Supersampling Enables Accurate Microstructural Bone Adaptation Simulations in Human *in vivo* HR-pQCT Images

Authors: Nicholas Ohs (1), Duncan C. Tourolle né Betts (1), Penny R. Atkins (1), Stephanie Sebastian (1), Bert van Rietbergen (2), Michael Blauth (3), Patrik Christen (1,4), Ralph Müller (1)

Affiliations:

1. *Institute for Biomechanics, ETH Zurich, Zurich, Switzerland*
2. *Orthopaedic Biomechanics, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
3. *Department for Trauma Surgery, Innsbruck University Hospital, Innsbruck, Austria*
4. *Institute for Information Systems, FHNW, Brugg, Switzerland*

Abstract

In silico trials of treatments in a virtual physiological human (VPH) would revolutionize research in the biomedical field. Hallmarks of bone disease and treatments can already be simulated in pre-clinical models and in *ex vivo* data of humans using microstructural bone adaptation simulations. The increasing availability of *in vivo* high resolution peripheral quantitative computed tomography (HR-pQCT) images provides novel opportunities to validate and ultimately utilize microstructural bone adaptation simulations to improve our understanding of bone diseases and move towards *in silico* VPH decision support systems for clinicians.

In the present study, we investigated if microstructural bone adaptation simulations of *in vivo* human HR-pQCT images yielded accurate results. Since high-resolution ground truth images cannot be obtained *in vivo,* we applied an *ex vivo* approach to study resolution dependence and the effect of super-sampling on morphometric accuracy. To address simulation initialisation issues, we developed an input regularisation approach to reduce initialisation shocks observed in microstructural bone adaptation simulations and evaluated supersampling as a way to improve the accuracy of model inputs. Finally, we compared our *ex vivo* results to simulations run on *in vivo* images to investigate whether *in vivo* image artefacts further affect simulation outcomes.

Keywords

Supersampling, HR-pQCT, Simulation, Mechanoregulation, Microstructure, Bone Adaptation

38 # Introduction

39 Simulations are considered the third pillar of modern science next to models and experiments. In the
40 biomedical field the creation of a virtual physiological human (VPH) is seen as one of the most
41 important goals [1]. The vision of the VPH is to provide researchers with a model that allows rapid
42 hypothesis testing via *in silico* trials and provides doctors with a virtual patient as a decision-support
43 system for their daily work [2]. The role of bone, both structurally and physiologically, indicates that a
44 validated model for microstructural bone adaptation and (re)modelling is a significant component to
45 any VPH model. Previous studies have shown that various aspects of bone diseases and their
46 treatments can be simulated in pre-clinical models [3–6]. Importantly, these models can produce
47 results comparable to population data when *ex vivo* images are used as an input [7]. However, the
48 translation of microstructural simulations to clinical image data has largely been constrained by the
49 availability of high quality images and validation data.

50 With the introduction and increased use of HR-pQCT, large amounts of clinically relevant data have
51 been gathered which provide the basis to validate and parameterise *in silico* models of bone [8–17].
52 However, combining current microstructural bone adaptation simulations with HR-pQCT is non-
53 trivial. Existing simulations either utilize synthetic images [4] or high-resolution micro-CT images
54 which cannot be obtained clinically [3,5–7]. Furthermore, HR-pQCT images tend to have more noise
55 [18] and other potential imaging artefacts, such as those due to movement [19].

56 The reduction in resolution is a known obstacles for the translation of computational techniques
57 from the lab into the clinical setting [20–22]. Thus, the use of clinical images in microstructural bone
58 adaptation simulations requires us to first understand the convergence of existing algorithms with
59 respect to image resolution [22,23] and second, evaluate whether supersampling of HR-pQCT data to
60 the resolution of desktop micro-CT images on which the algorithms have been validated produces
61 accurate results [18,24]. Supersampling of magnetic resonance imaging (MRI) data has been shown
62 to produce micro-FE results in good agreement with those from micro-CT images of a higher

resolution [18]. While it is clear that supersampling does not yield the same effects as scanning at a higher resolution [25], as supersampling cannot compensate for information missing in the image, techniques, like mesh refinement, are widely used in numerical applications to improve simulation accuracy by providing a better digital representation of the information contained in the images.

While several models exist [4,5,26–28], we chose the advection based remodelling simulation of Adachi [28] to test the appropriateness of HR-pQCT data input as it has been used previously on pre-clinical data [3,6] and *ex-vivo* large data [7]. In comparison to Ruimerman et al. [4], the model simulation is also deterministic which simplifies the comparison of results.

In previous studies using the algorithm by Adachi et al. [28], initial iterations, which showed aberrant results, were regarded as part of the model initialization and excluded from analysis [3,7]. However, the exclusion of the initial iterations leads to a divergence between the clinical *in vivo* and the *in silico* baseline models, which precludes direct comparison. The aberrant results of the initial iterations are caused by an initialisation shock, which is common when modelling coupled systems, like that of advection and finite-element methods by Adachi [28] or in coupled ocean-climate models, and are related to mismatches between experimental input data and simulation parameters [29,30]. In the context of microstructural bone adaptation simulations, the results of these mismatches can be observed in the large sudden changes in parameters, such as the total bone volume or the overall structural stiffness. Reducing this shock behaviour allows for the inclusion of all simulation iterations, such that identical baseline models can be used for the clinical and *in silico* models and results can be directly compared.

The second initialisation issue stems from the fact that a single threshold cannot yield both correct morphometric indices and mechanical properties on HR-pQCT images [31]. However, *in silico* microstructural bone adaptations rely on having correct digital representations of morphometrics *and* mechanics as the simulation couples these two properties.

87 The goal of the present study is to determine if microstructural simulations of *in vivo* human HR-
88 pQCT images yield accurate results, such that they could be used as part of a VPH. Since we cannot
89 obtain high-resolution ground truth images *in vivo,* we applied an *ex vivo* approach to study the
90 resolution dependence and the effects of super-sampling on morphometric accuracy. However, we
91 first had to address the two initialisation issues: initialisation shocks and disagreement of mechanics
92 and morphometrics in model inputs. Thus, we developed an input regularisation approach to reduce
93 initialisation shocks observed in microstructural bone adaptation simulations and studied
94 supersampling as a method to improve accuracy of model inputs with respect to both mechanics and
95 morphometry. Finally, we compared our *in silico* results to simulations run on *in vivo* images to
96 investigate whether additional *in vivo* image artefacts affect simulation outcomes.

97 Materials

98 High Resolution *Ex Vivo* Micro-CT Images

99 Five distal radii were obtained from female cadavers at the Amsterdam Medical Center as part of a
100 previous study [32]. The donors' ages varied between 58 and 95 years and the bone volume fraction
101 (BV/TV) of the samples varied from 7 to 20% where BV/TV was inversely related with age. The
102 medical history of the cadaveric specimens was unknown. High resolution CT images were obtained
103 at an isotropic voxel-size of 25 µm with a vivaCT 80 (70 kV, 114 µA, 300 ms integration time), a micro-
104 CT device by Scanco Medical AG (Switzerland). Images were Gauss-filtered (sigma = 1.2, support = 1)
105 and the trabecular region was hand-contoured by a trained operator for each scan using the
106 software of the scanner manufacturer.

107 *In Vivo* HR-pQCT Images

108 Five patients (four female, one male) were recruited at Innsbruck Medical University as part of a
109 radius fracture study. Patients provided informed consent and participated in a study approved by
110 the ethics committee of the Medical University of Innsbruck. The age of the patients ranged from 26
111 to 80 and their BV/TV from 21 to 7%. In this study, images of the unfractured, contralateral radius

112 were used. Scans were performed with an isotropic voxel-size of 61 µm using an XtremeCT II (68 kV,
113 1470 µA, 43 ms integration time), a clinical HR-pQCT device by Scanco Medical AG (Switzerland).
114 Images were Gauss-filtered (sigma = 1.2, support = 1) and the trabecular region was hand-contoured
115 by a trained operator using the software of the scanner manufacturer.

116 Generation of Low-Resolution Images

117 The high-resolution *ex vivo* micro-CT grey-scale images were downscaled to resolutions of 40, 61, and
118 82 µm. These three resolutions will be referred to as low-resolutions in this paper. Currently, 61 and
119 82 µm are the highest resolutions available for clinical CT scanners (Xtreme CT I and II, Scanco
120 Medical AG, Switzerland), these resolutions will be referred to as the clinically relevant resolutions.
121 Resizing was performed using the scikit-image [33] rescale function in Python [34] with third-order
122 interpolation and anti-aliasing enabled. The binary hand-contoured masks for the high-resolution
123 images were converted to a floating-point data-type and resized like the micro-CT images. Finally, a
124 threshold of 50% of the maximum image value was applied to obtain binary masks for the low-
125 resolution images.

126 Generation of Supersampled Images

127 Supersampled images were created from the low-resolution images by applying the scikit-image
128 resize Python function, again with third-order interpolation and anti-aliasing enabled, to a resolution
129 of 25 µm. When creating the supersampled images, the conversion of image dimensions between
130 the different resolutions was not unique (i.e. due to rounding of the integer image dimensions after
131 scaling with a floating point number, differences in size of one voxel could occur). To ensure that the
132 supersampled images have the exact same dimensions as the original images, the scikit image resize
133 function was used. The resize function is identical to the rescale function with the exception that it
134 resizes images to a target image dimension instead of resizing using a target scaling factor. For the
135 supersampled images, the same hand-contoured masks for the trabecular region were used as for
136 the high-resolution *ex vivo* micro-CT grey-scale images to avoid influences of differences in masks on

our results. Herein, the images supersampled from 40, 61, and 82 µm to 25 µm are referred to as s40 µm, s61 µm, and s82 µm, respectively.

The 61 µm clinical *in vivo* HR-pQCT images were re-sampled using the scikit-image rescale function, as was used for the *ex vivo* micro-CT images, to generate datasets at resolutions of 25, 40, and 82 µm. The masks of the HR-pQCT images were similarly rescaled.

Methods

Remodelling Simulations

Micro-FE Analysis

For the micro-FE analysis, we used the parallel octree solver parOsol [35] on the supercomputer Piz Daint at the Swiss National Supercomputing Centre (CSCS, Lugano, Switzerland). Output parameters of strain energy density (SED) and the apparent compressive stiffness along the longitudinal axis were evaluated. Boundary conditions were determined using a load estimation algorithm developed by Christen et al. [36]. This algorithm tries to linearly combine three different load cases to achieve the most homogeneous SED distribution possible across the given bone structure. The target mean SED value was 0.02 MPa, as has been used previously [22]. Furthermore, soft pads were added to the distal and proximal ends of the images with a pad-thickness of 246 µm and a Young's modulus of 15 MPa, which has previously been found to improve the load estimation [36]. For all experiments, we computed the load-estimation using the high-resolution files and applied the same loading conditions to the low-resolution and supersampled files. This method of load-estimation removes the voxel-size dependency of the algorithm as a confounding factor.

