology
15(1), 43-48, 2020

Monte Carlo Study of XSO4 (X= Ba, Be, Ca, Mg, and Sr) Response to Common Particles (Proton, Electron, and Neutron): A Shielding Application

Serkan GÜLDAL[1*]
[1] Department of Physics, Faculty of Arts and Sciences, Adiyaman University, Adiyaman, Turkey
*sguldal@adiyaman.edu.tr

(Geliş/Received: 13/02/2020; Kabul/Accepted: 20/02/2020)

Abstract: Subatomic particles are continuously studied by researchers parallel to developing detectors and theories. In experiments, accelerated particles penetrate the considered target and even shielding. Thus, the protection of researchers and equipment from high energy particles is an important problem in the field of high energy physics. Because of the need, researchers are looking for new materials to be used for shielding (besides detectors, dosimeters, and many other applications). In this study, we have tested the interaction of proton, electron, and neutron for various energy levels with XSO_4 (X= Ba, Be, Ca, Mg, and Sr) targets. XSO_4 is in the powder structure, so it is mobile unlike concrete. We show the energy deposition of particles in the defined geometry. The simulation results show $MgSO_4$ is significantly less responsive to proton and neutron beams. Thus, it is less likely a shielding material between the considered molecules. The other important conclusion, the most of the proton beams energy is absorbed by the listed molecules show similar responsiveness as concrete which makes these molecules possible alternatives as a shielding material in the high energy experiments.

Key words: Monte Carlo, SO_4, Shielding, High Energy

XSO4 (X= Ba, Be, Ca, Mg, ve Sr)'ün Bilinen Parçacıklara (Proton, Elektron, ve Nötron)Tepkisinin Monte Carlo Çalışması: Bir Zırh Uygulaması

Öz: Detektörlerdeki ve teorilerdeki gelişmelere paralel olarak atom altı parçacıklar çalışılmaya devam edilmektedir. Deneylerde hızlandırılmış parçacıklar çalışılan hedeflerin ve hatta zırh içine bile nüfuz etmektedir. Bu sebepten dolayı araştırmacıları ve ekipmanları yüksek enerjili parçacıklardan korumak yüksek enerji araştırmalarında önemli bir problemdir. Bu ihtiyaçtan dolayı, araştırmacılar (detektör, dozimetre, ve diğer uygulamaların yanı sıra) zırh olarak kullanılabilecek yeni malzemeler arayışındadır. Bu çalışmada proton, elektron ve nötron ışınlarına maruz bıraktığımız XSO4 (X= Ba, Be, Ca, Mg, ve Sr) moleküllerini çeşitli enerji seviyeleri için inceledik. XSO4 pudra yapısındadır, böylece betonun aksine taşınabilirdir. Tanımlanan geometride enerji dağılımları incelendi. Yapılan analizler gösterdi ki MgSO4 diğer moleküllere kıyasla proton ve nötron ışınımlarına belirgin bir şekilde daha az tepki göstermiştir. Diğer bir önemli sonuç ise proton ışınımın enerjisinin büyük bir kısmı analiz konusu olan moleküller tarafından tutulmuştur, bu davranış referans noktası olarak kullandığımız beton ile benzerlikler göstermesi deney düzeneklerinde bu moleküllerin alternatif zırh malzemesi olarak kullanılabileceğini düşündürmektedir.

Anahtar kelimeler: Monte Carlo, SO_4, Zırh, Yüksek Enerji

1. Introduction

Irradiation of molecules by subatomic particles studied by many researchers [1-3]. Accelerators are propagating high energy particles to a specific target(s) for various reasons such as generate particles which have a short lifetime or to simulate an event at the establishment of the universe [4]. In experimental studies, sensitive and high technology devices are necessary. For example, the latest detectors collect more reliable data in shorter times since information about the short-lived particles could be captured with relatively fast detectors [5]. However, one approach to design better detectors requires that various combinations of elements to be tested. These tests could be done by the well-developed models with an appropriate package program such as FLUKA.

