Accuracy of Linear Craniometric Measurements Obtained from Laser Scanning Created 3D Models of Dry Skulls

Diana Toneva, Silviya Nikolova, Ivan Georgiev and Assen Tchorbadjieff

Abstract The aim of this study was to establish the reliability of directly taken linear measurements on dry skulls and corresponding measurements taken on the 3D digital models created by laser scanning as well as to assess the agreement between both measuring methods. Four skulls were measured in two competitive methods—a direct measuring, based on the conventional craniometric method, and a digital measuring, accomplished on 3D models created by laser scanning. Thirteen cranial measurements were taken on both dry skulls and 3D models. The intra- and inter-examiner reliability was estimated using intraclass correlation coefficient. The agreement between both measuring methods was assessed applying the Bland-Altman method for replicated measurements. A Bland-Altman plot was constructed for each of the 13 parameters. The 3D model and directly taken measurements were assessed as highly reliable and reproducible, excepting the orbital height. Our results showed that 96 % of all digital measurements differ from the directly taken ones with less than 2 mm and respectively 67.6 % differ with less than 1 mm. Based on the results of the Bland-Altman plots, most of the measurements obtained by both measuring methods could be accepted as comparable, since the majority of differences were within the constructed limits of agreement. However, there were digital measurements, particularly these with landmarks situated on bone margins, which systematically overestimated the directly taken ones.

D. Toneva (✉) · S. Nikolova
Institute of Experimental Morphology, Pathology and Anthropology with Museum,
Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
e-mail: ditoneva@abv.bg

S. Nikolova
e-mail: sil_nikolova@abv.bg

I. Georgiev
Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
e-mail: ivan.georgiev@parallel.bas.bg

I. Georgiev · A. Tchorbadjieff
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
e-mail: atchorbadjieff@math.bas.bg

K. Georgiev et al. (eds.), *Advanced Computing in Industrial Mathematics*,
Studies in Computational Intelligence 681, DOI 10.1007/978-3-319-49544-6_18

Keywords Laser scanning · Craniometry · 3D models · Accuracy · Reliability

1 Introduction

The pursuit for a higher precision and quality in scientific research has been growing very fast in the past decades. This requires designing and implementation of new types of research equipment and development of new methods for research data analysis. The results from the new developed research methods are expected to be more precise and less disputable. However, as a first step every new method or equipment must be evaluated and compared to the theory and existing methods and data. This first and basic procedure is the inter-comparison between equivalent data obtained by new and old methods.

Because of the various applications and increasing accessibility of the imaging technologies in the medical and scientific fields, the accuracy of the digitally obtained metrical characteristics has been discussed in many studies. The most reported technology in the literature with regard to the reliability of the digital measurements is the computed tomography (CT) and especially cone beam computed tomography (CBCT), because of its growing application in the dental and orthodontic practice. The use of dry human skulls has been a traditional approach to validate new craniofacial imaging modalities [3]. The accuracy of linear measurements has been estimated on 2D tomographic slices and 2D cephalograms [11, 14, 16, 17, 19], but the results show some drawbacks, such as perspective limitations and positioning errors. Other studies have been dedicated to the assessment of the accuracy of cranial measurements made on 3D volumetric representations from CT and CBCT scans [2, 3, 10, 14, 15, 18, 23]. It has been established that the craniometric CBCT measurements are accurate and reliable, although most of the authors have noticed that the CT measuring data tend to be slightly lower than the conventional ones. Other authors have compared the measurements of the skulls with their replica models produced by rapid prototyping [9, 21], concluding that the models are extremely accurate, although slightly bigger than the original objects.

The CT technology is a suitable imaging method for diagnostic and therapeutic purposes in the medical practice and gives opportunity for various additional metrical analyses as well as for creating a large virtual database from patients' data. However, it should be applied more cautiously in the work with bone remains from archaeological excavations or forensic contexts. The exposure of a bone to clinical levels of radiation has been presumed to reduce the amount of amplifiable DNA [12] and thus, the extraction of ancient DNA from bone remains revealed in archaeological excavations could be obstructed in case of a foregoing CT-scanning. Therefore, if only surface data are needed for the purposes of an investigation, the laser scanning appears to be more suitable method for capturing data, than the hazardous exposure of the bones to X-rays.

The hand-held laser scanners appear to be very useful for digitizing bone samples because of their portability, easy manipulating with digital models and reduced risk of damages of the real objects [4]. The three-dimensional (3D) digital models created by laser scanning have an increasing application in the field of physical anthropology in recent years. Since the skull is the most investigated part of the skeleton, being an important source of information for a variety of anthropological studies, the agreement between the conventional and digitally taken standard cranial measurements is a crucial point for the future craniometric investigations. However, despite the widening usage of hand-held laser scanners in paleoanthropology and forensic anthropology and lots of applications of the created 3D bone models, there have not been many studies, concerning the precision of the cranial measurements obtained on laser scanning created 3D models [22, 23, 26]. The use of the cranial measurements in sex determination, race investigations, personal identification, etc., indicates the necessity of more studies comparing the accuracy between conventional and 3D digital measuring methods.