The micro-FE simulations for images with resolutions higher than 50 µm were run on a 50 µm hexahedral mesh since the mechanical signalling implemented in the microstructural bone adaptation simulation is roughly equivalent to a blurring with a sigma of 100 µm. Therefore, the additional resolution in the SED would not yield differing results. The use of the 50 µm hexahedral

mesh also reduced the computational resources required to run simulations (i.e. for the 25 µm images, this reduction was an order of magnitude).

Remodelling Algorithm

The strain-adaptive *in silico* microstructural bone adaptation simulation by Adachi et al. [28] was re-implemented in Python using NumPy [37] and pybind11 [38]. In short, this algorithm is iterative; for each step, SED (a result of the micro-FE simulation) is translated, via a mechanostat, into the velocity field of an adapted advection equation. Within this advection equation, the mass transfer is constrained within a proximity of the bone surfaces, which results in changes to the bone microstructure. Due to the implementation, all changes are limited to the trabecular region of the simulated structure.

Binary Model Generation from Micro-CT Data

The high-resolution images were segmented using a threshold of 450 mg HA/cm^3 [22] and the bone volume over total volume (BV/TV) for the trabecular regions was computed for reference. Finally, voxels identified as bone were set to 750 mg HA / cm^3 and background voxels were set to 0 mg HA / cm^3.

Regularized Model Generation

To ensure that the remodelling simulation operates only on the bone surface, the input to the algorithm is required to be binary except for surface voxels that can be represented with intermediate values. To compare the effects of using a conventional binary input or using an input allowing partially filled voxels at the surface layer, we implemented a regularization method that preserves information of the grey-scale image at the surface of bone structures (Figure 1, left). First, a regularization threshold was applied to each high-resolution grey-scale image. Then, surface voxels (empty voxels in direct face-to-face contact with full voxels) of the intermediate binary structure were identified using a Von Neumann neighbourhood. For each surface voxel, grey-scale values from the original grey scale image were converted to a value in the range of zero to one relative to the

regularization threshold. The regularization threshold was chosen such that the grey-scale BV/TV of the resulting structure was identical to the one computed for the respective conventional binary structure. Finally, the entire structure was multiplied by the same density value as the conventional binary input (750 mg HA / cm^3).

Parameters for Microstructural Bone Adaptation Simulation

Simulation parameters were chosen such that both formation and resorption were observed in the simulations and were used for all samples and resolutions. For this reason, a narrow lazy zone (0.0196 MPa to 0.0204 MPa) was chosen. The maximum velocity of the mechanostat was set to an arbitrary value of 12 µm/month. The slopes of the mechanostat were set to 8000 µm/year/MPa. The chosen value for slope resulted in generally high velocities and greater changes per time unit, due to the very narrow linear regime of the mechanostat. The choice of simulation parameters allowed for large differences between the different resolutions.

To ensure that the choice of time step between consecutive micro-FE calls did not alter the results, a time step of approximately 1.9 months was chosen. The simulated time period was set to 5 years, resulting in a shorter final iteration step.

Study Design

The validity of using *in vivo* HR-pQCT data as an input for advection based microstructural bone adaptation simulations was investigated using four virtual experiments. Experiment A addresses the issue of initialisation shocks and compares the current approach of generating simulation input models with a novel regularization method. The regularized approach retains grey-scale information, allowing the simulation to initialise with a structure closer to the original one. The goal was to compare the behaviour of the two approaches during the initial iteration steps to identify the approach that exhibits the least amount of initialization shock. Experiment B compared the mechanical and morphological properties of regularized input models generated from high-resolution images, which had been downsampled or down- and then supersampled, that were bone

211 volume fraction matched to the original high-resolution image. The aim of this experiment was to
212 quantify differences between regularized input models with respect to mechanics and
213 morphometrics that may confound simulations. In experiment C, microstructural bone adaptation
214 simulations were run on models of all three *ex vivo* image sets (high-resolution, downsampled, and
215 down- plus supersampled) to assess if observed differences in the simulations were due to a lack of
216 fidelity in the input data or the numerical grid. Finally, in experiment D, the *in vivo* images were
217 rescaled to the same resolutions used in experiment C and the convergence was quantified with
218 respect to resolution. The results were compared to those from experiment C to assess what effect
219 differences in image quality and factors other than resolution have on the outcome of the simulation.

220 Experiment A: Effect of Regularized Input Models on Initialisation Shock

221 Two simulation input model generation approaches were compared (Figure 1, right), the current
222 state of the art through binary representation and a regularized model with partially filled voxels at
223 the surface. Both models were generated from the high resolution *ex vivo* micro-CT image data set.
224 The partially filled voxels approximate a bone surface with sub-voxel precision. Microstructural bone
225 adaptation simulations were run and the discontinuity in BV/TV and compressive stiffness for the
226 initial simulation steps were quantified for both methods.

227 Experiment B: Effect of Supersampling on Mechanical and Morphometric Accuracy

228 For experiments B, regularized input models were produced for all three micro-CT image sets: high-
229 resolution, low-resolution, and supersampled low-resolution images. In the following, these are
230 called reference, low-resolution, and supersampled regularized input models. For all images, the
231 reference BV/TV was always the one obtained from the respective binary high-resolution image.

232 To quantify the agreement in mechanical properties between reference, low-resolution, and
233 supersampled regularized input models of the same bone structure, micro-FE analyses were
234 performed on the data set at each of these resolutions, respectively (Figure 2). SED distributions,
235 mean SED, mean static parameters (BV/TV, trabecular number (Tb.N), trabecular thickness (Tb.Th),

236 trabecular spacing (Tb.Sp), and structural model index (SMI)), and standard deviations for all mean
237 values were computed for comparison between the three regularized input model types.
238 Additionally, the Kolmogorov-Smirnov statistic was computed between the SED of each sample of
239 the low-resolution and supersampled regularized input models and the corresponding reference
240 regularized input model to quantify mechanical agreement. Finally, the adequacy of the chosen
241 threshold was assessed through comparison of the SED distribution from FE analysis of the low-
242 resolution regularized input models for a range of thresholds (575 - 775 mg HA/cm^3). The magnitudes
243 of the peak SED were compared to the reference SED distribution.

244 Experiment C: Effect of Supersampling on Morphometric Accuracy throughout a 245 Microstructural Bone Adaptation Simulation

246 Microstructural bone adaptation simulations were run on the input models generated for
247 experiment B. The simulations run on the reference regularized input models were considered the
248 best approximations of the *in vivo* remodelling process and were used as reference to quantify
249 errors. These simulations are referred to as the reference simulations (Figure 3). Static parameters
250 (BV/TV, Tb.N, Tb.Th, Tb.Sp, and SMI) and formed and resorbed volume over time were computed.

251 Experiment D: Effect of *In Vivo* Image Artefacts on Convergence of Supersampled HR-pQCT 252 Simulation

253 Microstructural bone adaptation simulations were run on regularized input models of the *in vivo* HR-
254 pQCT images and their rescaled version (Figure 3) using the same simulation parameters and
255 calculating the same static parameters as in experiment C. The reference BV/TV for the regularized
256 input model generation for all resolutions was based on the respective original HR-pQCT resolution
257 binary structure.

258 Evaluation and Statistics

259 All simulations and evaluations were performed within the trabecular mask. SED was evaluated for
260 non-empty voxels. SED distributions were represented using the SciPy Gaussian kernel density

261 approximation [39]. NumPy [37] was used to compute the Kolmogorov-Smirnov (KS) statistic, mean
262 and standard deviations, as well as BV/TV, which was computed through integration of the model
263 density within the trabecular mask. All other static parameters (Tb.N, Tb.Th, Tb.Sp, SMI) were
264 computed using the scanner manufacturer's image processing language (IPL) [40]. Before calling the
265 IPL functions, models were up-scaled to 25 µm to remove the voxel-size dependency of IPL functions
266 as a confounding factor. Finally, formed and resorbed volume over time was computed by
267 integration of positive and negative density changes using NumPy.

268 Comparisons for experiment A were done using a paired Student t-test. For experiments B and C, to
269 determine significance, two-way analysis of variance (two-way ANOVA) was performed for each
270 measured parameter as an omnibus test with the two categorical groups: resolution and
271 supersampling. If heteroscedasticity was detected using a Levene test, heteroscedasticity consistent
272 covariance matrices of type HC3 were used. Post-hoc group comparisons were done using paired
273 Student t-tests and p-values were corrected using the Holm-Bonferroni correction for multiple
274 comparisons. For experiment D, paired Student t-tests were performed and p-values were corrected
275 using the Holm-Bonferroni correction for multiple comparisons. The level of significance was set to
276 0.05. For the Student t-tests and Levene tests, scipy 1.3.1 was used. The ANOVA was done using
277 statsmodels [41] 0.10.2.

278 Results

279 Experiment A: Effect of Regularized Input Models on Initialisation Shock

280 For the conventional binary input models, the change in apparent compressive stiffness after the first
281 iteration was a factor of 5.9±0.8 larger than the maximum of all other iteration steps (Figure 4). For
282 BV/TV, a factor of 2.3±1.6 increase was observed in the first iteration (Figure 4). Visually, we observe
283 that for some samples a clear shock in BV/TV was present. In contrast, for the first iteration of the
284 regularized input model approach, the change in apparent compressive stiffness is indistinguishable
285 from the rest of the simulation with a computed factor of 0.5±0.5 increase, which is significantly

286 lower than for the threshold method ($p<0.001$). For BV/TV, we obtained a factor of 2.1±0.6 increase
287 (Figure 4) which is not significantly different from the threshold method.

288 Since the regularized input model approach removed the re-initialization shock in apparent stiffness
289 across all samples, we performed all other experiments using this approach.

290 Experiment B: Effect of Supersampling on Mechanical and Morphometric Accuracy

291 For all tested thresholds, the SED distributions for the low-resolution regularized input models did
292 not visually match the SED distribution of the reference regularized input models (Figure 5).
293 Qualitatively, the difference in SED distribution increased with voxel-size. For a BV/TV matched
294 threshold, the peaks of the distributions aligned visually. The agreement with the reference SED
295 distributions for supersampled regularized input models was almost an order of magnitude better
296 than for the low-resolution regularized input models ($p<0.05$ (40 µm), $p<0.01$ (61 µm and 82 µm);
297 Table 1) with deviations in mean SED of less than 5% and the KS statistic being below 0.03 (Table 1).

