Int. J. Morphol.,
40(1):148-156, 2022.

F Approach Algorithm in Missing Landmark Problem

Aproximación al Algoritmo F en Punto de Referencia Perdido

Fatma Ezgi Can[1] & Ilker Ercan[2]

CAN, F. E. & ERCAN, I. F Approach algorithm in missing landmark problem. *Int. J. Morphol., 40(1)*:148-156, 2022.

SUMMARY: Missing data may occur in every scientific studies. Statistical shape analysis involves methods that use geometric information obtained from objects. The most important input to the use of geometric information in statistical shape analysis is landmarks. Missing data in shape analysis occurs when there is a loss of information about landmark cartesian coordinates. The aim of the study is to propose F approach algorithm for estimating missing landmark coordinates and compare the performance of F approach with generally accepted missing data estimation methods, EM algorithm, PCA based methods such as Bayesian PCA, Nonlinear Estimation by Iterative Partial Least Squares PCA, Inverse non-linear PCA, Probabilistic PCA and regression imputation methods. Landmark counts were taken as 3, 6, 9 and sample sizes were taken as 5, 10, 30, 50, 100 in the simulation study. The data are generated based on multivariate normal distribution with positively defined variance-covariance matrices from isotropic models. In simulation study three different simulation scenarios and simulation based real data are considered with 1000 repetations. The best and the most different result in the performance evaluation according to all sample sizes is the Min (F) criteria of the F approach algorithm proposed in the study. In case of three landmarks which is only the proposed F approach and regression assignment method can be applied, Min (F) criteria give best results.

KEY WORDS: Cartesian coordinates; Geometric Morphometry; Landmark; Missing data; Shape analysis.

INTRODUCTION

Missing data arises frequently in scientific studies. In statistical shape analysis, missing data occurs at landmark coordinates. Shape, one of the most fundamental properties of biological structures, makes the overall appearance of the structures unique (Cho *et al*., 2019). In other words, shape is the physical property of objects whose appearance plays a major role in statistical analysis (Xu & Hong, 2017). The term of shape is generally used to describe appearance of an object (Anwary, 2012). Shape is all the geometrical information that remains when location, scale, and rotational effects are filtered out from an object (Kendall, 1977; Dryden & Mardia, 1998).

Statistical shape analysis includes methods using geometric information obtained from objects. Landmarks are the most important input for using geometric information. Each landmark has cartesian coordinates as ordered pairs in two dimensional plane or ordered triplet in three dimensional space (Ercan *et al*., 2012). Cartesian coordinates of objects in two-dimensional and three-dimensional space can be obtained by determining the landmarks, which are the most important inputs used in shape analysis.

Missing data in data sets can cause significant problems in statistical studies. Missing data in studies may cause biased estimates of parameters, loss of information, decreased statistical power, increased standard errors. Failing to eliminate missing data properly can cause the unsuitable data for a statistical procedure and violations of statistical analyses assumptions (Dong & Peng, 2013). It is seen that the problem of missing data is much more important, especially when multivariate analysis will be applied.

Missing data in shape analysis occurs when there is a loss of information in landmark coordinates. Data loss occurring in landmark coordinates in health and anthropology studies may be caused by fractures in the examined bone structure or deterioration in image quality. Therefore, missing data in cartesian coordinates of landmarks cause that landmark unusable and the unit of interest to exclude the survey. For these reasons missing data in shape analysis is important due to leading loss of data and shape integrity.

The missing landmark problem, especially in forensic medicine, paleontology and archeology, becomes

[1] Department of Biostatistics, Faculty of Medicine, Izmir Kâtip Celebi University, Izmir, Turkey.
[2] Department of Biostatistics, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey.

Received: 2021-03-02 Accepted: 2021-04-12

even more important when the number of samples included in the study is low. Methods of Expectation Maximization (EM) algorithm, multiple regression imputation and principal component analysis (PCA) are commonly used to estimate missing data (Couette & White, 2010). Although EM algorithm is one of the most used methods of missing data estimation, it has been observed that methods based on distributional models such as likelihood and multiple assignment are mostly used in recent years (Pigott, 2001).

The general and basic idea in developing missing data estimation methods is to estimate the missing data in data sets. Nowadays, data analysis is also done through shape and image. Along with the technological development, missing data has begun to be occured in figures and images with the development of statistical shape analysis approaches. The missing data problem in shape analysis appears as a missing data in cartesian coordinates due to landmark losses. A special method developed for shape-based missing landmarks could not be seen. The aim of study is to propose F approach algorithm for estimating missing landmark coordinates and compare the performance of F approach with generally accepted methods, EM algorithm, PCA based methods such as Bayesian PCA (BPCA), Nonlinear Estimation by Iterative Partial Least Squares PCA (INIPALS), Inverse non-linear PCA (NLPCA), Probabilistic PCA (PPCA) and regression imputation methods.