In this study, we aimed to establish the reliability of the directly taken linear measurements on dry skulls and corresponding measurements taken on the 3D digital models created by laser scanning as well as to assess the agreement between both measuring methods. There have been used different statistical methods for assessment of the methods agreement—by absolute difference, different correlation coefficients, linear regression, Bland-Altman plot, etc. We chose to apply the Bland-Altman method for measuring agreement using replicated measurements, which has been rarely performed because of its more complicated computation, but appearing to be most suitable for our purpose. Moreover, the traditional Bland-Altman plot, being a very illustrative method, has been widely used in clinical research.

2 Material and Methodology

2.1 Material

The subject of this inter-comparison study were four skulls from the osteological collection at the Institute of Experimental Morphology, Pathology and Anthropology with Museum, Bulgarian Academy of Sciences. The skulls belonged to adult individuals from both sexes [6].

2.2 Methodology

2.2.1 Measurements

The skulls were measured in two competitive methods—a direct measuring, based on the conventional craniometric method, and a digital measuring, accomplished on

3D models created by laser scanning. Thirteen cranial measurements were measured on both dry skulls and 3D models. The measurements represent direct distances between definite craniometric landmarks described according to Martin and Saller [20] (Table 1). The conventional measurements were taken with standard sliding and spreading calipers.

The 3D models for the digital measuring were created using hand-held laser scanner Creaform VIUscan^{TM}. The skulls were scanned without mandibles (Fig. 1). The scanning was set at a resolution of 0.7 mm and a texture resolution of 150 DPI. The accuracy of the laser scanner was to 0.050 mm. The surface image data collected by the laser scanning were post-processed in the scanner software platform VXelements^{TM}. The measurements on the 3D models (.stl) were taken using the free software Geomagic Verify Viewer (3D Systems, Inc).

All of the dry skulls and 3D models were measured three times by two examiners to test the intra- and inter-examiner reliability. Each set of 13 measurements were taken on a separate day in a random order. The replicated measurements were taken independently of each other.

2.2.2 Data

A total of 624 measurements were performed on the dry skulls and 3D models by both examiners. The data acquired from the direct and digital measuring method consisted of two series of 312 measurements. Each method series was separated by two equal parts corresponding to the measuring data of both examiners. For the purpose of comparing both methods, the data for each of the 13 measurements were grouped, as the measurements of each examiner obtained by the one method were paired with these received by the other method in accordance with the succession of the readings (1-st, 2-nd, 3-rd measuring).

Means of the measurements performed by both examiners on the dry skulls and 3D digital models were calculated over all 13 measurements (Table 2). It should be noticed that the data from the digital measurements were averaged from data with a higher precision, and the measured distances were with values to one hundredth of a millimeter, due to the advantages provided by the contemporary digital technologies. Conversely, the precision of the data acquired from the direct measurements was constrained up to the scale bar of the calipers, which is to a millimeter.

2.2.3 Statistics

The first step in the analysis was to evaluate consistency in the results between involved examiners. The intra- and inter-examiner reliability was assessed using intraclass correlation coefficient (ICC). Since in our study each of a random sample of n targets was rated independently by k judges and each target was rated by each of the same k judges, the "two-way mixed" model of ICC [27] was used:

