3D printing of anatomical models for surgeons: an investigation on repeatability

Jean H. D. Fasel, Didier D. Malis, Carina Wiederer & Niels Hagenbuch

International Journal on Interactive Design and Manufacturing (IJIDeM)

ISSN 1955-2513
Volume 12
Number 2

Int J Interact Des Manuf (2018)
12:621-627
DOI 10.1007/s12008-018-0478-4

Springer

International Journal on Interactive Design and Manufacturing (IJIDeM) (2018) 12:621–627
https://doi.org/10.1007/s12008-018-0478-4

ORIGINAL PAPER

3D printing of anatomical models for surgeons: an investigation on repeatability

Jean H. D. Fasel[1] · Didier D. Malis[2] · Carina Wiederer[3] · Niels Hagenbuch[4]

Received: 7 March 2018 / Accepted: 3 April 2018 / Published online: 12 April 2018

Abstract
As part of the trend towards personalised medicine, surgeons are increasingly using 3D printed replicas for preoperative planning. This raises the question of how reliable these models are. This paper examines the repeatability of manufacturing human mandibles. Five polyamide replicas were produced using selective laser sintering and digitised using structured light scanning. Quantitative comparisons were made using Mimics Software. The differences were analysed graphically, using histograms and kernel density estimates. The mean differences ranged between +0.0274 (SD 0.0671) mm and −0.0284 (SD 0.0629) mm. The median of absolute differences was 0.0308 mm, i.e. 50% of absolute differences were smaller than 31 μm. For the 22,811,168 differences measured, all were between +1.9836 and −2.0526 mm. The proportion of absolute differences below 0.10 mm was between 82.09 and 98.84%, and between 94.43 and 99.90% when using a threshold of 0.20 mm. 99.95% of the absolute differences were below 1.00 mm. In conclusion: 3D printed models may not be identical, even when based on the same imaging study and patient; on the other hand, identical replicas can be obtained with a constant production chain; we recommend that four distinguishing criteria should be used in future investigations: qualitative and quantitative accuracy, repeatability and reproducibility.

Keywords Additive manufacturing · 3D printing · Anatomical models · Repeatability · Preoperative planning

1 Introduction

The use of additive manufacturing techniques in medicine is growing rapidly. In surgery, 3D printing is most commonly used to fabricate anatomical replicas for preoperative planning [1]. However, there are questions about the quality of models made using current production processes. It has been shown that replicas produced for a single patient, based on identical source files, can vary markedly [2]. Reproduction artifacts are observed [3], and the accuracy of the models is often unsatisfactory [4].

During additive manufacturing many factors combine to produce the final outcome. Given the growing trend towards personalised medicine and individually tailored procedures there is a need to systematically investigate how different parameters in the workflow contribute to discrepancies and affect the quality of the final model.

The primary focus of this study was to evaluate the variability of models produced with a constant workflow. Our working hypothesis was that an unchanged production process should produce identical replicas. This paper investigates mandibular replicas obtained by selective laser sintering (SLS), as a common example from maxillofacial surgery.

2 Methods

2.1 Replica production

A digital model (in STL format) of an adult human mandible from our anatomy museum was obtained by micro-computed

✉ Jean H. D. Fasel
jean.fasel@unige.ch

1 Clinical Anatomy Research Group, Anatomy Sector, Departments of Cell Physiology and Metabolism, and Surgery, University Medical Centre and Hospitals, University of Geneva, 1 Rue M Servet, 1211 Geneva, Switzerland

2 Department of Oral and Maxillofacial Surgery, Clinique des Grangettes, 110 Route de Chêne, 1224 Chêne-Bougeries, Switzerland

3 Materialise GmbH, Gilching, Germany

4 Constat, Spiez, Switzerland

tomography (μCT; GE phoenix v/tome/x m 240, isotropic voxel size 65 μm). Five replicas from this file (denoted R2-R6) were produced in PA2200 (polyamide 12) by selective laser sintering (SLS). The printer used was an EOS P385, with PSW 3.2 software (Krailling, Germany). Manufacturer specifications state its precision to be 0.10 mm in the x- and y-axis (printing plane), and 0.15 mm in the (vertical) z-axis.