We have used FLUKA Monte Carlo package for our simulations [6, 7]. FLUKA is a comprehensive tool to study interactions of high energy particles with matter [8]. Thus, it has a wide variety of applications that are continuously studied such as shielding, detector design. It is continuously developed and used for simulations of experimental research in CERN (Conseil Européen pour la Recherche)[9]. The experimental researches investigate highly energized dense particles. The particles which have the potential to harm the around of the experimental setup are stoped by shielding [10]. We have investigated the molecules (XSO4 (X= Ba, Be, Ca, Mg, ve Sr)) respect to portland concrete which is a highly utilized shielding material.

* Corresponding author: sguldal@adiyaman.edu.tr. ORCID Number of authors: [1] 0000-0002-4247-0786

The main objective in this investigation is the show the relational energy deposition of XSO_4 (X= Ba, Be, Ca, Mg, and Sr) targets with Portland concrete for different energy levels of subatomic particles (proton, electron, and neutron) [11].

Simulation details have been explained in Materials and Method section. After Materials and Method section, we summarize results and give a brief discussion of the outputs. Lastly, we conclude this study with the conclusion of the findings with future studies.

2. Materials and Method

Geometry is a combination of a target, environment and black hole (Figure 1). The environment is defined as a vacuum. Outside of the vacuum is surrounded by a black hole. Target is a rectangular prism (40 cm x 40 cm x 100 cm) which is relatively thin for shielding applications, but it is enough big to extract response of the material under the common subatomic particles [12]. Geometry is used in the simplest form possible to obtain the pure response of the targets' molecules. Target materials are made of by the following molecules, XSO_4 (X= Ba, Be, Ca, Mg, and Sr) and concrete Portland (which is made of Hydrogen: 0.01, Carbon: 0.001, Oxygen: 0.529107, Sodium: 0.016, Magnesium: 0.002, Aluminum: 0.033872, Silicon: 0.337021, Potassium: 0.013, Calcium: 0.044, Iron: 0.014).

Beam gun is placed 10 cm away from the target. It is a Gaussian wave and pointed to positive z-direction for all particles. Also, full width at half maximum of beam wave is 1 cm by 1 cm for x and y directions respectively and energy levels change between 1 MeV- 120 MeV. This beam setup allows us to imitate experimental designs.

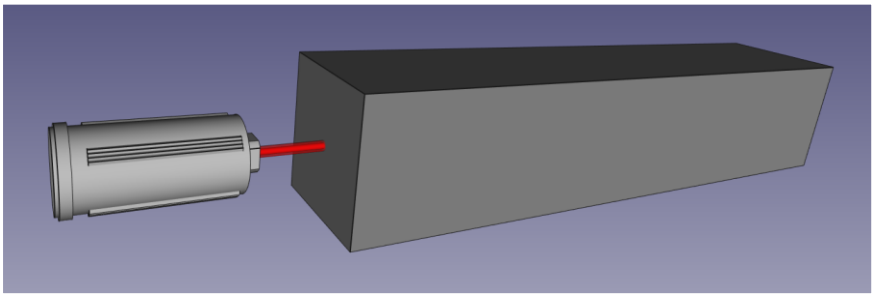

Figure 1. The simulation setup shows the beam and the target prism.

We use FLUKA Monte Carlo package for our simulations. Monte Carlo method randomly generates the beams with the limit of given properties (Shape: Gaussian, Energy: 1MeV- 120MeV, and Particle: e^-, p^+, n). Propagation and focus of Gaussian beam are well known, and a relatively lower diffraction rate are some of the advantages [13]. Utilized energy range corresponds to the energy of traditional laser-plasma accelerators [14, 15]. The well-studied charged and neutral particles are selected for the beam to broaden the effect of the study. Lastly, 5000 primaries produced a smoother distribution of the beam.

Targets are made out of alkaline earth metals with SO_4 for shielding applications. Specifically, $BaSO_4$ had been doped to cement with various concentrations to improve shielding [16, 17]. $BeSO_4$ is been investigated experimentally against high energy irritation to estimate the "age" of cosmic rays by means of ^{10}Be isotopes [18, 19]. $CaSO_4$, also, has been subject to radioactivity studies [20, 21]. $MgSO_4$ is effectively used as a drug component [22]. $SrSO_4$ is a promising candidate for a dosimeter [23, 24].

Structural properties are given in Table 1. The density of the target is an effective parameter for the simulations. The selected molecules highly durable with respect to the high temperatures.