MATERIAL AND METHOD

Landmark. A landmark is a point of correspondence on each object that matches between and within populations (Lele & Richtsmeier, 2001; Dryden & Mardia). Each landmark has Cartesian coordinates in the form of a triple ordered in a two-dimensional plane or a triple in three-dimensional space.

There are different classifications of landmarks in the literature. Dryden & Mardia classified landmarks into three groups as anatomical landmarks, mathematical landmarks, and pseudo landmarks. Lele & Richestmeier divided the landmarks into three different groups as traditional, fuzzy and structured landmarks.

Landmark-Based Approaches in Geometric Morphometry. Landmark-based geometric morphometry is a powerful approach that explains the biological nature of the shape, shape variability, and the relationship of shape with other factors. Graphical representations that emerge from shape differences as a result of analysis are visually attractive and intuitive. Traditional morphometry involves summarizing morphology in terms of length measurements, proportions or angles, whereas in landmark-based geometric morphometry, the shape is summarized using two or three dimensional Cartesian coordinates in terms of landmark configuration. Geometric morphometry is powerful and popular because it determines the spatial relationship between landmark data obtained from organisms (Webster & Sheets, 2010).

Some of the landmark-based methods used in geometric morphometry are thin plate spline analysis, finite element morphometry, Procrustes analysis and Euclidean Distance Matrix analysis (EDMA). Procrustes analysis and EDMA are frequently used among these methods (Ercan *et al*., 2012).

Missing Data Estimation Methods. Missing data can causes major problems in many studies. Ignoring the missing data disrupts the randomness of the sample and eliminates the possibility of generalizing the results (Rubin, 1976; Little & Rubin, 1987; Dong & Peng). The density of missing data can cause a decrease in power in statistical inferences and deviations in parameter estimates.

In morphological and especially paleontological studies, missing data in landmark coordinates frequently occur due to fossilization and time-dependent erosion (Fig. 1). Geometric morphometric methods require all measured specimens to have the same number of landmarks at homologous positions (Mitteroecker & Gunz, 2009). Therefore the approaches of paleontologists are to work with existing landmarks, select samples that do not contain missing landmarks or exclude examples that contain missing landmarks. As a result of these approaches, paleontologists are faced with either the loss of morphological information or to reduce the sample size. While these situations are not a problem when working with large samples, these approaches may cause problems in applying statistical analysis in small samples (Couette & White; Ozdemir *et al*., 2010).

Many missing data assignment methods are available in the literature for missing data estimation. Mean substition, EM algorithm and multiple regression assignment methods are among the most used missing data assignment methods (Adams *et al*., 2004; Schmitt *et al*., 2015). Apart from these methods, approaches to estimate missing data by modifying PCA have also been proposed (Nounou *et al*., 2002; Scholz *et al*., 2005; Stacklies *et al*., 2007).

In statistical shape analysis, missing data problem arises when cartesian coordinates of the landmarks of interest cannot be determined. Failure to determine the landmark coordinates causes the relevant landmark excludes from the study.