Table 1 Craniometric measurements

Measurements	Definition	Landmarks—definition/Type[a]
Cranial length (CL)	Direct distance from glabella to opisthocranion	Glabella (g)—the most prominent point in the midsagittal plane between the superciliary arches (Type III)
Cranial base length (CBL)	Direct distance from nasion to basion	Opisthocranion (op)—the most posterior point of the skull in the midsagittal plane (Type III)
Length of *foramen magnum* (LFM)	Direct distance from basion to opisthion	Nasion (n)—the point of intersection between the frontonasal suture and the midsagittal plane (Type I)
Occipital breadth (OCB)	Direct distance between both asterion	Basion (ba)—the midline point on the anterior margin of the *foramen magnum* (Type II)
Breadth of *foramen magnum* (BFM)	Greatest distance between the lateral margins of *foramen magnum*[b]	Opisthion (o)—the midline point at the posterior margin of the *foramen magnum* (Type II)
Cranial height (CH)	Direct distance from bregma to basion	Asterion (ast)—the point of intersection of the lambdoid, occipitomastoid and parietomastoid sutures (Type I)
Facial length (FL)	Direct distance from basion to prosthion	Bregma (b)—the point of intersection between the coronal and sagittal sutures (Type I)
Upper facial breadth (UFB)	Direct distance between both frontomalare temporale	Prosthion (pr)—the most anterior point in the midline on the alveolar processes of the maxillae (Type II)
Upper facial height (UFH)	Direct distance from nasion to prosthion	Frontomalare temporale (fmt)—the most laterally positioned point on the fronto-zygomatic suture (Type II)
Middle facial breadth (MFB)	Direct distance between both zygomaxillare	Zygomaxillare (zm)—the most lower point on the zygomaticomaxillary suture (Type II)
Length of the nasal bones (LNB)	Direct distance from nasion to rhinion	Rhinion (rhi)—the point of intersection between the internasal suture and the upper end of the piriform aperture (Type II)
Orbital breadth (OBB)	Direct distance from maxillofrontale to ektoconchion	Maxillofrontale (mf)—the point of intersection between the maxillofrontal suture and the medial rim of the orbit (Type II)
Orbital height (OBH)	Direct distance between the superior and inferior orbital margins, measured perpendicular to the orbital breadth[b]	Ektoconchion (ek)—the point of intersection between the lateral rim of the orbit and the line beginning from mf crossing the orbit parallel to the upper orbital margin (Type III)

[a]The type of the landmarks is given with Roman numerals (Type I, II, and III), according to the categorical classification of the landmarks [7]
[b]BFM and OBH are measured between Type III landmarks

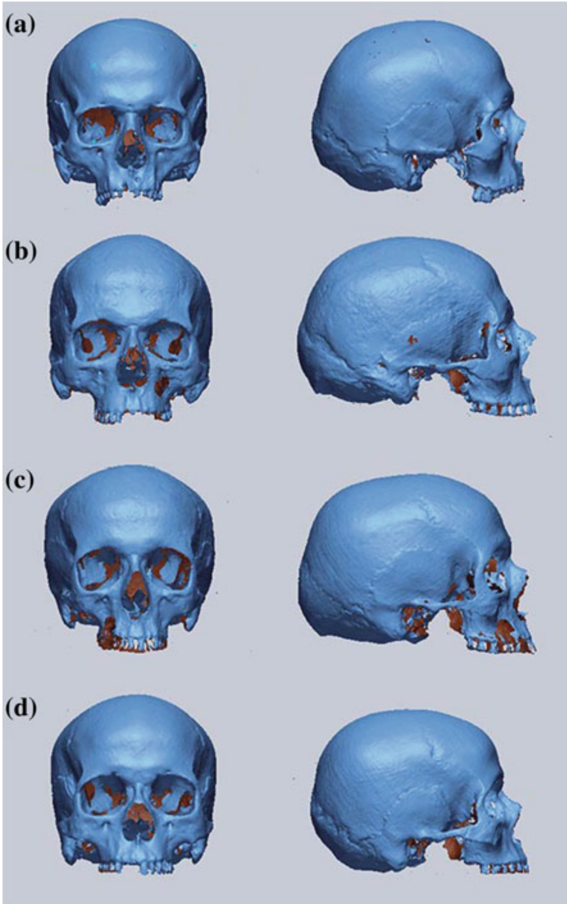

Fig. 1 Frontal and lateral views of the 3D models of the four skulls (**a**–**d**)

$$ICC\,(3,1) = \frac{BMS - EMS}{BMS + (k-1)\,EMS}$$

where with *BMS* is denoted between-targets mean k square with $n - 1$ degree of freedom, yielded from two-way ANOVA. The notation *EMS* is used for within-target residual sum of squares, with $(n - 1)(k - 1)$ degree of freedom.

The intra-examiner reliability was calculated for each measurement on the base of the triple measuring of the four skulls and 3D models, respectively ($k = 3$, $n = 4$). The inter-examiner reliability was calculated for each measurement separately for the digital and direct measuring methods based on the juxtaposition of the individual values of the triple measurements of both examiners for the four samples ($k = 2$, $n = 12$) (Table 3).

In the inter-methods comparison, the accuracy was assessed by an examination of the differences between digital and direct measurements performed by both examin-

Table 2 Means of the digital and direct measurements performed by both examiners. Means of the measurements for each measuring method based on the combined measurements of both examiners and corresponding mean differences for each parameter