Table 1. Densities of examined molecules are shown [25].

MOLECULE	DENSITY (GR/CM^3)	MELTING POINT (^{O}C)
$BaSO_4$	4.49	1580
$BeSO_4$	2.5	1127
$CaSO_4$	2.96	1460
$MgSO_4$	2.66	1137
$SrSO_4$	3.96	1606

3. Results and Discussion

Five sulfate-based molecules` and concrete Portland's responses have been measured against three well-studied particles (e^-, p^+, n). The beams send to target from 6 different angles which are between the target surface and incoming beam (Figure 2).

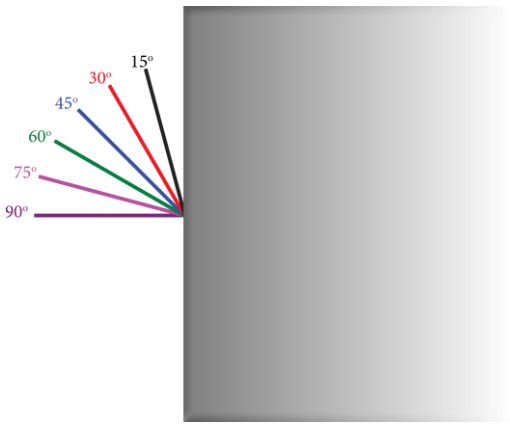

Figure 2. 6 different beam angles used to irritate targets

There is only one study for the considered molecules which is related to the dosimeters [26]. We have collected the information about deposited total energy in the material after every particle hits the specified target. Although, there is no significant difference between the responses of the molecules; $MgSO_4$ shows distinguishable behavior respect for the other molecules. It is described with detail in the following paragraphs.

In Figure 3, $BaSO_4$ target is irritated by e^-, p^+, and n. Electron beams' energies are stored in the target for relatively higher energies. Low energy beams are bounced from the surface of the target based on the energy decomposition. Proton beams have a linear relation with the beams' energies regardless of the beam angles. Proton energy residue is lower than electron energy because of the higher mass. Neutron beams behave differently than the other particles. Its energy is the least captured for 40 MeV. For the higher energy levels, angles' effect becomes more visible.

In Figure 4, $BeSO_4$ target's responses shown against various energy levels. Electron beams are deposited lesser for 10 MeV since relatively lower energy electrons scattered from the surface. Closer beams' angles to perpendicular have higher momentum respect to the surface normal caused higher interaction with the target. Neutron beams are less capable of escaping from the target for 10 MeV. However, the neutron beams with higher energies bounce from the surface by leaving 2% of their energies. Moreover, for 90 MeV and higher energies, the effect of the incoming beam angle becomes more visible.

In Figure 5, the response of $MgSO_4$ is presented. $MgSO_4$ stores lesser energy for the higher energized electron beams. For the lower energy, less than 10 MeV, the molecular mass of the molecules plays a crucial role in energy absorption. Diverse characteristics of $MgSO_4$ is also visible for the neutron beams. Beam angle becomes more visible for higher energy beams. The beams that are perpendicular to the surface and closer to the perpendicular angle are captured relatively higher level.

In Figure 6, the energy deposition for $CaSO_4$ target is shown for electron, proton, and neutron beams. The beams that come from lower angles are stored relatively lower. 10 MeV is a pivot point for storing energy in the target. Electron beams reach their highest ratio, but neutron beams reachest their lowest ratio. Additionally, electron beams reach their highest (~100%) for the perpendicular positions.

In Figure 7, $SrSO_4$ target's responses are presented for the beams which are released from the various angles. Electron is the most stored beam between all three beams. The stored proton beam energies are directly related to the energy of the beams regardless of the angle.