2.2 Qualitative replica evaluation

The qualitative assessment consisted of visually identifying all structures itemised in the current international anatomical nomenclature [5] on each replica. The identical reproduction of these elements was examined by comparative inspection of the replicas.

2.3 Quantitative replica evaluation

The quantitative investigation was based on structured light surface scanning (ATOS Core 200, GOM, Braunschweig, Germany). The resulting digital model of each replica was then compared with that of each other replica, as follows: the two models were computationally overlaid using "N-Points-Registration" and "Global Registration" algorithms (Mimics Innovation Suite Research, Version 19.0, Materialise, Leuven, Belgium). The deviations between the aligned replicas were calculated using the "Create Part Comparison Analysis" operation of the Analysis module (precision 0.0001 mm). All replica pairs were compared twice: i.e. by taking one model as the reference and the other as the object under evaluation, and vice versa. For each comparison, the distances between more than 1 million points were computed. The results are presented numerically, as histograms, and colour-coded images.

2.4 Statistical analysis

20 directional comparisons of the five replicas were made. The differences measured were analysed graphically, using histograms and kernel density estimates (Gaussian kernel, bandwidth 0.00245, chosen empirically). The proportion of the absolute differences below the limits 0.10 mm (precision of the 3D printer in the horizontal plane), 0.15 mm (thickness of a printed layer), 0.20, 0.50 and 1.00 mm were calculated. As summary statistics, the mean and standard deviation are reported for signed differences, and median and interquartile range for absolute differences. The calculations and plots were done in R (version 3.4.1) [6], using the lattice package (version 0.20-35) [7].

3 Results

3.1 Qualitative results

The qualitative comparison found that all structures denoted in the international anatomical nomenclature [5] were present and reproduced in an identical manner in all the replicas. The individual characteristics were consistently recreated. For example, the following was observed in the teeth: post extraction edentulous spaces for the molars 36 and 46; a fine smooth alveolar ridge, without scars, corresponding to space 46; a larger ridge in space 36, the location of the precedent placement of the roots still adumbrative; mesial inclination of the teeth on both sides distal to the edentulous space; numerous cortical fenestrations of the mandible, suggesting generalised periodontitis; remaining molars with classical dental anatomy, including 4 distinguished cuspids.

Anatomical variations were also uniformly reproduced on all replicas, for example accessory foramina above the genial spines [8]. Similarly, artifacts, such as marks from drill holes and SLS printing lines, were consistently rendered on all replicas.

3.2 Quantitative results

The signed differences between the replicas are summarised in Table 1. The summary statistics for the absolute differences are shown in Table 2.

The mean differences (in mm) were between $+0.0274$ (standard deviation SD 0.0671; for R4 vs. R6) and -0.0284 (SD 0.0629; for R6 vs. R4). The median absolute differences lay between 0.0237 (interquartile range from 0.0109 to 0.0434; for R2 vs. R5 and R5 vs. R2) and 0.0370 (0.0168 to 0.0740; for R6 vs. R5), in other words, between 24 and 37 μm.

The median of the mean differences (in mm) was 0.0002 (interquartile range from -0.0134 to 0.0136). The median of the medians of the absolute differences was 0.0339 (0.0272 to 0.0355).

The total of 22,811,168 measured differences was between $+1.9836$ and -2.0526 mm (both R2 vs. R5). The absolute differences lay between 6×10^{-9} and 2.0526 mm. The median of all absolute differences was 0.0308 mm. Alternatively expressed, 50% of all absolute differences were smaller than 31 μm.

The distribution of the differences was centred around 0 mm and followed a symmetric, unimodal bell shape (Fig. 1). A direct comparison of all 20 densities (Fig. 2) showed close similarity with respect to shape. Almost all values were within -0.2 and 0.2 mm.