298 Resolution, applied supersampling, and the interaction between resolution and supersampling had a
299 significant effect on all measured static parameters, except for the interaction of the effects for SMI
300 and the effect of supersampling on Tb.Th (Table 1).

301 Without supersampling, significant deviations in the mechanical and static parameters were
302 observed for all lower resolutions (40, 61, and 82 µm) (Table 1). The only exception being SMI which
303 showed no significant differences. No differences in BV/TV were observed, as BV/TV was matched.
304 With supersampling, deviations in the static parameters were significantly lower for each of the
305 lower resolutions (s40, s61, and s82 µm, respectively) (Table 1).

306 The average regularization threshold for the high-resolution images was 563.6±6.6 mg HA / cm^3. The
307 average regularization threshold for the supersampled images were 561.7±6.8, 530.9±9.6, and
308 496.6±13.7 mg HA / cm^3 for the three supersampled resolutions (s40, s61, and s82 µm), respectively.
309 For the images without supersampling, the average regularization thresholds were 601.8±10.0,
310 605.2±14.7, and 590.6±18.9 mg HA / cm^3 for 40, 61, and 82 µm, respectively.

Therefore, the regularization thresholds obtained for the supersampled regularized input models were roughly linearly related to resolution. This linear relationship did not hold true for the regularization thresholds for the low-resolution regularized input models.

In summary, without supersampling we observed deviations of more than 10% (up to 27%) for all static parameters, except Tb.Th, for the clinically relevant voxel-sizes for the regularized input models. In contrast, using super-sampling, deviations were less than 8% for all static parameters for the highest available clinical voxel-size (61 µm). Agreement with the reference SED improved an order of magnitude when using super-sampling.

Experiment C: Effect of Supersampling on Morphometric Accuracy Throughout a Microstructural Bone Adaptation Simulation

For the reference simulations, BV/TV was initially reduced by 1.8 to 14.8%, followed by an increase in BV/TV of 3.8 to 38.8% (Figure 7). Tb.N decreased by 5.2 to 22.8%. Trabecular thickness increased by 29.1 to 42.4%, except for one sample for which it decreased by 8.8%. SMI changed by -15.0 to 18.5%; two samples experienced an increase and three samples a decrease in SMI. Overall the range in SMI across all samples was reduced over the course of the simulation by 34.7%.

Resolution, supersampling, and their interaction had a significant effect ($p<0.001$) on all measured static and dynamic parameters for the maximum deviations observed during the simulation.

Low-Resolution Simulations

Differences in bone-structure between the reference simulation and the low-resolution simulation were visible and increased over the course of the simulation (Figure 6). Deviations in static parameters significantly increased over the course of the simulation (Table 1). BV/TV, and Tb.N were underestimated for all low-resolution simulations compared to the reference simulations ($p<0.01$). SMI and Tb.Sp were both overestimated ($p<0.05$), while Tb.Th did not follow a clear trend (Figure 8).

Bone formation and resorption rates were significantly different between reference and low-resolution simulations (Table 1). Visually, the formation rate for the low-resolution simulations peaked at a later time point and had lower peak values compared to the reference simulation (Figure 9). For the resorption rate, peak delay and widening was also observed for the low-resolution simulations (Figure 9). However, the magnitude of the resorption rate peaks increased with voxel-size and the initial resorption rate decayed slower compared to formation rates, for which the opposite was observed.

In summary, deviations in static parameters doubled over the course of the simulation for most parameters. For the dynamic parameters, we noticed a delay from the start of the simulation to when peak formation and resorption occur. Additionally, a lower change in formation and resorption rates was observed for all low-resolution simulations compared to the reference simulations.

Supersampled Simulations

Visually, differences in bone structure over the course of the simulation compared to the reference simulations were drastically reduced for the supersampled simulations (Figure 6). Accuracy in BV/TV was improved by an order of magnitude for the highest available clinical resolution (s61 µm) compared to the low-resolution simulations. The accuracy of all other static parameters was also significantly improved (Table 1, Figure 8). The maximum deviations in static parameters were not significantly greater than those of the input models for all supersampled resolutions.

For the dynamic parameters, accuracy of bone formation and resorption per time unit was significantly increased for the supersampled images (Table 1). The formation and resorption rates peaked at the same time-point across all resolutions, within the temporal resolution of the simulation. The peaks of both rates were of similar magnitude across all resolutions. Visually, the overall shapes of the formation and resorption curves were similar across all resolutions for the duration of the simulation period, with no noticeable widening or shift of peaks (Figure 9).

In summary, the deviations in static parameters did not increase beyond the initial standard deviation throughout the duration of the simulation for the highest clinical resolution when using supersampling. The behaviour of the dynamic parameters was very similar across all resolutions and matched the behaviour of the reference simulation.

Experiment D: Effect of *In Vivo* Image Artefacts on Convergence of Supersampled HR-pQCT Simulations

After an initial drop in BV/TV of 7.2 to 24.2%, the simulations on the *in vivo* HR-pQCT data supersampled to 25µm showed varying behaviour. Three samples showed an increase in BV/TV, one sample had a close to stable BV/TV over time and one sample experienced a further reduction in BV/TV before BV/TV began increasing after half of the simulation time. Tb.N. decreased for all samples over time by 12.0 to 42.9%, whereas Tb.Th. increased over time by 13.8 to 65.1%. Tb.Sp also increased by 13.6 to 78%. SMI decreased for all samples from 7.8 to 21.8%. The spread of SMI values across all samples decreased over the course of the simulation by 47.8%.

Comparing the convergence of the different static parameters with respect to resolution between the low-resolution simulations from experiment C and the simulations from experiment D, no significant differences could be found except for Tb.Th at 61 and 82 µm resolutions ($p<0.001$) (Figure 10). Hence, the effects of noise and other additional imaging artefacts from *in vivo* HR-pQCT were smaller than the effects of model resolution, which dominated the convergence errors observed in the static parameters (Figure 10).

Discussion

The objectives of this study were to investigate whether microstructural simulations of *in vivo* human HR-pQCT images yielded accurate results and were a viable tool as part of a VPH.

Experiment A: Effect of Regularized Input Models on Initialisation Shock

Since one of the ideas of the VPH is to provide doctors with a decision support system [2], the goal of every microstructural bone adaptation simulation must be to achieve parity between the simulated structure and the structure observed *in vivo.* We observed that when using conventional model inputs, the apparent compressive stiffness showed an initialization shock behaviour (Figure 4), with a change approximately six times larger than any other change in stiffness over the course of the simulation. While initialisation shocks have not been studied in the context of microstructural bone adaptation simulations, Mulholland et al. [30] has defined a cause of initialisation shocks in the context of ocean-climate models that can be related to microstructural bone adaptation algorithms. They describe how the removal of certain model components can result in abrupt changes in the dynamics of the system. The analogy for microstructural bone adaptation simulations are the mismatch of applied boundary conditions and the true, but unknown, *in vivo* boundary conditions. This mismatch can also be interpreted as the removal of certain boundary condition forces at the beginning of the *in silico* adaptation. We tackled this challenge by employing the load estimation algorithm by Christen et al. [36] which tries to estimate the *in vivo* applied loads more closely than the uniaxial compression boundary conditions typically used with HR-pQCT radius data [42]. Another potential cause for initialisation shocks could be the abrupt change of surface geometry from *in vivo* microstructural bone adaptation to *in silico* bone adaptation simulations. Using a regularized input, we used information in the grey-scale image that is normally cut off, improving the input model generation to reduce the initialisation shock and the associated effect on the results. For all simulations, the developed regularized input model approach removed the shock behaviour in apparent compressive stiffness (Figure 4). While the magnitude of the change in BV/TV for the initial iteration step did not change with the new regularized input model, change in BV/TV looks smoother using this approach (Figure 4), and larger changes in BV/TV are expected for this type of simulation, as over the course of the simulation the structure adapts more and more to the applied boundary conditions, yielding less changes as the structure reaches a local minimum. We conclude that the first

406 iteration step does not have to be excluded if this new approach is used, allowing direct comparisons
407 to *in vivo* measurements. Furthermore, the fact that the initialisation shock was removed might
408 indicate that the regularized input model is a better representation mechanically of the *in vivo* bone
409 structure than the conventional binary version.

410 Experiment B: Effect of Supersampling on Mechanical and Morphometric Accuracy

411 Another obstacle to overcome when running microstructural bone adaptation simulations on *in vivo*
412 HR-pQCT images was finding an accurate digital representation of a bone captured *in vivo* with HR-
413 pQCT with respect to mechanics and morphometry [31], which is an obvious requirement of bone
414 adaptation simulations. We found that with the use of supersampling, the choice of a single
415 threshold provided regularized input models that agreed well for BV/TV, mechanical properties, and
416 other tested morphometric parameters (Tb.N, Tb.Sp, SMI). Furthermore, the thresholds for the
417 different supersampled resolutions followed roughly a linear trend. Therefore, even in the absence of
418 a high resolution ground truth, an appropriate threshold for accurate morphometrics and mechanics
419 can be chosen based on this linear relationship. Our results for images that were not supersampled
420 agreed with previous research [31], which showed no agreement between mechanics and
421 morphometrics for various thresholds (Figure 5). This also holds true for thresholds optimized to
422 match BV/TV (Figure 5) which has been used in a previous study by Christen et al. in which they
423 investigated the voxel size dependence of a micro-FE based load estimation algorithm [22].
424 Therefore, supersampling might also prove to be useful in other studies using images with HR-pQCT
425 resolution.

426 Experiment C: Effect of Supersampling on Morphometric Accuracy Throughout a
427 Microstructural Bone Adaptation Simulation

428 To study the accuracy of microstructural bone adaptation simulations for images with HR-pQCT
429 resolutions, we used high-resolution micro-CT images as ground truth, as realistic bone structures
430 have already been simulated using such images [7]. The use of different low-resolution voxel-sizes

431 (40, 61, and 82 µm), resulted in deviations in morphometric parameters of more than 30% in
432 comparison to the reference simulations, even for the highest clinically available resolution (Table 1).
433 This is relevant, because deviations of even 15 %, e.g. in BV/TV, would indicate different diseases
434 with opposing effects on BV/TV [43]. Importantly, differences in parameters were not pure smooth
435 offsets which could be corrected via calibration curves as is possible with other computational
436 techniques [23]. In comparison to the static parameters, the deviations were greater for the dynamic
437 parameters. A previous study, by Schulte et al. which evaluated capabilities of bone adaptation
438 simulations to predict dynamic rates on pre-clinical models, similarly found that dynamic parameters
439 were more challenging to capture than static morphometric parameters [44]. Our results confirmed
440 that microstructural bone adaptation simulations run on native clinical scanner resolutions suffer
441 from poor accuracy in morphometric parameters, limiting future use as a model for human bone
442 adaptation.