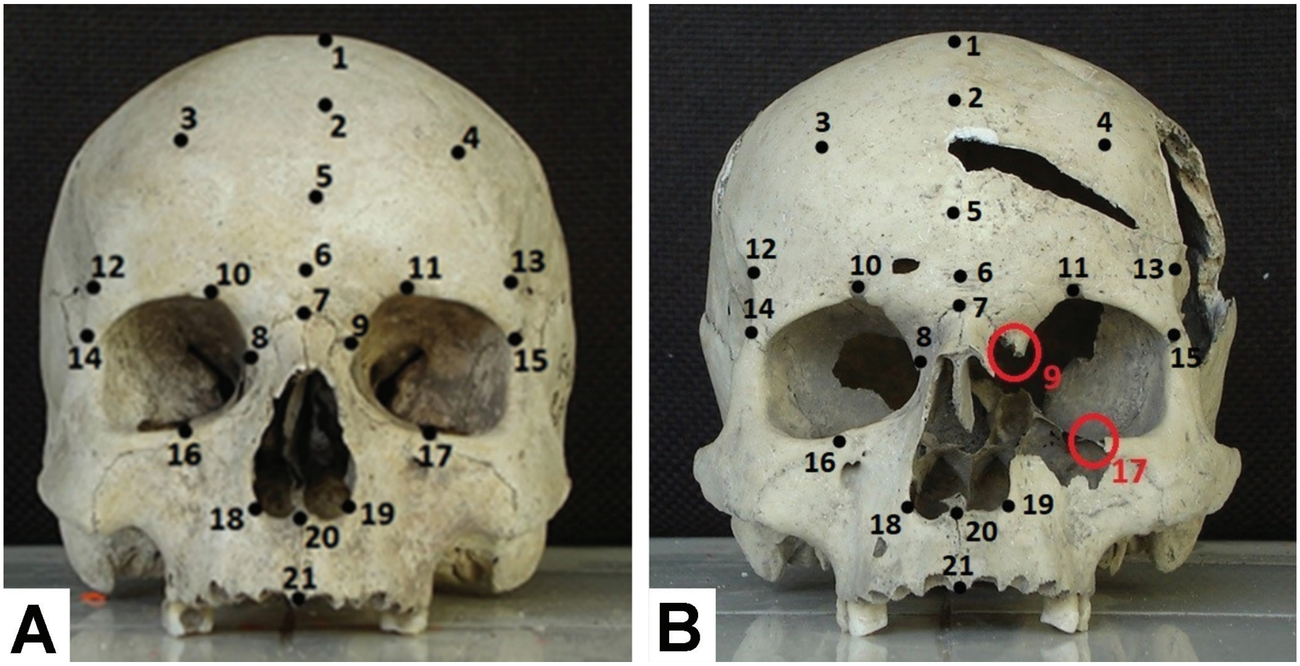

Fig. 1. Cranium shape. A) Cranium without missing. B) Cranium undergoing deformation landmarks.

The Proposed F Approach. We propose the F approach for estimating missing landmarks using Bookstein coordinates, circle equation and F statistics as an alternative to commonly used missing data estimation methods. Min (F) and Max (F) criteria were evaluated according to F statistics.

Let (x_j, y_j), $j = 1,2, ..., k$, be $k \geq 3$, landmark in a plane, Bookstein coordinates es are calculated as follows:

The Bookstein coordinates (u_j^B, v_j^B) obtained by moving the coordinates of an object, j = 3, ..., k, to the coordinates $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$ of the two landmarks that are referenced after translation, scaling and rotation (Equation-1).

$$u_j^B = \frac{\{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\}}{D_{12}^2} - \frac{1}{2}$$

$$v_j^B = \frac{\{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\}}{D_{12}^2}$$

Equation-1

Where j = 3, ..., $k, D_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$ and $-\infty < u_j^B, v_j^B < \infty$ (Ercan *et al*., 2015; Dryden & Mardia).

The proposed F approach algorithm for missing landmark estimation in our study is given below.

Step I: In the data set with k landmarks of n units, the m.th landmark that is missing in the related unit is determined.

Step II: The i.th and j.th landmarks are determined as two reference landmarks to be used in estimating the missing m.th landmark in the relevant unit.

Step III: The data set is transformed into Bookstein coordinates by taking the i.th and j.th landmarks as reference.

Step IV: In units where the m.th landmark is not missing, the distances between the i.th and j.th landmarks and the m.th landmark are calculated by using Euclidean distances.

Step V: The mean and standard error of the i-m (d_2) and j-m (d_3) distances are calculated according to the distances between the landmark calculated for each unit.

Step VI: 95 % confidence intervals are calculated for d_2 and d_3 distances.

Step VII: The lower limit values of both confidence intervals are accepted as the beginning of the iteration and the upper limit value as the end of the iteration.

Step VIII: Iteration coefficient are determined.

Step IX: Coordinates are estimated by using the circle equation whose two points are known in equation-2 for the m.th landmark missing in the relevant unit in each iteration (Fig. 2).