The proportion of absolute differences below 0.10 mm was between 82.09% (R6 vs. R5) and 98.84% (R3 vs. R2), and between 94.43% (R6 vs. R5) and 99.90% (R4 vs. R2)

Table 1 Summary statistics of the signed differences between the replicas (R2–R6)

	R2	R3	R4	R5	R6
R2 (n = 1′223′140)					
Mean		0.0157	− 0.0084	0.0032	0.0215
SD		0.0379	0.0475	0.0614	0.0591
Median		0.0161	− 0.0074	0.0005	0.0210
Maximum +		0.8912	1.3869	1.9836	1.4754
Maximum −		− 0.8405	−1.2556	− 2.0526	−1.5663
R3 (n = 1′182′340)					
Mean	− 0.0141		−0.0205	− 0.0104	0.0086
SD	0.0370		0.0602	0.0600	0.0643
Median	− 0.0148		−0.0170	− 0.0156	0.0067
Maximum +	1.1976		1.4896	1.9176	1.6233
Maximum −	− 1.2210		−1.5395	− 1.7901	−1.6551
R4 (n = 1′103′111)					
Mean	0.0098	0.0221		0.0129	0.0274
SD	0.0439	0.0590		0.0493	0.0671
Median	0.0087	0.0186		0.0112	0.0224
Maximum +	0.9781	1.0031		1.5355	1.4145
Maximum −	− 0.9536	−0.9371		− 1.3688	−1.2234
R5 (n = 1′030′464)					
Mean	− 0.0036	0.0117	− 0.0132		0.0176
SD	0.0453	0.0521	0.0442		0.0867
Median	− 0.0002	0.0172	− 0.0111		0.0196
Maximum +	0.9914	0.5844	1.1767		1.4119
Maximum −	− 1.0696	−0.7780	− 1.0468		−1.4424
R6 (n = 1′163′737)					
Mean	− 0.0225	−0.0087	− 0.0284	−0.0181	
SD	0.0539	0.0610	0.0629	0.0903	
Median	− 0.0216	−0.0066	− 0.0237	−0.0197	
Maximum +	0.7597	0.8169	0.8751	1.4042	
Maximum −	− 0.7535	−0.8125	− 1.1608	−1.2995	

The top row indicates the reference replica, the left column the replica tested. Next to the latter, the number of measuring points is given. Listed (in mm) are mean, standard deviation, median, maximum positive and negative differences

when using a threshold of 0.20 mm. With a limit of 0.50 mm, in 19 out of 20 comparisons, 99.91% of the absolute differences were smaller. In all comparisons, 99.95% of the absolute differences were below 1.00 mm.

4 Discussion

4.1 Terminology

The terms accuracy, precision, repeatability und reproducibility are not consistently used in the clinical literature, which can cause misinterpretations of study results. Based on Bartlett and Frost [9], Ender and Mehl [10], and Liu et al. [11] we propose that future investigations on the reliability of 3D printed preoperative models use the following definitions:

- "Qualitative accuracy" means closeness in shape, i.e. the evaluated replica shows all and only the anatomical structures on the reference object, visible to the naked eye.
- "Quantitative accuracy" refers to closeness in dimension, i.e. how closely the dimensional values measured on the replica correspond to the values obtained on the reference object.
- "Precision" of a production chain can be subdivided into "repeatability" and "reproducibility". Both express the degree of identity of the replicas, produced and measured under identical conditions for the former, and under differing conditions for the latter (different laboratories,

Table 2 Summary statistics of the absolute differences between the replicas (R2–R6)