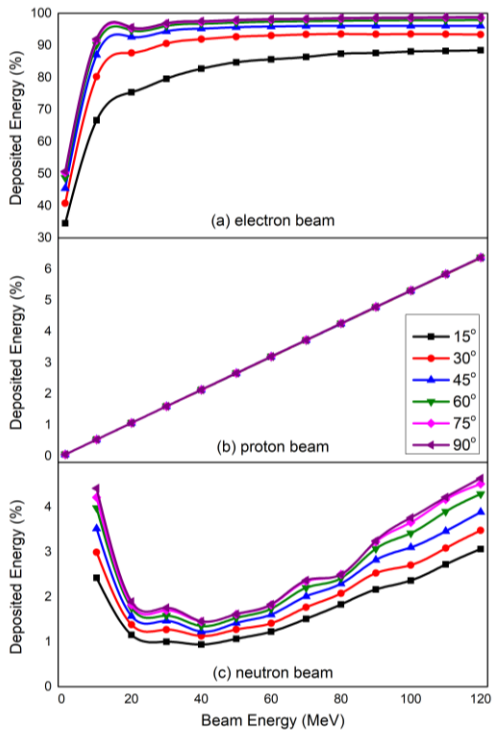

Figure 3. Stored energy ratio of $BaSO_4$ target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

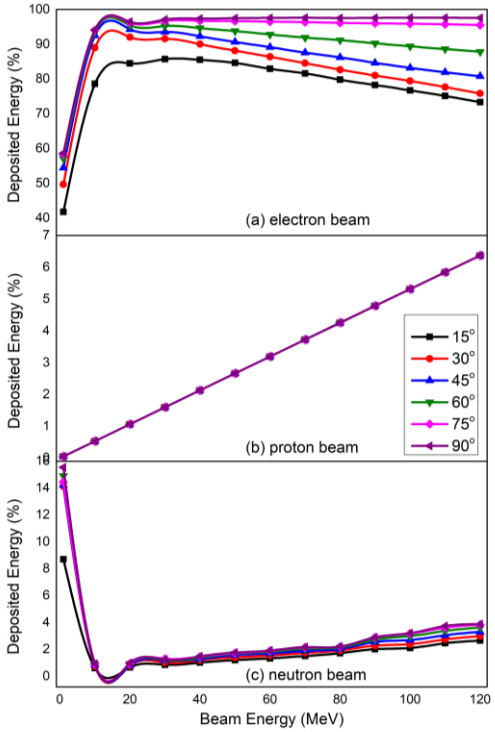

Figure 4. Stored energy ratio of $BeSO_4$ target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

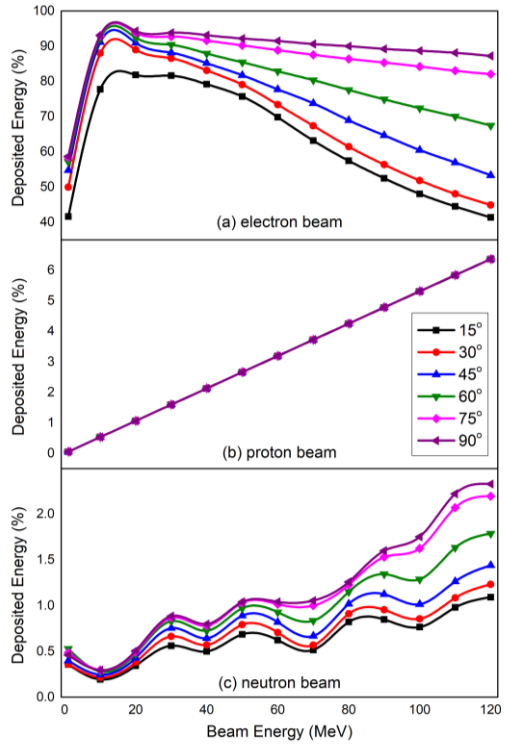

Figure 5. Stored energy ratio of $MgSO_4$ target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

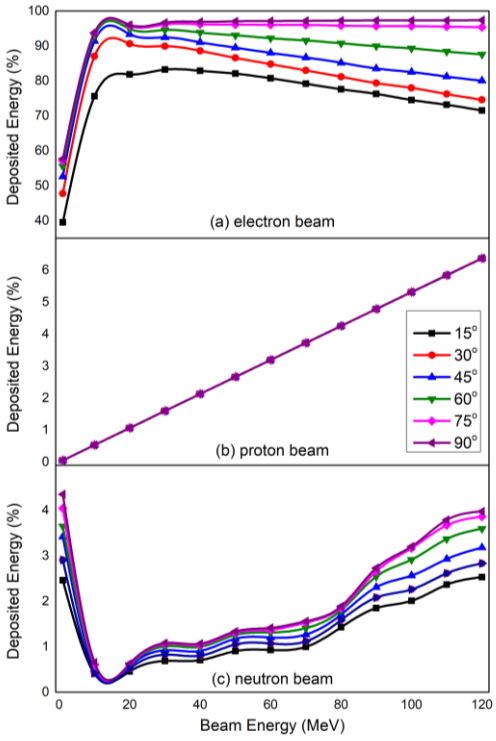

Figure 6. Stored energy ratio of $CaSO_4$ target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