Measurements	Digital measurements /mm/			Direct measurements /mm/			Mean difference /mm/
	Examiner I	Examiner II	Mean	Examiner I	Examiner II	Mean	
CL	180.48	180.46	180.47	180.92	180.83	180.88	−0.41
CBL	99.05	99.18	99.12	99.75	99.58	99.67	−0.55
LFM	35.98	35.77	35.88	35.21	34.63	34.92	0.96
OCB	110.65	110.51	110.58	110.38	110.42	110.40	0.18
BFM	30.84	30.61	30.73	29.58	29.21	29.40	1.33
CH	131.53	131.60	131.56	132.46	131.50	131.98	−0.42
FL	70.44	70.41	70.43	69.79	70.50	70.15	0.28
UFB	95.00	95.03	95.01	95.67	95.13	95.40	−0.38
UFH	101.59	101.86	101.73	102.17	102.25	102.21	−0.48
MFB	91.51	92.04	91.77	92.33	92.42	92.38	−0.60
LNB	20.90	21.12	21.01	21.21	21.58	21.40	−0.38
OBB	38.78	38.62	38.70	39.92	38.79	39.35	−0.66
OBH	33.89	34.00	33.95	33.25	32.67	32.96	0.99

Table 3 Intra- and inter-examiner reliability (ICCs)

Measurements	Intra-examiner reliability				Inter-examiner reliability	
	Digital measurements		Direct measurements		Digital measurements Examiner I/ Examiner II	Direct measurements Examiner I/ Examiner II
	Examiner I	Examiner II	Examiner I	Examiner II		
CL	0.998765	0.999417	0.999104	0.998946	0.998350	0.998027
CBL	0.998730	0.995827	0.999319	0.999095	0.997115	0.998389
LFM	0.959601	0.973110	0.988976	0.984081	0.963756	0.970985
OCB	0.992870	0.985779	0.993300	0.990424	0.982697	0.989921
BFM	0.994893	0.984771	0.972303	0.982387	0.985435	0.965900
CH	0.999877	0.999809	0.999381	0.997948	0.999635	0.996953
FL	0.958926	0.971117	0.958993	1	0.836640	0.979043
UFB	0.997212	0.995597	0.992610	0.991939	0.996457	0.986715
UFH	0.963741	0.843542	0.981203	0.980186	0.833535	0.971626
MFB	0.992704	0.997732	0.998818	0.995975	0.987033	0.987061
LNB	0.994014	0.987694	0.990733	0.997278	0.990700	0.989085
OBB	0.919665	0.956470	0.961538	0.848708	0.804901	0.834675
OBH	0.696814	0.595882	0.500000	–	0.323467	0.406780

–ICC value could not be calculated for OBH from the direct measurements of the second examiner. It was due to the lack of between-subject variance. However, the model was correct and all values on the OBH for the triple measurements of all four skulls were with a difference of no more than 1 mm

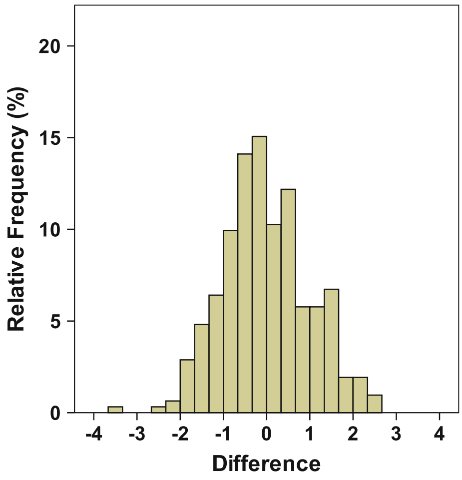

Fig. 2 Histogram of the differences of all paired measurements (n = 312)

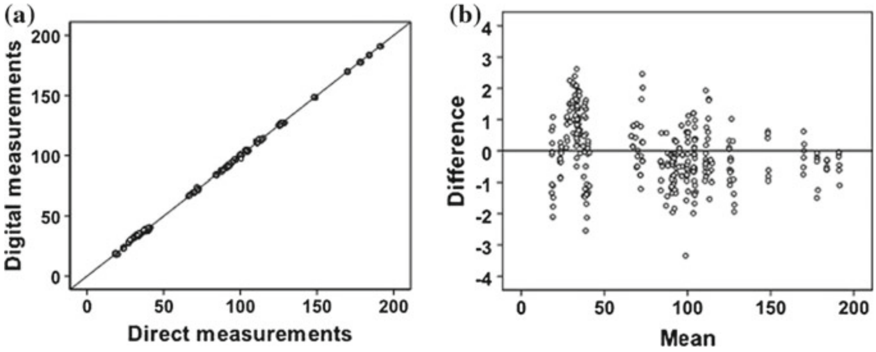

Fig. 3 Scatterplots of **a** the direct versus the digital cranial measurements with the line of equality; **b** the means versus the differences between digital and direct measurements

ers. The differences between the two methods were measured as absolute differences in millimeters /mm/. Firstly, the inter-comparison was assessed based on the mean differences between both measuring methods for all 13 parameters (Table 2). As a second step, the comparison of the methods was founded on the individual differences from the paired triple measurements of the two examiners on all four samples. The frequency of the differences between all paired measurements was illustrated by a percentage histogram (Fig. 2). A scatterplot was used to graph the correlation between the data of direct and digital measurements (Fig. 3a). A graph plotting the means against the differences between both measuring methods was used to illustrate the systematic diversion of the differences in accordance to the magnitude of the measurements [5] (Fig. 3b). The assumption of normality for the digital-direct differences of each parameter (n = 24) was tested by the Shapiro-Wilk test.