	R2	R3	R4	R5	R6
R2 (n = 1′223′140)					
Median		0.0245	0.0279	0.0237	0.0343
1. Quartile		0.0117	0.0130	0.0109	0.0161
3. Quartile		0.0415	0.0485	0.0434	0.0608
< 0.10 mm		98.67%	97.34%	95.54%	91.26%
< 0.15 mm		99.58%	99.52%	98.63%	97.89%
< 0.20 mm		99.76%	99.73%	99.39%	99.51%
< 0.50 mm		99.97%	99.95%	99.82%	99.92%
< 1.00 mm		100.00%	100.00%	99.95%	99.99%
R3 (n = 1′182′340)					
Median	0.0243		0.0353	0.0332	0.0346
1. Quartile	0.0115		0.0161	0.0159	0.0160
3. Quartile	0.0411		0.0655	0.0560	0.0635
< 0.10 mm	98.84%		89.21%	93.78%	88.99%
< 0.15 mm	99.72%		97.37%	98.32%	97.10%
< 0.20 mm	99.85%		99.58%	99.52%	99.61%
< 0.50 mm	99.98%		99.96%	99.91%	99.94%
< 1.00 mm	100.00%		100.00%	99.98%	99.99%
R4 (n = 1′103′111)					
Median	0.0283	0.0360		0.0273	0.0360
1. Quartile	0.0131	0.0163		0.0125	0.0161
3. Quartile	0.0494	0.0668		0.0489	0.0708
< 0.10 mm	97.35%	88.62%		96.70%	85.15%
< 0.15 mm	99.74%	97.15%		99.33%	94.05%
< 0.20 mm	99.90%	99.65%		99.73%	99.00%
< 0.50 mm	99.99%	99.99%		99.92%	99.95%
< 1.00 mm	100.00%	100.00%		99.99%	100.00%
R5 (n = 1′030′464)					
Median	0.0237	0.0335	0.0268		0.0364
1. Quartile	0.0110	0.0160	0.0123		0.0164
3. Quartile	0.0433	0.0561	0.0480		0.0717
< 0.10 mm	95.85%	94.12%	96.87%		83.27%
< 0.15 mm	98.99%	98.56%	99.41%		90.38%
< 0.20 mm	99.76%	99.73%	99.85%		95.04%
< 0.50 mm	99.99%	100.00%	99.98%		99.95%
< 1.00 mm	100.00%	100.00%	100.00%		100.00%
R6 (n = 1′163′737)					
Median	0.0351	0.0350	0.0363	0.0370	
1. Quartile	0.0164	0.0162	0.0162	0.0168	
3. Quartile	0.0622	0.0648	0.0708	0.0740	
< 0.10 mm	90.96%	88.75%	85.67%	82.09%	
< 0.15 mm	98.11%	97.33%	94.81%	89.42%	
< 0.20 mm	99.76%	99.77%	99.25%	94.43%	
< 0.50 mm	100.00%	99.99%	99.99%	99.94%	
< 1.00 mm	100.00%	100.00%	100.00%	100.00%	

The top row indicates the reference replica, the left column the replica tested. Next to the latter, the number of measuring points is given. Shown (in mm) are median, first, and third quartile. The proportion of absolute differences below the limits 0.10, 0.15, 0.20, 0.50 and 1.00 mm are listed (in %)