In Figure 8, the Energy deposition behavior of Portland Concrete is shown as the reference point for the rest of the study. The concrete is commonly used as the shielding material. There is no result fully fits with the reference study. Thus, they are not equivalent to the concrete. However, the electron behavior of $BeSO_4$ and $CaSO_4$ resembles better for higher energy beams. The stored proton beams change from ~0% to ~6% for all the targets including the Portland Concrete. The closest neutron beams' depositions are detected in $BeSO_4$.

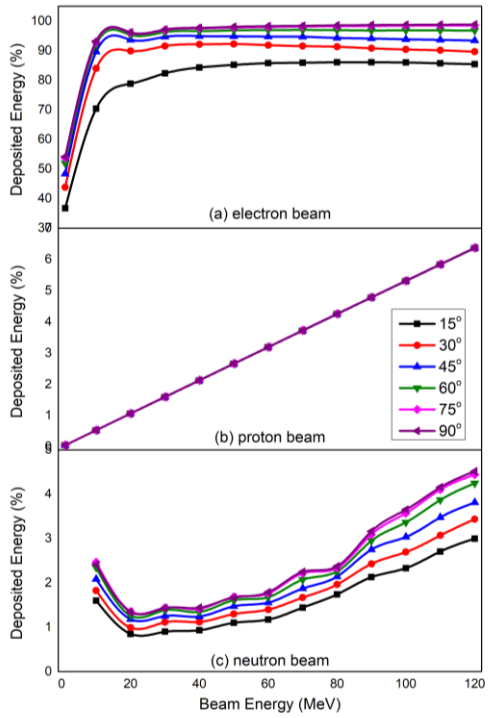

Figure 7. Stored energy ratio of $SrSO_4$ target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

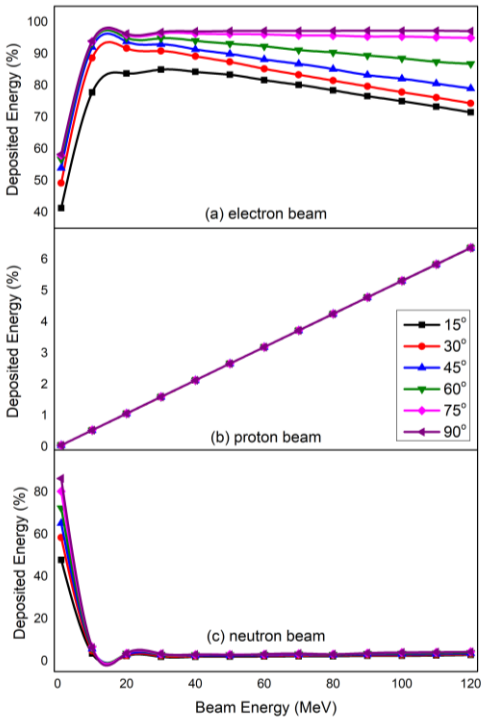

Figure 8. Stored energy ratio of Portland Concrete target (a) for an electron beam, (b) for a proton beam, and (c) for a neutron beam

4. Conclusion

In this work, we have examined the reaction of XSO_4 (X= Ba, Be, Ca, Mg, and Sr) against to e^-, p^+, n beams (1MeV- 120MeV and from 15° to 90°). The deposition of energy shows us how much energy stored in the target. Deposited energy in the targets shows the measurable quantity about the intensity of the interactions. $MgSO_4$ shows distinguishable behavior with respect to the other targets. For the higher energies, the beam angle plays a crucial role by the means of storing energy.

In future studies, we are planning to broaden this study to different particles and ions as the high energy beams. Additionally, we will test more promising molecules to be used in high energy experiments.