443 Running the same simulations on the supersampled images (s40, s61, s82 µm) resulted in a drastic
444 reduction in static parameter deviations to less than 10%. For BV/TV, these deviations were near 1%
445 (Table 1), which is similar to the reproducibility limit of BV/TV for the clinical setting (0.84-1.14%)
446 [45,46]. These deviations may have been much smaller than for the simulations without
447 supersampling since the regularized input model generated from the supersampled images captured
448 the reference model more closely, than the low-resolution regularized input model (Table 1). Any
449 deviations of the regularized input model lead to error accumulation throughout the simulation,
450 which ultimately lead to larger morphometric deviations of the final structure (Figure 8). Only the
451 dynamic parameters obtained from the supersampled simulations closely followed the temporal
452 profile of the high-resolution simulations capturing peak positions, width, height, and overall curve
453 profile, indicating that there may have also been an intrinsic voxel-size dependence of the algorithm
454 independent of the initial model.

455 Supersampling improved the accuracy of microstructural bone adaptation simulations run on images
456 with clinical resolutions. This improved accuracy allowed for accurate predictions of static and
457 dynamic parameters relative to the reference simulations run on micro-CT images.

458 Experiment D: Effect of *In Vivo* Image Artefacts on Convergence of Supersampled HR-
459 pQCT Simulations

460 Finally, we investigated the effects of clinically observed image artefacts, e.g. higher noise levels, by
461 using *in vivo* HR-pQCT images. Comparing the simulations run on these images to those run on
462 supersampled versions of the same images, no significant difference was observed in the
463 convergence of the different static parameters (Figure 10), except for Tb.Th. Since the *ex vivo* and *in*
464 *vivo* dataset are not identical, it is possible that this difference in Tb.Th is due to unknown
465 physiological differences between the subject groups. Furthermore, with respect to the overall
466 deviation observed in Tb.Th, the observed significant difference is still small. On top of that, Tb.Th is
467 known to be difficult to capture with HR-pQCT resolution [21,47]. Our findings support this as
468 supersampling does not significantly improve the accuracy of Tb.Th for the regularized input models
469 we generated, indicating that the image resolution of HR-pQCT images is not sufficient to contain
470 enough information to extract Tb.Th as accurate as other morphometric parameters. Hence, the
471 observed significant difference could be due to the limitations of the method to extract trabecular
472 thickness. Irrespective of the cause, the small deviations in Tb.Th are likely not clinically relevant, due
473 to the aforementioned limited accuracy with which Tb.Th can be measured with HR-pQCT.

474 In Conclusion, image resolution dominates the accuracy of morphological simulation outcomes of
475 microstructural bone adaptation simulations with no other effects having a significant influence.
476 Since we have demonstrated that the effects of voxel-size can be drastically reduced with
477 supersampling in experiment C, we conclude that within the variation of our data, supersampling on
478 HR-pQCT data improves accuracy comparably to supersampled down-scaled micro-CT data.

Limitations

This study is, however, not without limitations. One limitation of this study was the lack of a high resolution ground truth scan of the patient radii. However, images of cadaveric specimens do not capture the artefacts associated with *in vivo* HR-pQCT images, such as motion artefact. Further, micro-CT images cannot be obtained from patients due to the radiation dosage. Thus, we utilized both high-resolution cadaveric images and clinically acquired *in vivo* HR-pQCT images of patients to assess these factors independently.

An additional limitation of this study is the sample size (n=5 for both *ex vivo* and *in vivo* experiments). However, the small spread in deviations across subjects observed from the results of the supersampled simulations indicates that a larger sample size may not be warranted. Importantly, the inclusion of additional samples would have required an excess of computational resources due to the high resolution of the simulations.

Conclusions

In conclusion, we found model resolution to be the dominating image property which drove convergence errors in microstructural bone adaptation simulations. Importantly, supersampling drastically reduced this dependency, resulting in simulation outcomes that, even for clinically available resolutions, were similar to those from high-resolution images. Initialisation errors were avoided with the use of supersampling and the proposed regularization method, which generated model input that closely represented the true bone structure with respect to both mechanics and morphometry. With these results, we conclude that microstructural bone adaptation simulations can be run on *in vivo* HR-pQCT images and yield realistic results, given a validated set of parameters. These simulations provide a powerful tool to study disease related bone microstructure changes in patients as part of the VPH vision.

Acknowledgements

Funding for the DACH Fx Project from the Swiss National Science Foundation (Lead Agency, 320030L_170205), German Research Foundation (IG 18/19-1, SI 2196/2-1), and Austrian Science Fund (I 3258-B27) is gratefully acknowledged. This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s841.

References

1. Fenner JW *et al.* 2008 The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **366**, 2979–2999. (doi:10.1098/rsta.2008.0089)

2. Viceconti M, Hunter P. 2016 The Virtual Physiological Human: Ten Years After. *Annu. Rev. Biomed. Eng.* **18**, 103–123. (doi:10.1146/annurev-bioeng-110915-114742)

3. Schulte FA, Zwahlen A, Lambers FM, Kuhn G, Ruffoni D, Betts D, Webster DJ, Müller R. 2013 Strain-adaptive in silico modeling of bone adaptation - A computer simulation validated by in vivo micro-computed tomography data. *Bone* **52**, 485–492. (doi:10.1016/j.bone.2012.09.008)

4. Ruimerman R, Hilbers P, Van Rietbergen B, Huiskes R. 2005 A theoretical framework for strain-related trabecular bone maintenance and adaptation. *J. Biomech.* **38**, 931–941. (doi:10.1016/j.jbiomech.2004.03.037)

5. Müller R. 2005 Long-term prediction of three-dimensional bone architecture in simulations of pre-, peri- and post-menopausal microstructural bone remodeling. *Osteoporos. Int.* **16**, 25–35. (doi:10.1007/s00198-004-1701-7)

6. Levchuk A, Zwahlen A, Weigt C, Lambers FM, Badilatti SD, Schulte FA, Kuhn G, Müller R. 2014 The Clinical Biomechanics Award 2012 - Presented by the European Society of Biomechanics: Large scale simulations of trabecular bone adaptation to loading and treatment. *Clin. Biomech.* **29**, 355–362. (doi:10.1016/j.clinbiomech.2013.12.019)

526 7. Badilatti SD, Christen P, Parkinson I, Müller R. 2016 Load-adaptive bone remodeling
527 simulations reveal osteoporotic microstructural and mechanical changes in whole human
528 vertebrae. *J. Biomech.* **49**, 3770–3779. (doi:10.1016/j.jbiomech.2016.10.002)

529 8. Kirmani S *et al.* 2009 Bone structure at the distal radius during adolescent growth. *J. Bone*
530 *Miner. Res.* **24**, 1033–1042. (doi:10.1359/jbmr.081255)

531 9. Nicks KM, Amin S, Atkinson EJ, Riggs BL, Melton LJ, Khosla S. 2012 Relationship of age to bone
532 microstructure independent of areal bone mineral density. *J. Bone Miner. Res.* **27**, 637–644.
533 (doi:10.1002/jbmr.1468)

534 10. Nishiyama KK, Macdonald HM, Hanley DA, Boyd SK. 2013 Women with previous fragility
535 fractures can be classified based on bone microarchitecture and finite element analysis
536 measured with HR-pQCT. *Osteoporos. Int.* **24**, 1733–1740. (doi:10.1007/s00198-012-2160-1)

537 11. Nishiyama KK, Shane E. 2013 Clinical imaging of bone microarchitecture with HR-pQCT. *Curr.*
538 *Osteoporos. Rep.* **11**, 147–155. (doi:10.1007/s11914-013-0142-7)

539 12. Zhu TY *et al.* 2014 Alterations of bone density, microstructure, and strength of the distal
540 radius in male patients with rheumatoid arthritis: A case-control study with HR-pQCT. *J. Bone*
541 *Miner. Res.* **29**, 2118–2129. (doi:10.1002/jbmr.2221)

542 13. Yu WS, Chan KY, Yu FWP, Ng BKW, Lee KM, Qin L, Lam TP, Cheng JCY. 2014 Bone structural
543 and mechanical indices in Adolescent Idiopathic Scoliosis evaluated by high-resolution
544 peripheral quantitative computed tomography (HR-pQCT). *Bone* **61**, 109–115.
545 (doi:10.1016/j.bone.2013.12.033)

546 14. Boutroy S, Bouxsein ML, Munoz F, Delmas PD. 2005 In vivo assessment of trabecular bone
547 microarchitecture by high-resolution peripheral quantitative computed tomography. *J. Clin.*
548 *Endocrinol. Metab.* **90**, 6508–6515. (doi:10.1210/jc.2005-1258)

549 15. Boutroy S, Van Rietbergen B, Sornay-Rendu E, Munoz F, Bouxsein ML, Delmas PD. 2008 Finite

550 element analysis based on in vivo HR-pQCT images of the distal radius is associated with wrist
551 fracture in postmenopausal women. *J. Bone Miner. Res.* **23**, 392–399.
552 (doi:10.1359/jbmr.071108)

553 16. Sornay-Rendu E, Boutroy S, Munoz F, Delmas PD. 2007 Alterations of cortical and trabecular
554 architecture are associated with fractures in postmenopausal women, partially independent
555 of decreased BMD measured by DXA: The OFELY study. *J. Bone Miner. Res.* **22**, 425–433.
556 (doi:10.1359/jbmr.061206)

557 17. Sornay-Rendu E, Boutroy S, Duboeuf F, Chapurlat RD. 2017 Bone Microarchitecture Assessed
558 by HR-pQCT as Predictor of Fracture Risk in Postmenopausal Women: The OFELY Study. *J.*
559 *Bone Miner. Res.* **32**, 1243–1251. (doi:10.1002/jbmr.3105)

560 18. Rajapakse CS, Magland J, Zhang H, Liu XS, Wehrli SL, Guo XE, Wehrli FW. 2009 Implications of
561 noise and resolution on mechanical properties of trabecular bone estimated by image-based
562 finite-element analysis. *J. Orthop. Res.* **27**, 1263–1271. (doi:10.1002/jor.20877)

563 19. Pialat JB, Burghardt AJ, Sode M, Link TM, Majumdar S. 2012 Visual grading of motion induced
564 image degradation in high resolution peripheral computed tomography: Impact of image
565 quality on measures of bone density and micro-architecture. *Bone* **50**, 111–118.
566 (doi:10.1016/j.bone.2011.10.003)

567 20. Tjong W, Kazakia GJ, Burghardt AJ, Majumdar S. 2012 The effect of voxel size on high-
568 resolution peripheral computed tomography measurements of trabecular and cortical bone
569 microstructure. *Med. Phys.* **39**, 1893–1903. (doi:10.1118/1.3689813)

570 21. Manske SL, Zhu Y, Sandino C, Boyd SK. 2015 Human trabecular bone microarchitecture can be
571 assessed independently of density with second generation HR-pQCT. *Bone* **79**, 213–221.
572 (doi:10.1016/j.bone.2015.06.006)

573 22. Christen P *et al.* 2016 Voxel size dependency, reproducibility and sensitivity of an in vivo bone

loading estimation algorithm. *J. R. Soc. Interface* **13**, 20150991. (doi:10.1098/rsif.2015.0991)

23. Muller R, Koller B, Hildebrand T, Laib A, Gianolini S, Ruegsegger P. 1996 Resolution dependency of microstructural properties of cancellous bone based on three-dimensional mu-tomography. *Technol. Health Care* **4**, 113–119.