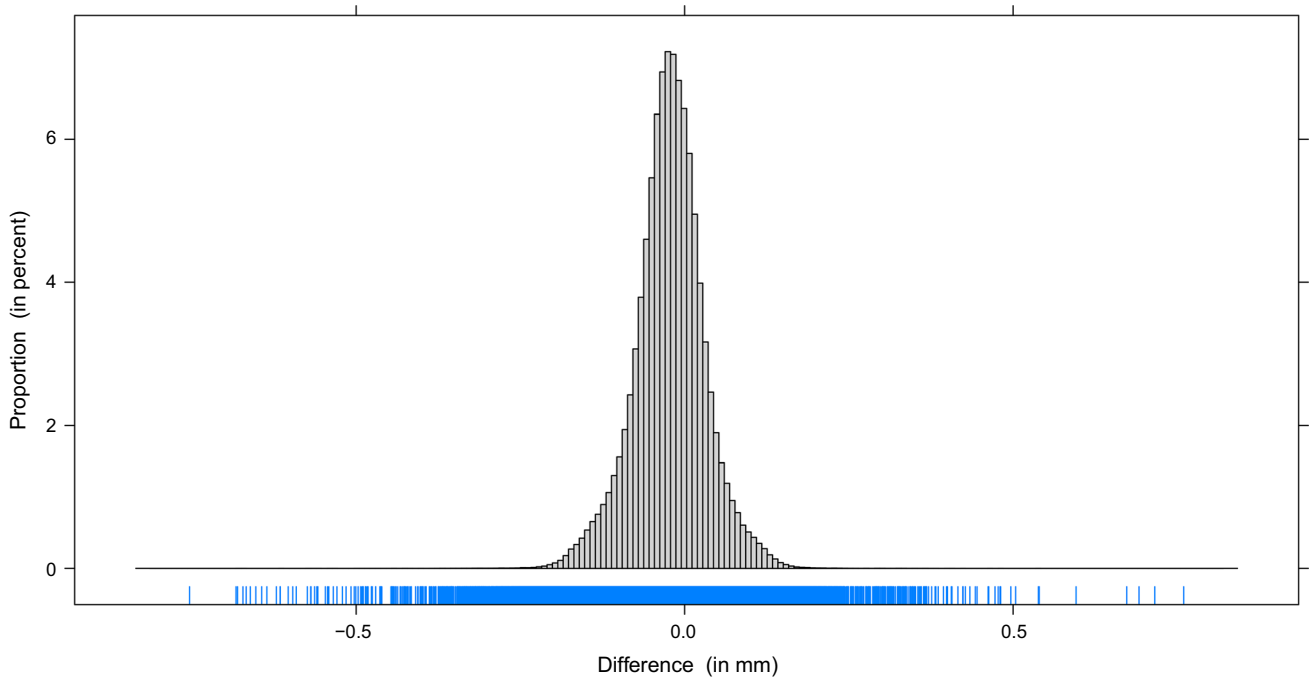

Fig. 1 Histogram showing comparison of replica 6 versus replica 2. Differences (in mm) obtained from 1,163,737 measuring points (200 bins, width 0.0082 mm).

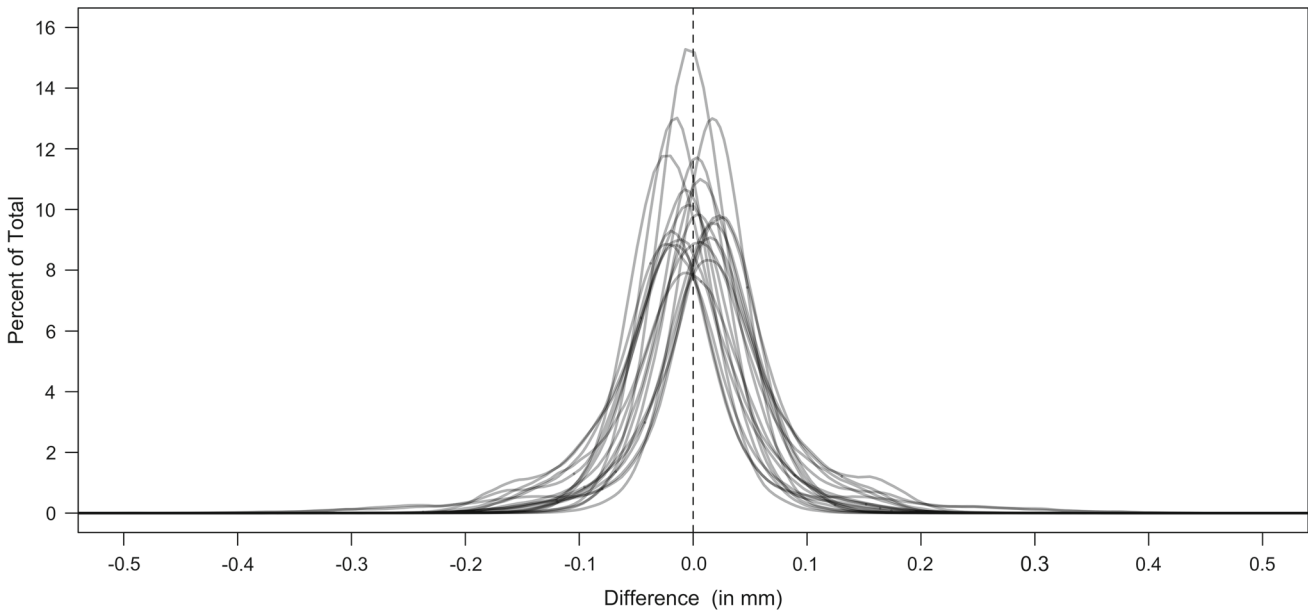

Fig. 2 Density estimators of differences for the 20 comparisons. Almost all values are within − 0.2 and 0.2 mm.

different operators, different measurement techniques etc).

Using this terminology, the present investigation is a repeatability study for a part of the production chain (identical STL file, identical work flow, identical measurement method), using qualitative and quantitative accuracy of the replicas between themselves as criteria.

4.2 Results

The examined replicas were qualitatively accurate, without exception, according to the above definition. Individual characteristics, for example in the case of the teeth, as well as anatomical variations were uniformly reproduced across the replicas. This was also the case for artifacts, such as the orientation of the printing lines.