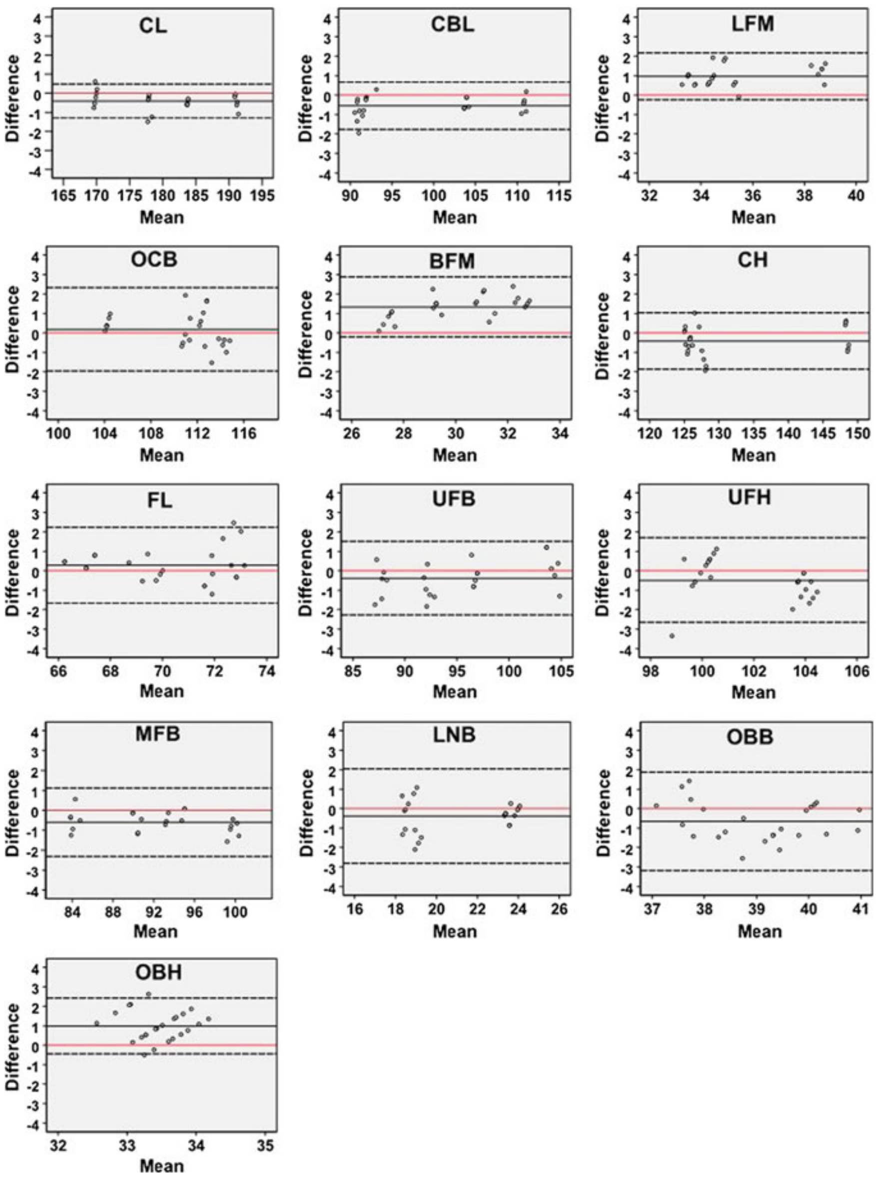

Fig. 4 Bland-Altman plots representing the bias (*continuous black line*) and the 95 % limits of agreement (*dotted lines*), based on the replicated measurements of each cranial measurement

The digital and direct measuring methods were compared applying the Bland–Altman method [1], which plots the means of the results of both methods (x-axis) and the differences between the two methods (y-axis). The methods agreement was quantified by constructing 95 % limits of agreement (LoA) for each of the 13 measurements (Fig. 4). Because of the repeated measurements, the LoA were built in the assumption of compound variance of the observed variance of the differences between the within-subject means σ_d and the within-subject variances σ_{Xw} and σ_{Yw} from measurements by the used methods (X and Y) with equal number of replicates m. The variance of the differences between means was yielded by the following formula [5]:

$$\sigma_{X-Y}^2 = \sigma_d^2 + \left(1 - \frac{1}{m}\right)\sigma_{Xw}^2 + \left(1 - \frac{1}{m}\right)\sigma_{Yw}^2$$

The values for the standard deviation were computed using the Linear Mixed-Effects Models library in R [8]. The used data set consisted of 4 columns of parameters—method, item, replication number and distance measurement in mm.