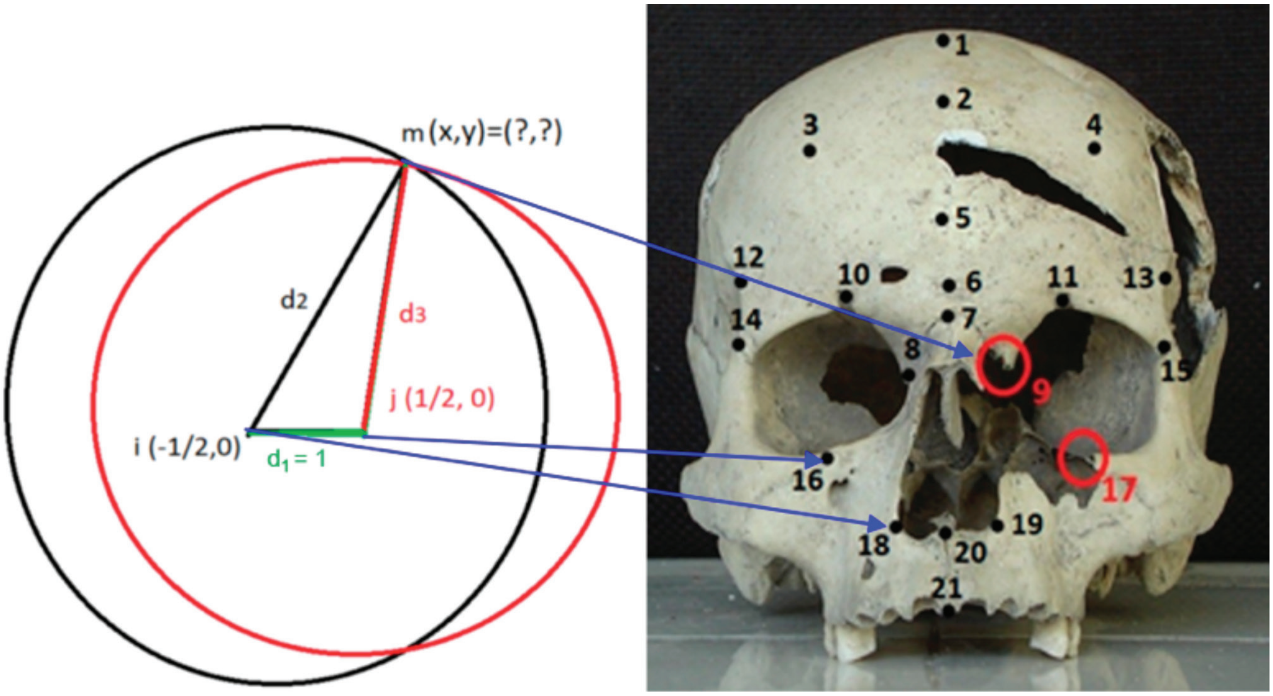

Fig. 2. Estimating coordinates using the equation of a circle with two points known

$$(x-\left(-\frac{1}{2}\right))^2+(y-0)^2=d_2^2$$

$$(x-\frac{1}{2})^2+(y-0)^2=d_3^2$$

Equation-2

Step X: F statistics are calculated for predicted m.th landmark coordinates x_m and y_m.

Step XI: The iteration is repeated until the d_2 and d_3 distances reach the upper limit value.

Step XII: Min(F) and Max(F) statistics are calculated considering all iterations.

Step XIII: According to the Min(F) and Max(F) statistics, the corresponding x and y coordinates are considered as missing landmark coordinates.

Step XIV: The coordinates of all landmarks are transformed from Bookstein coordinates to their original coordinates.

Comparison of other missing data estimation methods (Simulation methodology and scenarios). We generated landmarks coordinates based on Ozdemir *et al*. (Equation-2). They examined the differences in cranial shape variation from skeletal collections belongs to the Late Byzantine and modern human periods.

In our study we used three different landmark situation which is 3, 6 and 9 landmark. Considered landmarks in our study are explained below (Fig. 1a).

· 3 landmark situation : 1st, 3rd, 4th landmarks.

· 6 landmark situation : 1st, 3rd, 4th, 6th, 12th, 13th landmarks.

· 9 landmark situation : 1st, 3rd, 4th, 6th, 12th, 13th, 16th, 17th, 20th landmarks

A variety of sample sizes (5, 10, 30, 50, 100) were considered for three scenarios and data generated from multivariate normal distribution with positively defined variance-covariance matrix. 1000 repetitions were made in the simulation study. In order to compare performances of missing data estimation methods, missing landmark is created by extracting cartesian coordinates of the i.th landmark from the landmark data sets derived in two dimensions.

To compare the performance of the methods in estimating missing landmark, the root mean square error (RMSE) criterion was used by considering the distance between the x and y points of the original data and the predicted data (Fig. 3).

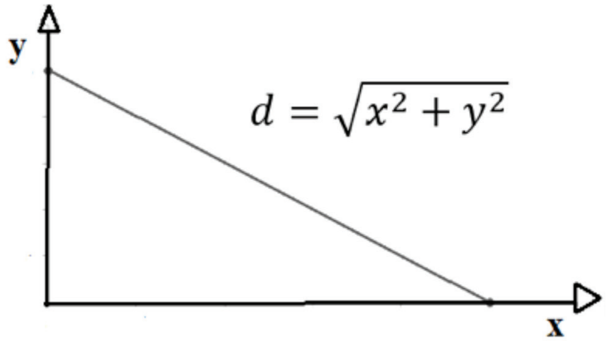

Fig. 3. The distance, d, between the x and y points used in the calculation of the RMSE criterion

$$RMSE = \sqrt{\frac{\sum_{i=1}^{r}(d_i^{estimated} - d_i^{observed})^2}{r}}, \text{ i=1,2,3, …, r.}$$

Equation-3

Where r is the number of repetitions and d is hypotenuse.

The simulations were performed by using the “mvtnorm”, “shapes”, “readxl”, “dplyr”, “Amelia”, “mice” and “pcaMethods” packages in R-3.4.0 (Stacklies *et al*.; R Development Core Team, 2010; Honaker *et al*., 2011; van Buuren & Groothuis-Oudshoorn, 2011; Dryden 2017.; Wickham & Bryan, 2017; Wickham *et al*., 2019; Genz *et al*., 2020).

Simulation Based Real Data. In real application, mean vectors of landmarks and modified variance-covariance matrix from Ozdemir *et al*. are used. Results of real application and comparisons of our methods are given Table I. Scatter graph of F statistic values calculated for proposed F approach in the study with scenario 1, 3 landmarks and n = 50 is given in Figure 4.

Simulation Results. In the simulation study, 3 different simulation scenarios were created by using variance-covariance matrices based on isotropic models.

* Scenario 1: Variance was 0.5, the non-diagonal values were 0.