24. Schulte FA, Christen P, Badilatti SD, Parkinson I, Khosla S, Goldhahn J, Müller R. 2019 Virtual supersampling as post-processing step preserves the trabecular bone morphometry in human peripheral quantitative computed tomography scans. *PLoS One* **14**, 1–11. (doi:10.1371/journal.pone.0212280)

25. Cooper D, Turinsky A, Sensen C, Hallgrimsson B. 2007 Effect of voxel size on 3D micro-CT analysis of cortical bone porosity. *Calcif. Tissue Int.* **80**, 211–219. (doi:10.1007/s00223-005-0274-6)

26. Beaupre GS, Orr TE. 1990 An approach for time-dependent bone modeling and remodeling—application: A preliminary remodeling simulation - Beaupré - 2005 - Journal of Orthopaedic Research - Wiley Online Library. *J. Orthop. …* , 662–670. (doi:10.1167/iovs.13-11708)

27. Huiskes R, Rulmerman R, Van Lenthe GH, Janssen JD. 2000 Effects of mechanical forces on maintenance and adaptation of form in trabecular bone. *Nature* **405**, 704–706. (doi:10.1038/35015116)

28. Adachi T, Tsubota KI, Tomita Y, Scott JH. 2001 Trabecular surface remodeling simulation for cancellous bone using microstructural voxel finite element models. *J. Biomech. Eng.* **123**, 403–409. (doi:10.1115/1.1392315)

29. Balmaseda M, Anderson D. 2009 Impact of initialization strategies and observations on seasonal forecast skill. *Geophys. Res. Lett.* **36**, 1–5. (doi:10.1029/2008GL035561)

30. Mulholland DP, Laloyaux P, Haines K, Balmaseda MA. 2015 Origin and impact of initialization shocks in coupled atmosphere-ocean forecasts. *Mon. Weather Rev.* **143**, 4631–4644.

(doi:10.1175/MWR-D-15-0076.1)

31. Alsayednoor J, Metcalf L, Rochester J, Dall'Ara E, McCloskey E, Lacroix D. 2018 Comparison of HR-pQCT- and microCT-based finite element models for the estimation of the mechanical properties of the calcaneus trabecular bone. *Biomech. Model. Mechanobiol.* **17**, 1715–1730. (doi:10.1007/s10237-018-1051-6)

32. De Jong JJA, Arts JJ, Meyer U, Willems PC, Geusens PP, Van Den Bergh JPW, Van Rietbergen B. 2016 Effect of a cast on short-term reproducibility and bone parameters obtained from HR-pQCT measurements at the distal end of the radius. *J. Bone Jt. Surg. - Am. Vol.* **98**, 356–362. (doi:10.2106/JBJS.O.00127)

33. Van Der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. 2014 Scikit-image: Image processing in python. *PeerJ* **2014**, 1–18. (doi:10.7717/peerj.453)

34. Python Software Foundation. 2020 Python Language Reference, Version 3.7. See https://www.python.org.

35. Flaig C. 2012 A highly scalable memory efficient multigrid solver for µ-finite element analyses. ETH Zurich. (doi:10.3929/ethz-a-007613965)

36. Christen P, Ito K, Knippels I, Müller R, van Lenthe GH, van Rietbergen B. 2013 Subject-specific bone loading estimation in the human distal radius. *J. Biomech.* **46**, 759–766. (doi:10.1016/j.jbiomech.2012.11.016)

37. Van Der Walt S, Colbert SC, Varoquaux G. 2011 The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30. (doi:10.1109/MCSE.2011.37)

38. Jakob W, Rhinelander J, Moldovan D. 2017 pybind11 - Seamless operability between C++11 and Python.

39. Virtanen P *et al.* 2019 SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python. ,

622 1–22.

623 40. Hildebrand T, Laib A, Müller R, Dequeker J, Rüegsegger P. 1999 Direct Three-Dimensional
624 Morphometric Analysis of Human Cancellous Bone: Microstructural Data from Spine, Femur,
625 Iliac Crest, and Calcaneus. *J. Bone Miner. Res.* **14**, 1167–1174.
626 (doi:10.1359/jbmr.1999.14.7.1167)

627 41. Seabold S, Perktold J. 2010 statsmodels: Econometric and statistical modeling with python. In
628 *9th Python in Science Conference*,

629 42. Burghardt AJ, Link TM, Majumdar S. 2011 High-resolution computed tomography for clinical
630 imaging of bone microarchitecture. *Clin. Orthop. Relat. Res.* **469**, 2179–2193.
631 (doi:10.1007/s11999-010-1766-x)

632 43. Zhang ZM, Li ZC, Jiang LS, Jiang SD, Dai LY. 2010 Micro-CT and mechanical evaluation of
633 subchondral trabecular bone structure between postmenopausal women with osteoarthritis
634 and osteoporosis. *Osteoporos. Int.* **21**, 1383–1390. (doi:10.1007/s00198-009-1071-2)

635 44. Schulte FA, Zwahlen A, Lambers FM, Kuhn G, Ruffoni D, Betts D, Webster DJ, Müller R. 2013
636 Strain-adaptive in silico modeling of bone adaptation - A computer simulation validated by in
637 vivo micro-computed tomography data. *Bone* **52**, 485–492. (doi:10.1016/j.bone.2012.09.008)

638 45. MacNeil JA, Boyd SK. 2008 Improved reproducibility of high-resolution peripheral quantitative
639 computed tomography for measurement of bone quality. *Med. Eng. Phys.* **30**, 792–799.
640 (doi:10.1016/j.medengphy.2007.11.003)

641 46. Mueller TL, Stauber M, Kohler T, Eckstein F, Müller R, van Lenthe GH. 2009 Non-invasive bone
642 competence analysis by high-resolution pQCT: An in vitro reproducibility study on structural
643 and mechanical properties at the human radius. *Bone* **44**, 364–371.
644 (doi:10.1016/j.bone.2008.10.045)

645 47. MacNeil JA, Boyd SK. 2007 Accuracy of high-resolution peripheral quantitative computed

646 tomography for measurement of bone quality. *Med. Eng. Phys.* **29**, 1096–1105.

647 (doi:10.1016/j.medengphy.2006.11.002)