3 Results

The ICCs calculated for each examiner and between examiners showed almost perfect intra- and inter-examiner reliability (Table 3). The only exception was observed for the OBH with ICC values indicating moderate intra-examiner reliability and fair inter-examiner agreement for both measuring methods.

The means of the digitally and directly taken measurements and the corresponding mean differences for all 13 parameters are given in Table 2. Eight of the craniometric measurements on the 3D digital models were slightly smaller than the directly measured ones, as the mean differences ranged from −0.38 to −0.66 mm. The remaining five measurements, including the LFM, BFM, OCB, FL and OBH, showed higher readings on the 3D digital models than these on the dry skulls. Unlike the OCB and FL, which had the lowest absolute mean differences, the mean differences of the other three features had the highest values of all calculated ones. The bigger values of the LFM, BFM, and OBH, obtained with the software probably were due to the location of the landmarks for these measurements on bone margins, which made very difficult their placement exactly on the margin in the 3D model. Thus, these measurements required the placement of the landmarks slightly above/below the margin within the bone area so as to be caught by the program, and respectively the obtained values got higher than the directly measured ones. The digital measurements of the BFM and OBH exceeded the direct ones by an average of 1.33 mm and 0.99 mm, respectively.

The percentage histogram of the differences obtained from the paired repeated measurements of all parameters showed that they were normally distributed as the most common were differences within 1 mm (Fig. 2). Overall, our results showed that 96 % of all digital measurements differed from the directly taken ones with less than 2 mm and respectively 67.6 % differed with less than 1 mm. The scatterplot between the grouped data of both methods indicated the lack of a systematic bias, as almost all points lied on the line of equality (Fig. 3a). According to the graph plotting the means against the differences between both measuring methods (Fig. 3b), it could be observed that the divergence of the differences decreased with the magnitude of measurements. The most reasonable explanation for this trend was the observed difficulties in the precise measuring of the dimensions with small sizes. The hypothesis that the digital-direct differences of each parameter were normally distributed was confirmed by the Shapiro-Wilk test's p-values, which were higher than 0.05 ($p > 0.05$).

The Bland-Altman plot as a method for comparing different measuring methods was constructed for each of the 13 measurements. The biggest bias was observed for LFM, BFM and OBH (Fig. 4). The differences in seven of the craniometric measurements were entirely distributed in the LoA. Four of the measurements (CBL, CH, FL, and UFH) had 23/24 or 95.8 % of the differences falling within the limits. The CL and OBH showed 22/24 or 91.7 % of the differences distributed in the LoA. The analysis of the data that left outside the intervals of agreement showed that the outliers were produced completely random and there was not any relation to the selected methods of measuring or the personality of examiners. However, it should be notices that this method of analysis is not very sensitive to one or two large outlying differences and there is no need to be removed from the analysis [5].

The smallest width of the 95 % LoA was established for the CL (1.78 mm) and the biggest one was observed for the OBB (5.04 mm). The very wide LoA observed for the LNB and OBB were due respectively to the bigger between-subject variance and the bigger within-subject variance of the direct method. Most of the Bland-Altman plots evidenced for an agreement between both measuring methods, except for the LFM, BFM and OBH, because of the very big bias. However, only five measurements had a width of the LoA less than 3 mm (CL, CBL, LFM, CH and OBH), which indicated quite wide LoA for the most measurements. A reason for the wide LoA intervals in our study could be the small sample size. Besides, it has been noticed that when the method is applied using replicated measurements, the LoA intervals are wider compared to the variant with the means of each measuring method [5].

4 Discussion

The reliability and accuracy of the digital measurements obtained on laser scanning created 3D models are key points at the time of an increasing usage of hand-held laser scanners in paleoanthropology and forensic anthropology. Concerning the reliability of the 3D laser scanning method, there have been established excellent results for the linear craniometric measurements [22] as well as precise ones for the surface area and volume measurements [26]. Our study is not an exception with almost perfect intra- and inter-examiner reliability for nearly all digitally and directly taken measurements, except for the OBH indicating moderate intra-examiner reliability and fair inter-examiner agreement. The results obtained for this parameter could due to the type of landmarks (Type III) defining this measurement as well as to the small sample size. However, namely Type III landmarks have been reported to yield the most precise coordinate data on 3D laser scanner models [25], as the measurements between them have been found to be very consistent [23]. On the other hand, Type

III landmarks have been established to be lowly reproducible according to a study based on the coordinate landmarks data obtained by 3D digitizing [24]. Although, in our study, exactly a measurement between these landmarks showed poor results on the reliability, the other two measurements defined by the same type landmarks had excellent ones. So it should be considered that the separate Type III landmarks and the measurements defined by them could show varying reproducibility, depending not only on the choice of the 3D technique but also on the choice of the landmarks, and thus to lead to different results and conclusions.