* Scenario 2: Variance was 1, the non-diagonal values were 0.

* Scenario 3: Variance was 5, the non-diagonal values were 0.

The results of scenarios 2-4 were given in Tables II-IV.

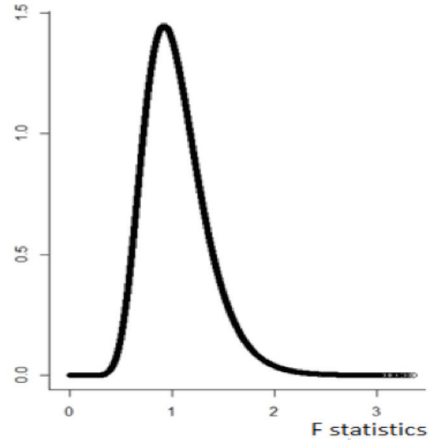

Fig. 4. F statistic values calculated for proposed F approach for scenario 1, 3 landmarks and n = 50

DISCUSSION

Missing data problem frequently arises in studies regardless of how meticulous and careful the research process is carried out. The problem of missing data sometimes occurs from inability to reach the units and also due to technical reasons and the nature of the units. Violations of statistical analysis assumptions and bias problems in estimates arises in studies due to the missing data problem. The frequent occurrence of the missing data problem in studies has lead researchers to study about missing data estimation methods.

When the studies for the missing landmark estimation for shape analysis were investigated in the literature, few studies were found. Bookstein *et al*. (1999) used thin-plate spline relaxation method to estimate the missing landmark in their studies on cranium profiles. On the other hand, some researchers have applied the analysis by removing missing units without considering the missing landmarks (Beumer *et al*., 2006).

In morphometry studies, data sets generally consist of anatomical structures such as skeletons and bones. Missing landmark coordinates occur due to time-dependent breaks in such anatomical structures. Unknown landmark coordinates causes that the landmark should be out of work. In paleontology and archeology studies, which usually have small sample size, each landmark is much more valuable. For these reasons, it is important to locate the landmarks and estimate the missing landmark coordinates in research. In small sample studies, in cases where it is not possible to replace the unit with the missing landmark, the missing landmark estimation becomes a necessity.

Table I. Comparisons of real data results.

L	n	Min(F)	Max(F)	Regression	BPCA	INIPALS	NLPCA	PPCA	EM
	5	2929.08	3012.848	27203.44	-	-	-	-	-
	10	12578.94	5420	5723.36	-	-	-	-	-
3*	30	1089.71	3153.89	515.73	-	-	-	-	-
	50	1507.41	2262.17	1995.17	-	-	-	-	-
	100	2279.73	3081.45	3323.36	-	-	-	-	-
	5	5394.45	1723.55	43269.19	3580.83	2955.15	15020.22	8689.11	-**
	10	14405.81	8589.75	3615.76	1427.57	1819.651	4771.66	2881.40	-**
6	30	1961.09	3438.84	3938.09	1527.54	1784.90	5072.69	2328.07	1732.62
	50	8131.83	11426.52	7767.12	5018.08	3982.79	26661.56	7409.17	3460.33
	100	9772.20	11514.82	13513.07	10073.97	11361.39	5621.94	8836.57	12795.69
	5	2154.62	3877.67	2349.131	9377.49	9055.60	15219.91	9695.944	-**
	10	789.12	1703.46	7028.255	17229.70	12757.08	53319.7	16237.29	-**
9	30	3966.03	618.97	5043.55	2470.69	935.41	8254.19	3728.74	3813.62
	50	2074.78	1368.92	946.61	515.44	508.68	715.88	515.48	608.47
	100	2054.31	1606.86	1343.91	9032.63	895.75	4480.71	8277.03	761.70

*In case of 3 landmarks, EM algorithm and PCA based methods can not be used due to the low number of landmarks. ** Sample size is too low to estimate. Underlined results shows best methods

Table II. Comparisons of results from scenario 1.

L	n	Min(F)	Max(F)	Regression	BPCA	INIPALS	NLPCA	PPCA	EM
	5	12487.78	4991.18	1157.142	-	-	-	-	-
	10	1680.09	6790.40	49.46	-	-	-	-	-
3*	30	127.17	39.88	140.97	-	-	-	-	-
	50	26.90	488.23	168.67	-	-	-	-	-
	100	26.09	28.36	18.06	-	-	-	-	-
	5	6365.38	3693.41	3731.33	1026.94	1188.06	9107.18	6596.22	-**
	10	13644.45	2286.48	9812.69	6202.70	1386.47	10194.98	6109.82	-**
6	30	159.09	101.13	52.46	13.05	14.37	26.21	16.15	10.03
	50	864.70	46.17	2544.64	71.55	59.29	340.70	75.56	17.19
	100	552.45	2287.70	1943.67	271.05	175.98	494.40	269.91	167.23
	5	35755.73	15407.93	7125.53	504654.40	629058.70	8212700	686608.50	-**
	10	5279.23	30913.93	20277.28	37131582	35387356	72368175	46541308	-**
9	30	55.43	119.87	105.14	1274.70	211.90	2452.31	1478.32	28506.56
	50	307.45	1774.17	450.88	236.71	237.08	239.08	237.11	328.73
	100	85.12	311.19	753.64	369.48	152.54	2333.57	439.96	1024.55

*In case of 3 landmarks, EM algorithm and PCA based methods can not be used due to the low number of landmarks. ** Sample size is too low to estimate. Underlined results shows best methods

Table III. Comparisons of results from scenario 2.