References

[1] Korkut H, Korkut T, Kara A, Yiğit M, and Tel E, "Monte Carlo Simulations of 17.9–22.3 MeV Energetic Proton Irradiation Effects on bcc-Zirconium Fusionic Materials," *Journal of Fusion Energy*, 2016; 35(3): 591-596.

[2] P.Limkitjaroenporn, Kaewkhao J, Chewpraditkul W, and Limsuwan P, "Mass Attenuation Coefficient and Effective Atomic Number of Ag/Cu/Zn Alloy at Different Photon Energy by Compton Scattering Technique," *Procedia Engineering*, 2012; 32: 847-854.

[3] Aygün B, Korkut T, Karabulut A, Gencel O, and Karabulut A, "Production and Neutron Irradiation Tests on a New Epoxy/Molybdenum Composite," *International Journal of Polymer Analysis and Characterization* 2015; 20(4): 323-329.

[4] Howell R H, Sterne P A, Hartley J, and Cowan T E, "High energy beam lifetime analysis," *Applied Surface Science*, 1999; 149(1-4): 103-105.
[5] Leach S A *et al.*, "Front-end electronics of the Compact High Energy Camera," presented at the Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2020/02/01/, 2020.
[6] Böhlen T T *et al.*, "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications," *Nuclear Data Sheets*, 2014; 120: 211-214.
[7] Ferrari A, Sala P R, Fasso` A, and Ranft J, "FLUKA: a multi-particle transport code," *CERN-2005-10*, 2005; INFN/TC_05/11, SLAC-R-773
[8] Hançerlioğullari A, Korkut T, and Madee Y G A, "The Neutron Macroscopic Cross Sections Calculation of Some Minerals by Using FLUKA Monte Carlo Method," *The Online Journal of Science and Technology* 2007; 7(3): 137-143.
[9] Fassò A, Ferrari A, Sala P R, and Ranft J, "FLUKA: Status and Prospects for Hadronic Applications," *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications*, (2001); 955-960.
[10] Sarıyer D and Küçer R, "Proton Hızlandırıcılarında Farklı Maddeler İçin Zırh Kalınlıklarının Analitik Yöntemle Belirlenmesi," *Determination of Shielding Thicknesses by Using Analytical Method for Different Materials in Proton Accelerators.*, Article 2013; 8, (1): 100-105.
[11] Callan E J, "Concrete for Radiation Shielding," *Journal Proceedings American Concrete Institute*, 1953; 50(9): 17-44.
[12] Nunomiya T *et al.*, "Measurements of Neutron Attenuation through Iron and Concrete at ISIS," *Journal of Nuclear Science and Technology*, 2014; 37 (1): 158-161.
[13] Shealy D and Hoffnagle J, *Laser beam shaping profiles and propagation*. 2006, pp. 5118-5131.
[14] Malka V, Faure J, Glinec Y, and Lifschitz A F, "Laser-plasma accelerator: status and perspectives," (in English), *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2006; 364(1840): 601-610.
[15] Glinec Y and Malka V, *High charge (0.5 nC) monoenergetic 170 +/-20 MeV electrons beam generated by 30 fs laser pulses*. 2004.
[16] Chanthima N, Prongsamrong P, Kaewkhao J, and Limsuwan P, "Simulated radiation attenuation properties of cement containing with BaSO4 and PbO," *Procedia Engineering*, 2012; 32: 976-981.
[17] Zezulová A, Staněk T, and Opravil T, "The Influence of Barium Sulphate and Barium Carbonate on the Portland Cement," *Procedia Engineering*, 2016; 151: 42-49.
[18] Moussavi-Zarandi A, "Determination of beryllium by use of photonuclear activation techniques," *Applied Radiation and Isotopes*, 2008; 66 (2): 158-161.
[19] Amin B S, Biswas S, Lal D, and Somayajulu B L K, "Radiochemical measurements of 10Be and 7Be formation cross sections in oxygen by 135 and 550 MeV protons," *Nuclear Physics A*, 1972; 195 (1): 311-320.
[20] Nambi K S V, "Self-dose and fading of CaSO4(Dy/Tm) and CaF2 (natural) TL phosphors in environmental radiation monitoring," *Nuclear Instruments and Methods in Physics Research*, 1982; 197 (2): 453-457.
[21] Bakshi A K, Suryanarayana P, Sharma R K, Maheshwari N K, Joshi V J, and Chougaonkar M P, "Measurement of gamma attenuation for ANR end shield model using CaSO4:Dy phosphor," *Annals of Nuclear Energy*, 2013; 55: 205-210.
[22] Thapa K and Jha R, "Magnesium sulphate: a life saving drug," (in eng), *JNMA; journal of the Nepal Medical Association*, 2008; 47(171): 104-108.
[23] Singh V P, Medhat M E, and Badiger N M, "Photon attenuation coefficients of thermoluminescent dosimetric materials by Geant4 toolkit, XCOM program and experimental data: A comparison study," *Annals of Nuclear Energy*, 2014; 68: 96-100.
[24] Rushdi M A H, Abdel-Fattah A A, and Soliman Y S, "Physico-chemical studies for strontium sulfate radiation dosimeter," *Journal of Radiation Research and Applied Sciences*, 2015; 8 (2): 221-225.
[25] Dongxue L and Liu Y, "Recent progress of dopant-free organic hole-transporting materials in perovskite solar cells," *Journal of Semiconductors*, 2017; 38 (1).
[26] McConn R J, Gesh C J, Pagh R T, Rucker R A, and Williams Iii R, "Compendium of Material Composition Data for Radiation Transport Modeling," United States2011-03-04 2011, Available: https://www.osti.gov/servlets/purl/1023125.