In terms of quantitative accuracy, the median absolute differences lay between 24 and 37 μm (Table 2). The maximum differences were +1.9836 and −2.0526 mm (Table 1). The proportion of differences with absolute values smaller than 0.10 mm was between 82.09 and 98.84%, and between 94.43 and 99.90%, when a 0.20 mm limit was used. Using a threshold of 0.50 mm, 19 of the 20 comparisons had over 99.91% agreement, while with a 1.00 mm limit, at least 99.95% of differences were smaller. Based on these results, we conclude that the replicas were quantitatively accurate, provided, however, that differences under 1.0 mm in at least 99.95% of more than 1 million measure points are accepted as the tolerance threshold.

4.3 Methods

4.3.1 Additive production methods

Many factors influence the final product obtained by additive manufacturing methods [12]. Especially critical are the radiological acquisition parameters, the choice of threshold for segmentation, the transformation into a printer compatible format, the type of additive technology and the printing material used. The high number of these influential parameters explains why replicas, even when derived from the same scan, can feature differences [2]. The intuitive assumption, that a given CT will necessarily result in identical replicas, can not be taken for granted. For the same reason, this study can only provide evidence for the repeatability of the production chain examined. The results cannot be generalised to include other workflows. Indeed, each production pipeline must be individually evaluated.

4.3.2 Investigation methods

Different methods are available for digitising and measuring physical objects to be investigated for manufacturing quality control and nominal-actual comparisons. These include tactile and non-tactile optical surface and micro-computed tomography scanning. Non-contact optical scanners allow a large number of points to be acquired and compared (in our study over 1 million in each comparison). One of the disadvantages of this method is that certain geometries cannot be properly captured [13]. Figure 3 illustrates an instance of such a shape, a deep narrow interdental crevice.

This limitation has to be taken into account in the interpretation of the results. For example, the largest observed absolute difference was 2.0526 mm (Table 1). It occured in the depth of the mandibular foramen, i.e. in an area not capturable by surface scanning (Fig. 4). Thus this deviation is likely to correspond to a methodological artifact and not to a real physical difference in the replicas. It could then be assumed that the maximum differences are actually smaller

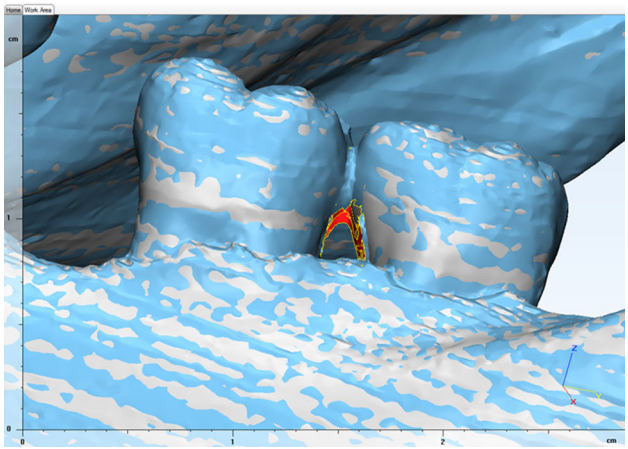

Fig. 3 Example of a geometry that cannot be captured by surface scanning methods, such as deep narrow fissures. The Figure illustrates the defect (red area) in an interdental space