L	n	Min(F)	Max(F)	Regression	BPCA	INIPALS	NLPCA	PPCA	EM
	5	2021.55	17511.33	3174.53	-	-	-	-	-
	10	635.80	2009.67	4113.83	-	-	-	-	-
3*	30	177.96	56.13	220.31	-	-	-	-	-
	50	74.94	573.34	261.16	-	-	-	-	-
	100	301.12	17.52	217.14	-	-	-	-	-
	5	6350.87	3777.63	1642.72	1033.11	1194.66	8608.343	7087.908	-**
	10	13574.19	2292.01	9541.46	6285.77	1397.07	10387.44	6197.602	-**
6	30	235.47	150.34	78.46	18.48	16.91	31.519	19.03	14.586
	50	641.50	32.10	1907.18	72.22	60.33	204.129	75.938	22.38
	100	457.46	2362.92	1991.35	216.41	222.22	445.717	214.44	239.021
	5	25201.79	11010.50	6727.152	357102.10	443267.20	3247662	484698.20	-**
	10	7201.97	40599.99	19769.99	18483807	17554313	58581131	23248569	-**
9	30	81.10	174.16	152.82	1270.03	215.65	2505	1502.86	35320.18
	50	576.31	4833.97	1261.71	573.26	573.24	573.45	573.26	1087.31
	100	18.70	495.50	601.87	415.39	24.29	2190.25	136.02	1130.67

*In case of 3 landmarks, EM algorithm and PCA based methods can not be used due to the low number of landmarks.** Sample size is too low to estimate. Underlined results shows best methods

Table IV. Comparisons of results from scenario 3.

L	n	Min(F)	Max(F)	Regression	BPCA	INIPALS	NLPCA	PPCA	EM
3*	5	12847.62	5269.93	1640.477	-	-	-	-	-
	10	672.29	8674.75	1331.86	-	-	-	-	-
	30	418.11	133.28	588.47	-	-	-	-	-
	50	834.99	276.62	231.01	-	-	-	-	-
	100	762.74	295.57	2162.87	-	-	-	-	-
6	5	5404.07	3667.86	1117.92	1022.46	1163.69	6184.50	6736.11	-**
	10	12910.18	2263.59	7540.58	6375.09	1446.63	11568.38	6505.51	-**
	30	528.41	339.01	178.80	37.11	45.93	77.27	54.37	32.56
	50	840.50	112.55	2734.09	174.79	237.57	253.65	165.67	156.93
	100	1174.18	2709.02	2367.70	160.81	394.09	1077.01	159.55	492.62
9	5	22104.13	9707.60	8116.199	157284	198363.10	796502.80	218668.20	-**
	10	7111.18	37641.03	4165.73	3620524	3379616	7347837	4583450	-**
	30	187.98	402.66	353.332	1261.98	235.29	2548.73	1551.78	48.63
	50	38.18	450.88	156.481	248.14	248.14	248.20	248.15	114.13
	100	464.63	753.64	770.22	770.22	435.87	1901.59	578.25	522.58

*In case of 3 landmarks, EM algorithm and PCA based methods can not be used due to the low number of landmarks. ** Sample size is too low to estimate
Underlined results shows best methods

Missing data estimation methods used in this study were compared with different methods in morphometric studies (Strauss *et al*., 2003; Couette & White). Couette & White compared the EM algorithm and multiple regression assignment methods to investigate the importance of missing data in three-dimensional morphometric data and showed that the EM algorithm and multiple regression assignment methods give similar results. Strauss *et al*. compared the EM algorithm and principal component estimation, and found that the methods give similar results. In our study while EM algorithm give similar results with Max(F) criteria approach, Min(F) criteria give much better results than EM algorithm.

Arbour & Brown (2014) compared the performances of BPCA, least squares regression, thin plate spline analysis and mean substition methods used in estimation of missing data in geometric morphometry. As a result of the comparison, they stated that BPCA and least squares regression are reliable methods. According to the simulation results in our study, the BPCA and regression assignment methods give low performance compared to the Min(F) and Max(F) criteria.

Brown *et al*. (2012) investigated the performance of missing data estimation methods in morphometric data and compared the mean substition, regression assignment and BPCA methods, and stated that when the missing data rate is low, the BPCA method gave the best performance and the mean substition method gave the worst performance. Neeser *et al*. (2009) stated that the regression assignment method gave the best results by comparing mean substition, thin plate spline analysis and multiple regression assignment methods for the reconstruction of fossil craniums.