C O P Y R I G H T R E L E A S E F O R M

TURKISH JOURNAL OF SCIENCE AND TECHNOLOGY (TJST)
Published by Firat University

Firat University, Fen Bilimleri Enstitüsü Müdürlüğü
Turkish Journal of Science & Technology Editörlüğü
Elazığ-TURKEY,
Manuscript title: .....
.....
.....
Full names of all authors (in order to appear on manuscript): .....
.....
Name, address etc. of corresponding author: .....
.....
.....
ID Number: ..... Telephone: .....
E-mail: ..... Mobile phone: .....

The author(s) warrant(s) that:
a) the manuscript submitted is his/her/their own original work;
b) all authors participated in the work in a substantive way and are prepared to take public responsibility for the work;
c) all authors have seen and approved the manuscript as submitted;
d) the manuscript has not been published and is not being submitted or considered for publication elsewhere;
e) the text, illustrations, and any other materials included in the manuscript do not infringe upon any existing copyright or other rights of anyone. Notwithstanding the above, the Contributor(s) or, if applicable the Contributor's Employer, retain(s) all proprietary rights other than copyright, such as
a) patent rights;
b) to use, free of charge, all parts of this article for the author's future works in books, lectures, classroom teaching or oral presentations;
c) the right to reproduce the article for their own purposes provided the copies are not offered for sale.
However, reproduction, posting, transmission or other distribution or use of the article or any material contained therein, in any medium as permitted hereunder, requires a citation to the Journal and appropriate credit to Firat University as publisher, suitable in form and content as follows:
Title of article, author(s), journal title and volume/issue, Copyright© year.
All materials related to manuscripts, accepted or rejected, including photographs, original figures etc., will be kept by Turkish Journal of Science and Technology editority for one year following the editor's decision. These materials will then be destroyed.
I/We indemnify Firat University and the Editors of the Journals, and hold them harmless from any loss, expense or damage occasioned by a claim or suit by a third party for copyright infringement, or any suit arising out of any breach of the foregoing warranties as a result of publication of my/our article. I/We also warrant that the article contains no libelous or unlawful statements and does not contain material or instructions that might cause harm or injury.
This copyright form must be signed by all authors. Separate copies of the form (completed in full) may be submitted by authors located at different institutions; however, all signatures must be original.

ID number: ..... ID number: .....
Full name (block letters) ..... Full name (block letters) .....
Signature ..... Date ..... Signature ..... Date .....
ID number: ..... ID number: .....
Full name (block letters) ..... Full name (block letters) .....
Signature ..... Date ..... Signature ..... Date .....
ID number: ..... ID number: .....

Turkish authors must supply their ID card number; foreign authors must supply their passport number (if possible)