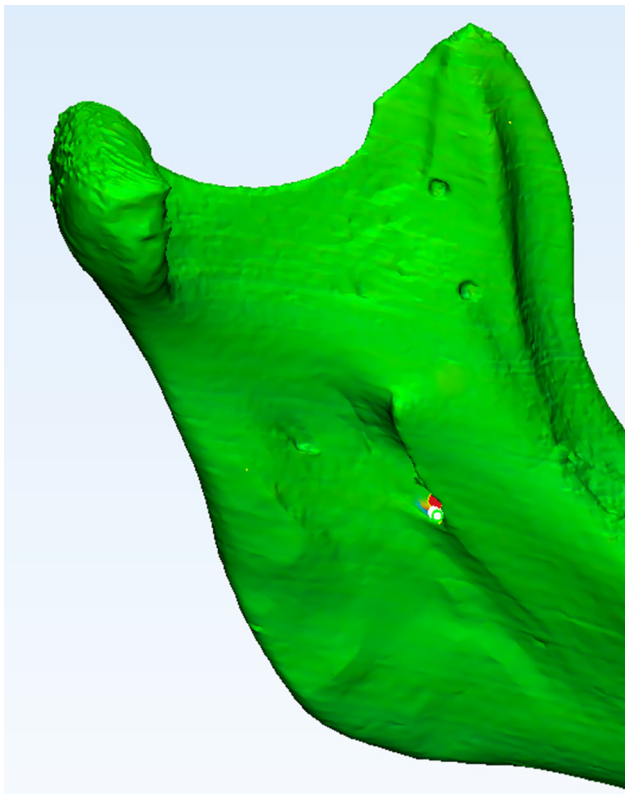

Fig. 4 Location of the largest deviation measured in the study (comparison between replica 2 vs. replica 5). The value (−2.0526 mm, red, Table 1) lies in the depth of the mandibular foramen. As such reliefs cannot be captured with surface scanning, this deviation is likely to correspond to a methodological artifact.

than those mentioned, and that the replicas are even more similar than the values in Tables 1 and 2 suggest. This is supported by the fact that almost all differences lie within a narrow range (between 0.2 and −0.2 mm), and that values found outside this range were extremely rare (Fig. 2). The dif-

ferences also centred around 0 mm, further evidence of the sameness of the replicas, which otherwise would have been expected to follow a wider or shifted distribution pattern.

5 Conclusions

- Surgeons must be aware that 3D printed models, based on one and the same image file from a single patient, may differ, contrary to intuitive assumptions.
- Reversely, it is possible to manufacture qualitatively and quantitatively identical replicas with a constant production chain.
- The proof of repeatability given in this study applies only to the workflow under evaluation. The results cannot be generally applied to other production processes. Each production chain must be specifically assessed.
- For future studies focusing on reliability of preoperative models, we suggest that the following terms should be used: Qualitative accuracy for closeness in shape. Quantitative accuracy for closeness in dimension. Repeatability / reproducibility for identity of results, obtained under identical / differing conditions.

References

1. Hoang, D., Perrault, D., Stevanovic, M., et al.: Surgical applications of three-dimensional printing: a review of the current literature and how to get started. Ann. Transl. Med. **4**, 456:1–19 (2016)
2. Huotilainen, E., Jaanimets, R., Valasek, J., et al.: Inaccuracies in additive manufactured medical skull models caused by the DICOM to STL conversion process. J. Craniomaxillofac. Surg. **42**, e259–e265 (2014)
3. Fasel, J.H.D., Beinemann, J., Schaller, K., et al.: A critical inventory of preoperative skull replicas. Ann. R. Coll. Surg. Engl. **95**, 401–404 (2013)
4. Martelli, N., Serrano, C., van den Brink, H., et al.: Advantages and disadvantages of 3-dimensional printing in surgery: a systematic review. Surgery **159**, 1485–1500 (2016)
5. FCAT (Federative Committee on Anatomical Terminology): Terminologia Anatomica. Stuttgart/New York, Thieme, pp. 15–16 (1998)
6. R Core Team: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing (2017)
7. Sarkar, D.: Lattice: Multivariate Data Visualization with R. Springer, New York (2008)
8. Liang, X., Jacobs, R., Lambrichts, I., et al.: Lingual foramina on the mandibular midline revisited: a macroanatomical study. Clin. Anat. **20**, 246–251 (2007)
9. Bartlett, J.W., Frost, C.: Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet. Gynecol. **31**, 466–475 (2008)
10. Ender, A., Mehl, A.: Accuracy in dental medicine, a new way to measure trueness and precision. J. Vis. Exp. **86**, e51374:1–13 (2014)
11. Liu, S., Mi, Z., Langenburg, G.M., et al.: Accuracy and reliability of feature selection by Chinese fingerprint examiners. Forensic Sci. Res. **2**, 203–209 (2017)
12. George, E., Liacouras, P., Rybicki, F.J., et al.: Measuring and establishing the accuracy and reproducibility of 3D printed medical models. RadioGraphics **37**, 1424–1450 (2017)
13. Fasel, J.H.D., Knoepfli, A.S., San Millan, D., et al.: Quantitative evaluation of 3D printed anatomical objects: a comparison of optical surface scanning and micro-computed tomography. SM J. Biomed. Eng. **3**, 1021:1–4 (2017)