According to the used inter-comparison technology, the accuracy of the cranial measurements was found to differ to a varying degree. Because of the manual measuring and the easier landmark identification directly on the real objects, the dry skulls and their produced prototypes have shown least differences in their metrical characteristics [23]. Comparing the digital technologies, the laser scans have provided more accurate measurements compared to the CT ones, due to the interpolation of the data between CT scan slices at the 3D rendering [23]. The 3D laser scanning method has been reported to give a slightly lower reading compared to the conventional measuring [22]. However, such a tendency has not been observed in other studies [23, present study]. Our results showed that there were even a few digital measurements such as LFM, BFM, and OBH, showing a systematic overestimation of the direct measurements.

It worth noting that the providing of an accurate digital metrical analysis requires a very good quality of the 3D models with well captured surface in the places of all investigated landmarks. The suture-based landmarks (or Type I) have been reported to be very problematic for identification on 3D models, as the measurements between such landmarks have shown a greater variation in the measurement error [23]. In our study, the measurements with one or two Type I landmarks did not indicate consistent big differences between both measuring methods, but some of them were also highly variable. As a whole, Type I landmarks have been reported to be more precisely identified when they are collected with a digitizer than on 3D models [24, 25].

The landmark location has been suggested to be a major source of variability in the measurements depending mostly on the human judgement [13]. This could be a reason to some extent for the inaccuracy in the measuring of the BFM and OBH in our study, which were described indefinitely as a largest diameter and a perpendicular to some another measurement (i.e. include only Type III landmarks), and thus, the personal assessment of the examiner is supposed to be a substantial factor. The lower intra- and inter-examiner reliability observed for the OBH could cause to a certain degree the inconsistency between both measuring methods, but this cannot be an explanation in the case of the BFM. However, as we previously noticed there were software caused difficulties in the landmark location for these measurements, so the measuring differences could not be specified as only human dependent. It should be also taken into account the big difference in the precision of the used measuring techniques, although our results showed that all digital-direct differences with few exceptions were within 2 mm, which have been considered as an acceptable amount of error in forensic anthropology [28].

5 Conclusion

The ICCs indicated that digital and directly taken measurements were highly reliable and reproducible, except for the OBH. Overall, almost all digital measurements differed from the directly taken ones with less than 2 mm and respectively 2/3 of them differed with less than 1 mm.

Based on the results of the Bland-Altman plots, most of the measurements obtained by both measuring methods could be accepted as comparable, since the majority of differences were within the LoA intervals. However, there were digital measurements, particularly these with landmarks situated on bone margins, which systematically overestimated the directly taken ones and should be considered with more attention when it concerns to a 3D digital measuring method.

Acknowledgements This study was supported by the "Program for Career Development of Young Scientists, Bulgarian Academy of Sciences", research grant DFNP—75/27.04.2016.

References

1. Altman, D.G., Bland, J.M.: Measurement in medicine: the analysis of method comparison studies. Statistician **32**, 307–317 (1983)
2. Baumgaertel, S., Palomo, J.M., Palomo, L., Hans, M.G.: Reliability and accuracy of cone-beam computed tomography dental measurements. Am. J. Orthod. Dentofac. Orthop. **136**(1), 19–25 (2009)
3. Berco, M., Rigali Jr., P.H., Miner, M.R., DeLuca, S., Anderson, N.K., Will, L.A.: Accuracy and reliability of linear cephalometric measurements from cone-beam computed tomography scans of a dry human skull. Am. J. Orthod. Dentofac. Orthop. **136**, 17.e1–17.e9 (2009)
4. Bibliowicz, J., Khan, A., Agur, A., Singh, K.: High-precision surface reconstruction of human bones from point-sampled data. In: ISHS 2011 Conference Proceedings: International Summit on Human Simulation, pp. 1–10 (2011)
5. Bland, J.M., Altman, D.G.: Measuring agreement in method comparison studies. Stat. Methods Med. Res. **8**, 135–160 (1999)
6. Buikstra, J.E., Ubelaker, D.H.: Standards for data collection from human skeletal remains. In: Arkansas Archeological Survey Research Series No. 44, Fayetteville (1994)
7. Bookstein, F.L.: Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, Cambridge (1991)
8. Cartensen, B., Simpson, J., Gurrin, L.C.: Statistical models for assessing agreement in method comparison studies with replicate measurements. Int. J. Biostat. **4**(1), 1–26 (2008)
9. Choi, J.Y., Choi, J.H., Kim, N.K., Kim, Y., Lee, J.K., Kim, M.K., Lee, J.H., Kim, M.J.: Analysis of errors in medical rapid prototyping models. Int. J. Oral Maxillofac. Surg. **31**(1), 23–32 (2002)
10. Damstra, J., Fourie, Z., Huddleston Slater, J.J., Ren, Y.: Accuracy of linear measurements from cone-beam computed tomography-derived surface models of different voxel sizes. Am. J. Orthod. Dentofac. Orthop. **137**(1), 16.e1–16.e6 (2010)
11. Gribel, B.F., Gribel, M.N., Frazao, D.C., McNamara Jr., J.A., Manzi, F.R.: Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans. Angle Orthod. **81**, 26–35 (2011)
12. Grieshaber, B.M., Osborne, D.L., Doubleday, A.F., Kaestle, F.A.: A pilot study into the effects of X-ray and computed tomography exposure on the amplification of DNA from bone. J. Archaeol. Sci. **35**(3), 681–687 (2008)