In the simulation study, considering all sample sizes, the Min (F) criterion of the proposed F-approach algorithm gave the best and the most different result in performance evaluation. When other methods are evaluated, the Max(F) criterion of the proposed F-approach algorithm and EM algorithms yielded similar results. Regression imputation method can be considered next succesful method for missing landmark estimation. PCA based methods, BPCA, INIPALS, NLPCA, PPCA, are not successful like the other methods considered in the study. However, in other studies on landmark-based missing data estimation, BPCA performed well when missing data rate was low (Brown *et al*.; Arbour & Brown).

Many of the methods frequently used in missing landmark estimation and the methods considered in this study are intense in statistical theory and therefore, performances of this methods may be reduced if some statistical assumptions are not achieved. The number of landmarks is also very important in these methods. In the methods examined in the study, except for the regression assignment method, there is a problem if the number of landmarks is 3. The F-approach algorithm that we proposed in the study can be used for missing landmark estimation in case there are 3 landmarks forming the shape. When we evaluated the simulation results for the proposed F-approach algorithm with the regression assignment method according to the case of three landmarks, Min(F) criterion gave the best results. However, Max(F) and regression assignment methods give similar results. The fact that the number of landmarks is more than 3 does not turn into an advantage in terms of other methods.

CAN, F. E. & ERCAN, I. F Enfoque de Algoritmo F en el punto de referencia perdido. *Int. J. Morphol., 40(1)*:148-156, 2022.

RESUMEN: Los datos faltantes pueden ocurrir en todos los estudios científicos. El análisis estadístico de formas involucra métodos que utilizan información geométrica obtenida de objetos. La entrada más importante para el uso de información geométrica en el análisis estadístico de formas son los puntos de referencia. Los datos que faltan en el análisis de formas se producen cuando hay una pérdida de información sobre las coordenadas cartesianas históricas. El objetivo del estudio es proponer el algoritmo de enfoque F para estimar las coordenadas de puntos de referencia faltantes y comparar el rendimiento del enfoque F con métodos de estimación de datos faltantes generalmente aceptados, algoritmo EM, métodos basados en PCA como Bayesian PCA, Estimación no lineal por Iterative Partial Least Squares PCA , PCA no lineal inverso, PCA probabilístico y métodos de imputación de regresión. Los recuentos de puntos de referencia se tomaron como 3, 6, 9 y los tamaños de muestra se tomaron como 5, 10, 30, 50, 100 en el estudio de simulación. Los datos se generan en base a una distribución normal multivariada con matrices de varianza-covarianza definidas positivamente a partir de modelos isotrópicos. En el estudio de simulación se consideran tres escenarios de simulación diferentes y se consideran datos reales basados en simulación con 1000 repeticiones. El mejor y más diferente resultado en la evaluación del desempeño según todos los tamaños de muestra es el criterio Min (F) del algoritmo de enfoque F propuesto en el estudio. En el caso de tres puntos de referencia, que es solo el enfoque F propuesto y se puede aplicar el método de asignación de regresión, los criterios Min (F) dan mejores resultados.

PALABRAS CLAVE: Coordenadas cartesianas; Morfometría geométrica; Punto de referencia; Datos faltantes; Análisis de forma.

REFERENCES

Adams, D. C.; Rohlf, F. J. & Slice, D. E. Geometric morphometrics: Ten years of progress following the ‘revolution’. *Ital. J. Zool., 71(1)*:5-16, 2004.

Anwary, A. R. *Statistical Shape Analysis for the Human Back. Master of Philosophy in Production and Manufacturing Engineering*. Wolverhampton, University of Wolverhampton, 2012.

Arbour, J. H. & Brown, C. M. Incomplete specimens in geometric morphometric analyses. *Methods Ecol. Evol., 5(1)*:16-26, 2014.

Beumer, G. M.; Tao, Q.; Bazen, A. M. & Veldhuis, R. N. J. *A Landmark Paper in Face Recognition*. Southampton, 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006.

Bookstein, F.; Schäfer, K.; Prossinger, H.; Seidler, H.; Fieder, M.; Stringer, C.; Weber, G. W.; Arsuaga, J. L.; Slice, D. E.; Rohlf, F. J.; *et al*. Comparing frontal cranial profiles in archaic and modern homo by morphometric analysis. *Anat. Rec., 257(6)*:217-24, 1999.

Brown, C. M.; Arbour, J. H. & Jackson, D. A. Testing of the effect of missing data estimation and distribution in morphometric multivariate data analyses. *Syst. Biol., 61(6)*:941-54, 2012.

Cho, M. H.; Asiaee, A. & Kurtek, S. Elastic statistical shape analysis of biological structures with case studies: a tutorial. *Bull. Math. Biol., 81(7)*:2052-73, 2019.

Couette, S. & White, J. 3D geometric morphometrics and missing-data. Can extant taxa give clues for the analysis of fossil primates? *Comptes Rendus Palevol, 9(6-7)*:423-33, 2010.