648

Tables

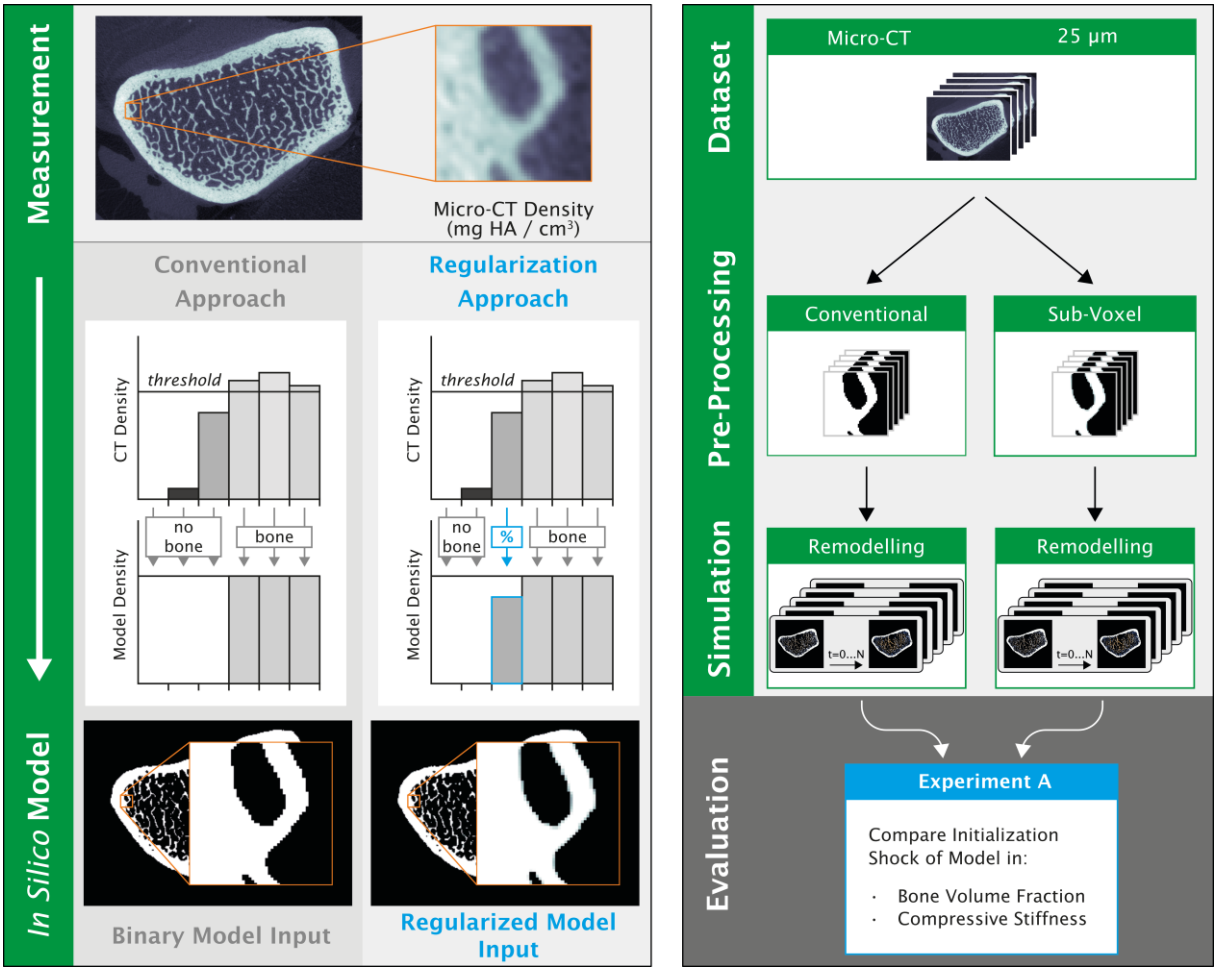
Table 1

			Downsampled			Added Supersampling		
			40 µm	61 µm	82 µm	s40 µm	s61 µm	s82 µm
Regularized Input Model Deviation	Mechanics	Mean SED	4.1 ± 1.3% *	17.2 ± 2.0% ***, ###	42.2 ± 7.1% **, ††	0.4 ± 0.6% ‡	2.4 ± 1.6% ‡‡	5.1 ± 3.4% ‡‡
		KS	0.022 ± 0.007 *	0.083 ± 0.023 *, #	0.257 ± 0.083 *, †	0.003 ± 0.004 ‡	0.012 ± 0.009 ‡	0.026 ± 0.021 ‡
	Static Parameters	Tb.N	-7.9 ± 2.7% **	-19.8 ± 4.5% **, ##	-30.1 ± 5.5% **, †††	-0.5 ± 0.1% **, ‡‡	-6.6 ± 1.3% **, ‡‡, ##	-13.4 ± 2.7% **, ‡‡, ††
		Tb.Sp	9.8 ± 2.6% **	26.8 ± 6.0% **, ##	46.5 ± 9.8% **, ††	0.5 ± 0.1% **, ‡‡	7.1 ± 1.4% **, ‡‡, ##	15.9 ± 3.4% **, ‡‡, ††
		Tb.Th	-3.0 ± 1.7% **	6.5 ± 3.8% **, ##	16.8 ± 5.8% **, ††	0.5 ± 0.1%	5.8 ± 1.2% ##	12.4 ± 2.7% *, ††
		SMI	5.5 ± 4.5%	13.9 ± 11.4%	20.5 ± 17.0%	0.03 ± 0.2%	1.8 ± 2.1%	6.6 ± 5.0%
Maximum Deviation Over the Course of the Simulation	Static Parameters	BV/TV	-6.7 ± 2.4% **, ¥¥	-21.0 ± 5.2% **, ¥¥, ##	-45.2 ± 6.2% ***, ¥¥¥, ††	-0.3 ± 0.4% ‡‡	-1.4 ± 1.7% ‡‡	-3.2 ± 3.7% ‡‡‡
		Tb.N	-11.1 ± 3.8% **, ¥	-33.8 ± 8.5% **, ¥¥, ##	-57.2 ± 7.4% ***, ¥¥¥, ††	-0.6 ± 0.2% **, ‡‡	-6.8 ± 1.8% **, ‡‡, ##	-14.9 ± 4.2% **, ‡‡‡, ††
		Tb.Sp	14.1 ± 4.2% *, ¥	57.9 ± 16.4% *, ¥, #	154.4 ± 33.3% **, ¥¥, ††	0.6 ± 0.3% *, ‡	7.5 ± 2.1% *, ‡, #	18.2 ± 5.8% *, ‡‡, †
		Tb.Th	-6.3 ± 0.8% ***, ¥	10.3 ± 4.0% *, ¥, ##	18.6 ± 6.6% *, †	0.5 ± 0.1% **, ‡‡‡	5.9 ± 1.2% **, ##	12.9 ± 3.0% **, ††
		SMI	12.6 ± 3.8% *, ¥	31.4 ± 8.9% *, ¥, #	48.4 ± 13.5% *, ¥¥, #, †	0.2 ± 0.2% ‡	2.8 ± 0.9% *, ‡, #	8.2 ± 2.3% *, ‡, †
	Dynamic Parameters	Bone Formation	70.9 ± 1.9% ***	80.3 ± 2.1% ***, ###	86.2 ± 2.7% ***, †	1.5 ± 0.7% ***, ‡‡‡	14.7 ± 2.1% ***, ‡‡‡, ###	29.0 ± 3.6% ***, ‡‡‡, †††
		Bone Resorption	73.0 ± 5.4% ***	186.7 ± 23.4% ***, ##	365.6 ± 61.3% **, ††	0.8 ± 0.9% ‡‡‡	4.7 ± 2.3% *, ‡‡‡, #	11.5 ± 7.2% ‡‡

Table 1: Comparison of microstructural bone adaptation simulation outcomes for regularized model inputs of micro-CT images that were downsampled to 40, 61, and 82 µm (left) and additionally supersampled back to 25 µm (s40, s61, and s82 µm) (right). Parameters were compared against the high-resolution micro-CT image reference simulation and relative deviations in percent are shown. Parameters were compared relative to the regularized input model before running the adaptation simulations (top) and for the maximum deviation over the duration of the simulation per parameter (bottom). We observe significantly less deviations of supersampled simulations across almost all parameters for the initial model as well as a drastically reduced increase in deviations over the course of the simulation. Strain-energy-density (SED), the Kolmogorov Smirnov statistic (KS) of the normalized SED distributions, trabecular number (Tb.N), trabecular spacing (Tb.Sp), trabecular thickness (Tb.Th), and the structure model index (SMI). Statistical significance is indicated as follows: Difference from zero (the reference simulation) (*), differences between the same resolution with and without super-sampling (‡), difference between 40 and 61 µm (#), difference between 61 and 82 µm (†), and difference between initial and maximum deviations (¥).

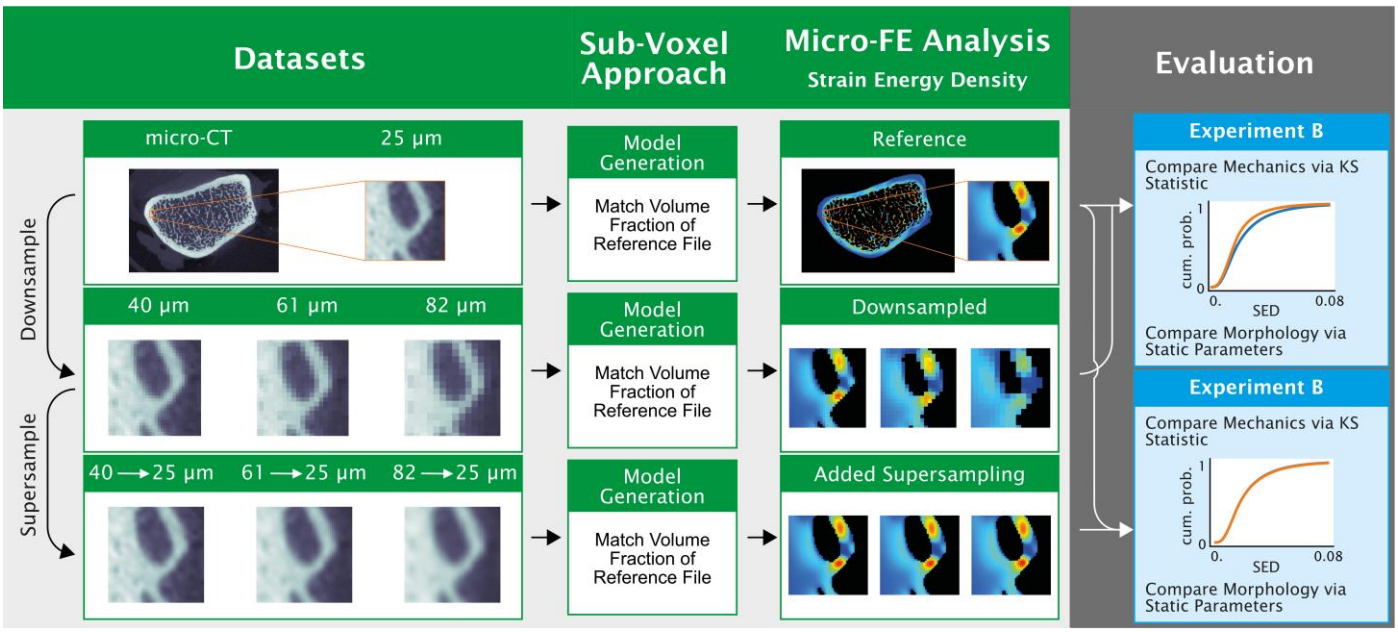
Figures

Figure 1



Experiment A evaluated the reduction of initialisation shock behaviour in bone volume fraction and compressive stiffness due to a novel input regularization approach compared to the conventional input approach. Left: Illustration of the two different threshold approaches used to convert a CT image into a valid input for the load adaptation simulation. The conventional threshold approach, and the regularization approach. Right: Overview of experiment A. Each picture per box represents a bone sample captured by micro computed tomography, and each card in the simulation boxes represent one full simulation of such a bone sample, respectively.