13. Harris, E.F., Smith, R.N.: Accounting for measurement error: a critical but often overlooked process. Arch. Oral. Biol. **54**(Suppl. 1), S107–S117 (2009)
14. Hassan, B., Setelt, P., Sanderink, G.: Accuracy of three-dimensional measurements obtained from cone beam computed tomography surface-rendered images for cephalometric analysis: influence of patient scanning position. Eur. J. Orthod. **23**, 1–6 (2009)
15. Kamburoğlu, K., Kolsuz, E., Kurt, H., Kiliç, C., Özen, T., Paksoy, C.S.: Accuracy of CBCT measurements of a human skull. J. Digit. Imaging **24**(5), 787–793 (2011)
16. Kumar, V., Ludlow, J.B., Mol, A., Cevidanes, L.: Comparison of conventional and cone beam CT synthesized cephalograms. Dentomaxillofacial Radiol. **36**, 263–269 (2007)
17. Lascala, C., Panella, J., Marques, M.: Analysis of the accuracy of linear measurements obtained by cone beam computed tomography (CBCT-NewTom). Dentomaxillofac Radiol. **33**, 291–294 (2004)
18. Lorkiewicz-Muszyńska, D., Kociemba, W., Sroka, A., Kulczyk, T., Żaba, C., Paprzycki, W., Przystańska, A.: Accuracy of the anthropometric measurements of skeletonized skulls with corresponding measurements of their 3D reconstructions obtained by CT scanning. Anthropol. Anz. **72**(3), 293–301 (2015)
19. Ludlow, J.B., Laster, W.S., See, M., Bailey, L.J., Hershey, H.G.: Accuracy of measurements of mandibular anatomy in cone beam computed tomography images. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. **103**, 534–542 (2007)
20. Martin, R., Saller, K.: Kraniometrische Technik. In: Martin, R., Saller, K. (eds.) Lehrbuch der Anthropologie, vol. I. Gustav Fischer, Stuttgart (1957)
21. Nizam, A., Gopal, R.N., Naing, L., Hakim, A.B., Samsudin, A.R.: Dimensional accuracy of the skull models produced by rapid prototyping technology using stereolithography apparatus. Arch. Orofac. Sci. **1**, 60–66 (2006)
22. Park, H.K., Chung, J.W., Kho, H.S.: Use of hand-held laser scanning in the assessment of craniometry. Forensic. Sci. Int. **160**, 200–206 (2006)
23. Richard, A.H., Parks, C.L., Monson, K.L.: Accuracy of standard craniometric measurements using multiple data formats. Forensic. Sci. Int. **242**, 177–185 (2014)
24. Ross, A.H., Slice, D.E., Williams, S.E.: Geometric Morphometric Tools for the Classification of Human Skulls. Department of Justice, Document 231195 (2010)
25. Sholts, S.B., Flores, L., Walker, P.L., Wärmländer, S.K.: Comparison of coordinate measurement precision of different landmark types on human crania using a 3D laser scanner and a 3D digitiser: implications for applications of digital morphometrics. Int. J. Osteoarchaeol. **21**, 535–543 (2011)
26. Sholts, S.B., Wärmländer, S.K., Flores, L.M., Miller, K.W., Walker, P.L.: Variation in the measurement of cranial volume and surface area using 3D laser scanning technology. J. Forensic. Sci. **55**(4), 871–876 (2010)
27. Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. **86**(2), 420–428 (1979)
28. Stull, K.E., Tise, M.L., Ali, Z., Fowler, D.R.: Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. Forensic. Sci. Int. **238**, 133–140 (2014)