Dong, Y. & Peng, C. Y. J. Principled missing data methods for researchers. *SpringerPlus, 2*:222, 2013.

Dryden, & Mardia, K. V. Statistical Shape Analysis. New York, John Wiley and Sons, 1998.

Dryden, I. L. & Mardia, K. V. *Statistical Shape Analysis with Applications in R*. 2nd ed. Hoboken, John Wiley & Sons, 2016.

Dryden, I. shapes: Statistical shape analysis. (Version R package version 1.2.0). Retrieved from https://cran.r-project.org/package=shapes, 2017.

Ercan, I.; Ocakoglu, G.; Sigirli, D. & Ozkaya, G. Statistical shape analysis and usage in medical sciences: review. *Turk. Klinik. J. Biostat., 4(1)*:27-35, 2012.

Ercan, I.; Sigirli, D. & Ozkaya, G. Examining the variations in the results of the Hotelling T (2) test in case of changing baseline landmarks in the Bookstein coordinates. *Interdiscip. Sci., 7(2)*:186-93, 2015.

Genz, A.; Bretz, F.; Miwa, T.; Mi, X.; Leisch, F.; Scheipl, F.; Bornkamp, B.; Maechler, M. & Hothorn, T. Computes multivariate normal and t probabilities, quantiles, random deviates and densities. mvtnorm: Multivariate Normal and t Distributions, 2020. Available from: http://cran.r-project.org/package=mvtnorm

Honaker, J.; King, G. & Blackwell, M. Amelia II: A program for missing data. *J. Stat. Softw., 45(7)*:1-47, 2011.

Kendall, D. The diffusion of shape. *Adv. Appl. Probab., 9(3)*:428-30, 1977.

Lele, S. R. & Richtsmeier, J. T. *An Invariant Approach to Statistical Analysis*. London, Chapman and Hall/CRC, 2001.

Little, R. J. A. & Rubin, D. B. *Statistical Analysis with Missing Data*. New York, Wiley, 1987.

Mitteroecker, P. & Gunz, P. Advances in Geometric morphometrics. *Evol. Biol., 36(2)*:235-47, 2009.

Neeser, R.; Ackermann, R. R. & Gain, J. Comparing the accuracy and precision of three techniques used for estimating missing landmarks when reconstructing fossil hominin crania. *Am. J. Phys. Anthropol., 140(1)*:1-18, 2009.

Nounou, M. N.; Bakshi, B. R.; Goel, P. K. & Shen, X. Bayesian principal component analysis. *J. Chemom., 16(11)*:576-95, 2002.

Ozdemir, S. T.; Ercan, I.; Ozkaya, G.; Cankur, N. S. & Erdal, Y. S. Geometric morphometric study and cluster analysis of late Byzantine and modern human crania. *Coll. Antropol., 34(2)*:493-9, 2010.

Pigott, T. D. A review of methods for missing data. *Educ. Res. Eval., 7(4)*:353-83, 2001.

R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010. Available from: https://www.R-project.org

Rubin, D. B. Inference and missing data. *Biometrika, 63*:581-92, 1976.

Schmitt, P.; Mandel, J. & Guedj, M. A comparison of six methods for missing data ımputation. *J. Biom. Biostat., 6(1)*:1-6, 2015.

Scholz, M.; Kaplan, F.; Guy, C. L.; Kopka, J. & Selbig, J. Non-linear PCA: A missing data approach. *Bioinformatics, 21(20)*:3887-95, 2005.

Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D. & Selbig, J. pcaMethods - A bioconductor package providing PCA methods for incomplete data. *Bioinformatics, 23(9)*:1164-7, 2007.

Strauss, R. E.; Atanassov, M. N. & De Oliveira, J. A. Evaluation of the principal-component and expectation-maximization methods for estimating missing data in morphometric studies. *J. Vertebr. Paleontol., 23(2)*:284-96, 2003.

van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw., 45(3)*:1-67, 2011.

Webster, M. & Sheets, H. D. A Practical Introduction to Landmark-Based Geometric Morphometrics. *Paleontol. Soc. Pap., 16*:163-88, 2010.

Wickham, H. & Bryan, J. readxl: Read Excel Files, 2017.

Wickham, H.; Franois, R.; Henry, L. & Müller, K. dplyr: A Grammar of data manipulation. R Studio, 2019. Available from: https://cran.r-project.org/web/packages/dplyr/index.html

Xu, L. & Hong, Y. Functional and Shape Data Analysis. *J. Qual. Technol., 49(4)*:419-20, 2017.

Corresponding author:
Fatma Ezgi Can
Izmir Kâtip Celebi University
Faculty of Medicine
Department of Biostatistics
Balatcık Mh
Havaalanı Sosesi Cd.
N^{o}: 33/2 35620 Cigli/Izmir
TURKEY

E-mail:fatmaezgi.can@ikc.edu.tr