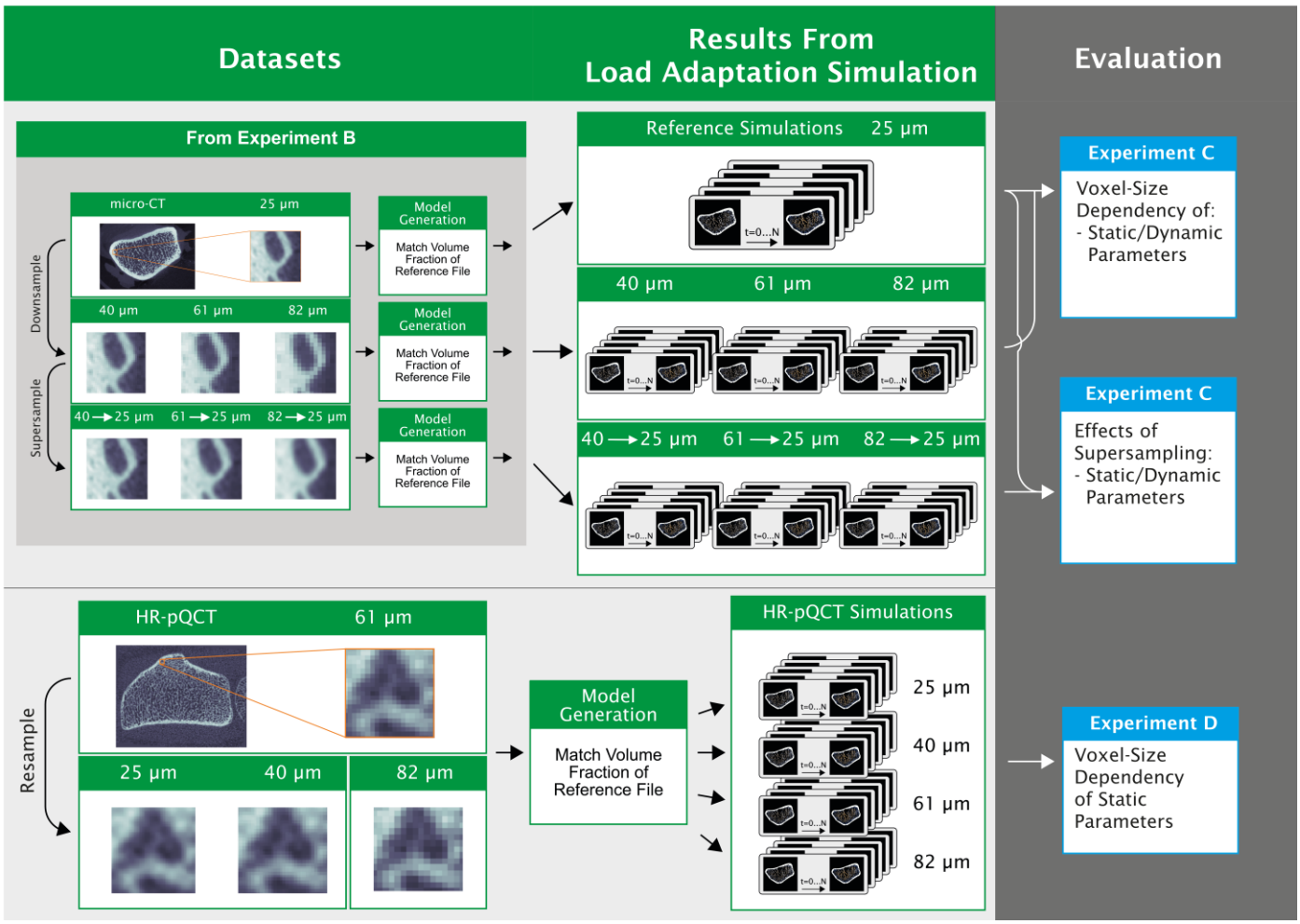
677 Figure 2



678

679 Experiment B evaluated the mechanical agreement between the different resolutions for matched
680 BV/TV. For each micro-CT image, high-, low-resolution, and supersampled images were created (left
681 column). Low-resolution: 40 µm, 61 µm, and 82 µm and supersampled images: Same downsampled
682 resolutions as the low-resolution image but supersampled back to 25 µm. Micro-finite-element
683 (micro-FE) models were generated using the regularization model generation approach of
684 experiment A (Figure 1) and matching bone volume fraction (BV/TV) for each image to the reference
685 (25 µm). A micro-FE analysis was run and strain energy density (SED) (shown in the jet colour-map)
686 and static parameters were computed for each model. SED distributions between low and
687 supersampled resolution images were compared using the Kolmogorov-Smirnov statistic to see
688 which one more closely matches the reference high-resolution regularized input model mechanically.
689 Mean static parameters were compared to see which one more closely matches the reference model
690 morphologically.

691 Figure 3

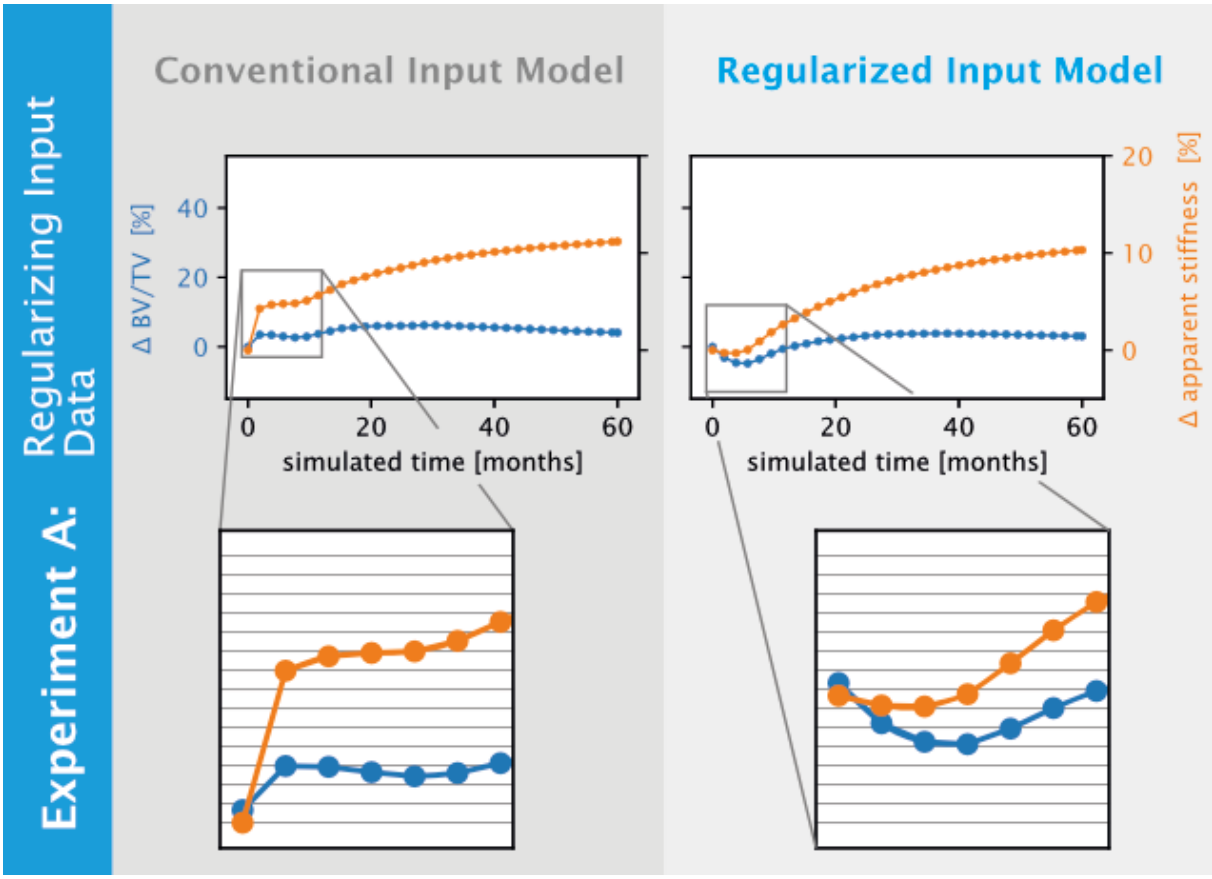


692

693 Overview of experiments C and D. Top: Regularized input models from experiment B were taken and
694 microstructural bone adaptation simulations were run for all input models. Bottom: A separate HR-
695 pQCT dataset was also converted to regularized input models and simulations were run for all input
696 models. Each simulation is represented by one card in the four boxes of the middle column. For
697 experiment C, static parameters and dynamic parameters were computed. The goal of experiment C
698 was to compare how accurate low-resolution versus supersampled resolution simulations were
699 relative to the reference simulations. For experiment D, static parameters were computed as well.
700 The goal of experiment D was to compare the voxel size dependency of down-scaled micro-CT and
701 resampled HR-pQCT images to determine if additional artefacts introduced by the HR-pQCT images
702 had a strong influence on the outcome of the bone-adaptation simulation.

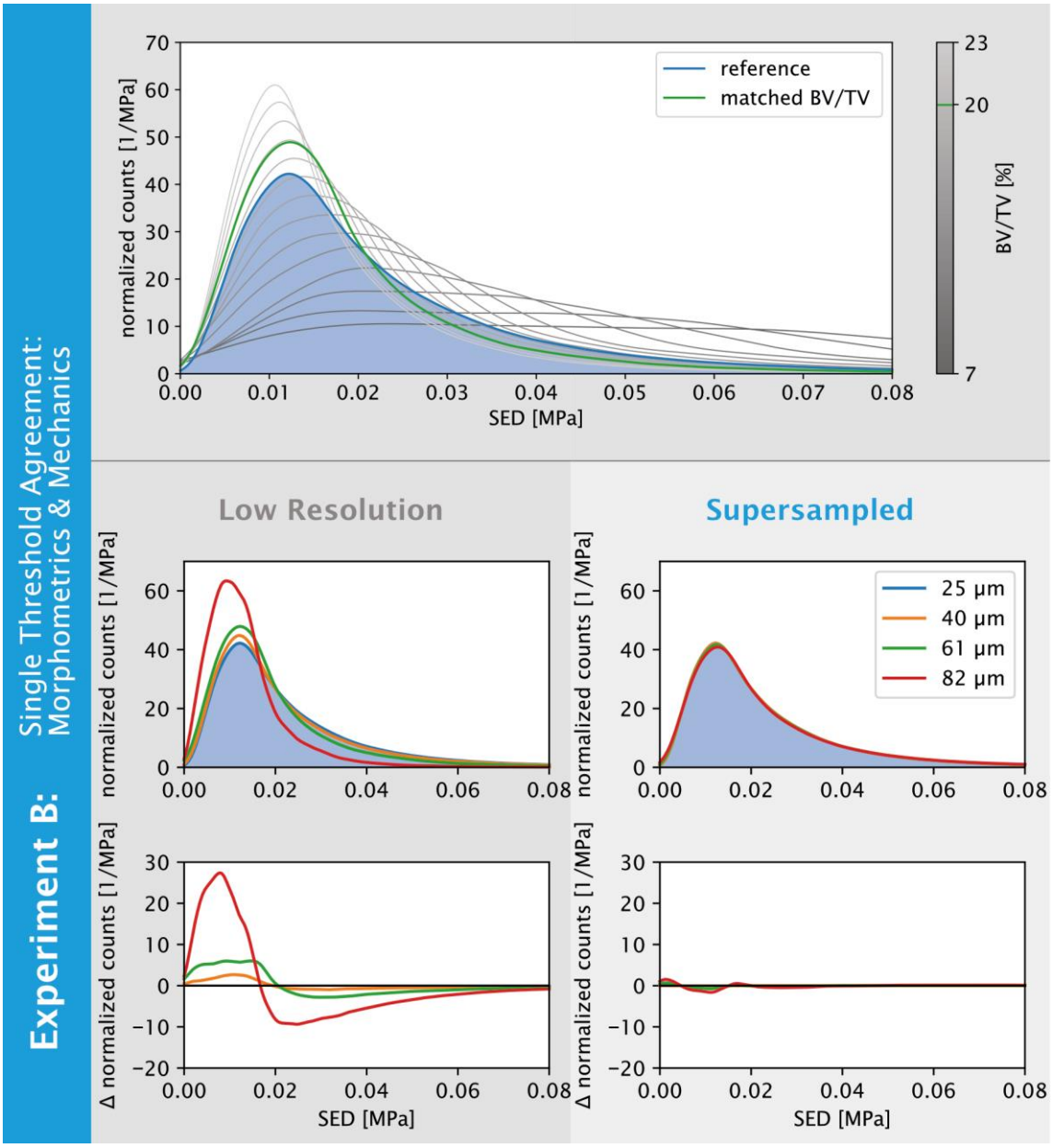
703

Figure 4



Results of experiment A for a representative sample. Comparison of relative change in bone volume fraction (BV/TV) and relative apparent stiffness. The initialization shock visible in BV/TV and stiffness (left) for the conventional model generation (Figure 1) was not present when using the regularization approach (Figure 1) (right).

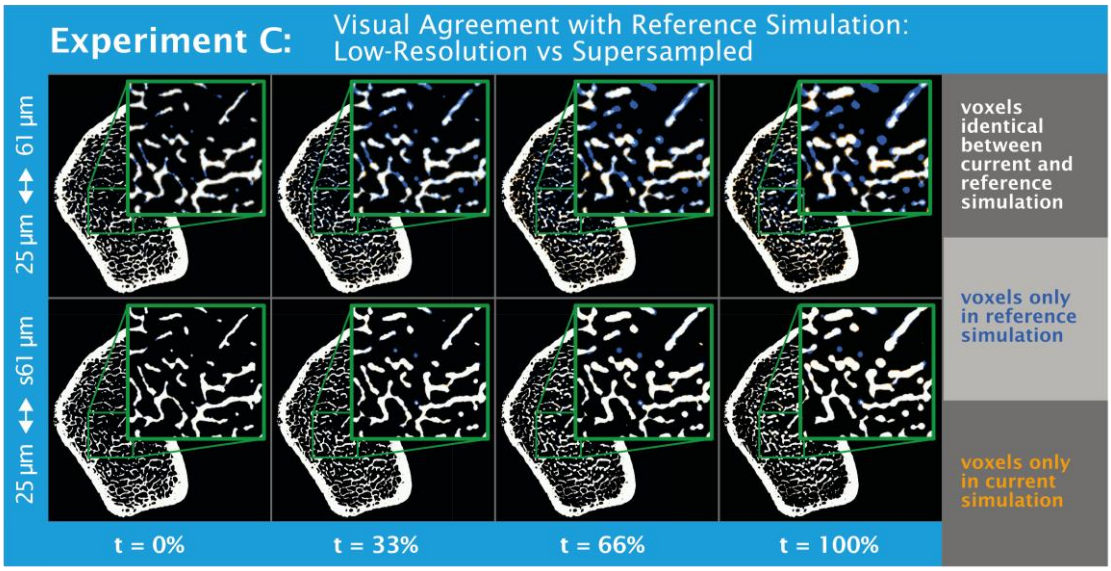
Figure 5



Results of experiment B for a representative sample. Top: The strain energy density (SED) distributions for the reference resolution (25 µm) (blue) was flatted compared to the down-scaled 61 µm model (green), which used a threshold chosen to match the bone volume fraction (BV/TV) of the reference resolution model. Thresholds corresponding to BV/TV values between 7 and 23% are shown in different shades of grey with a worse agreement with the reference SED distribution. Bottom: Comparing across all resolutions, the supersampled resolutions clearly captured the

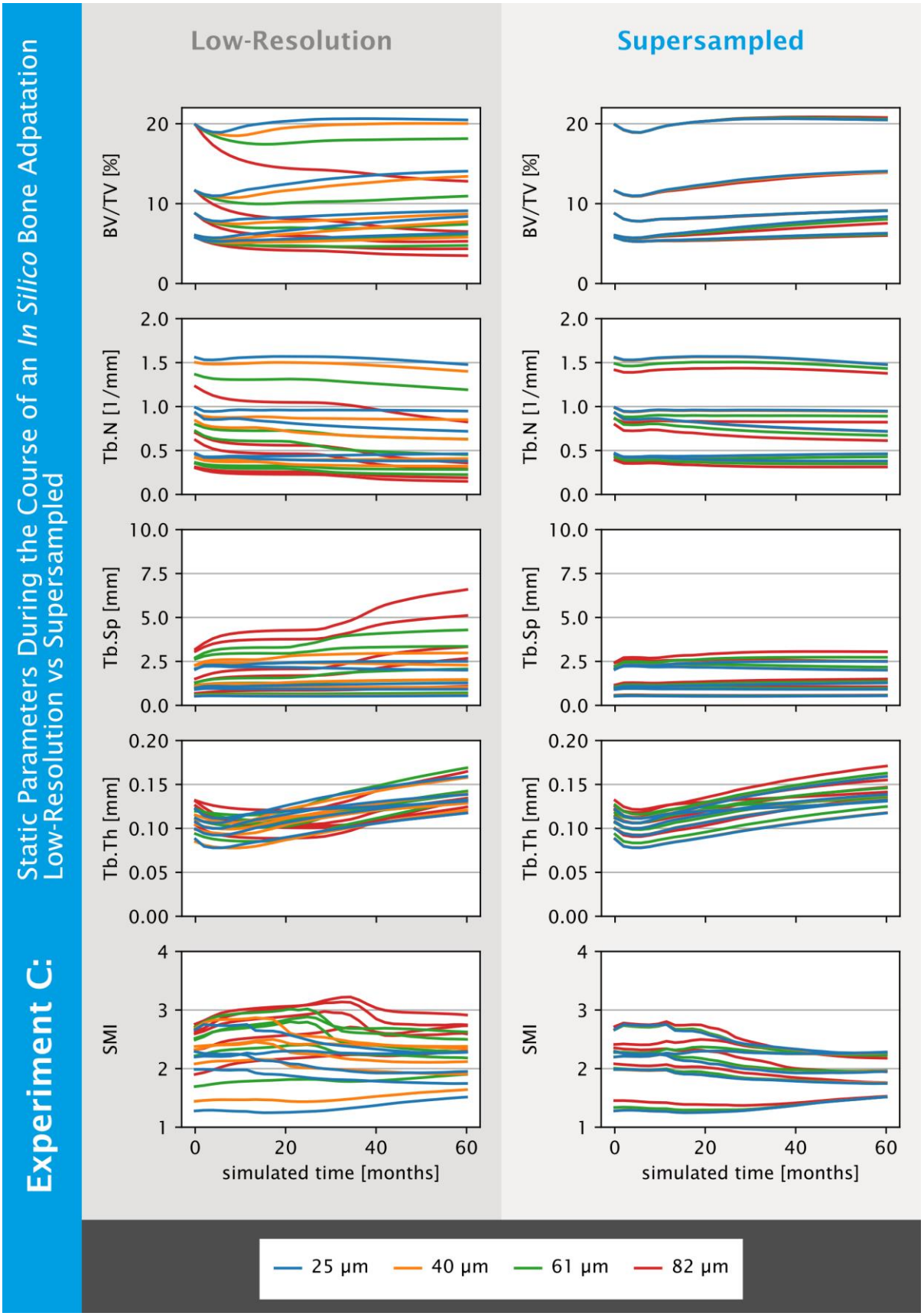
distribution of the reference model across all resolutions, with deviations being an order of magnitude smaller compared to models generated without supersampling.

Figure 6



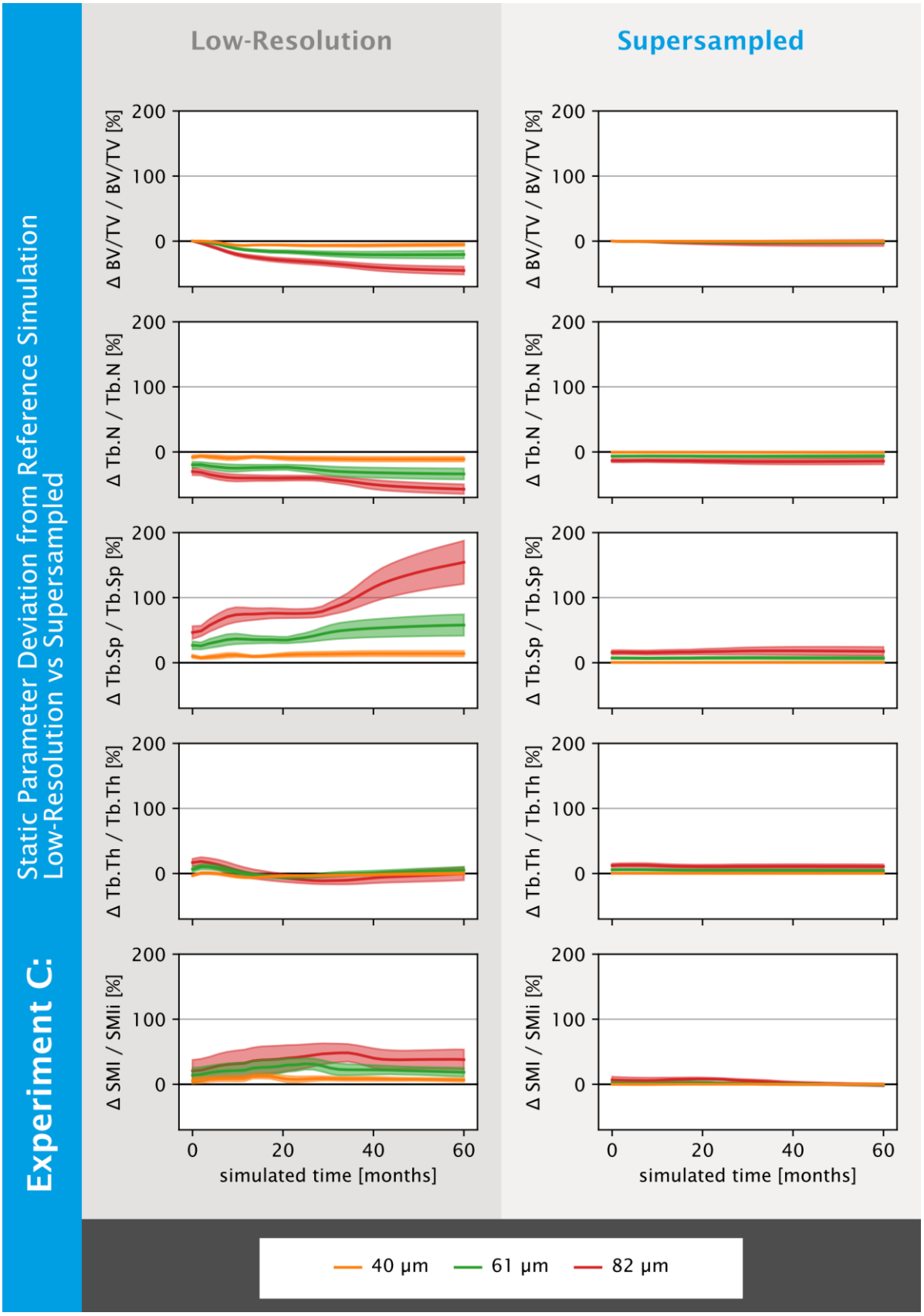
Visual difference after 0, 33, 66, and 100% of the simulated time for a representative sample of experiment C. Top: comparison of reference to low-resolution simulation. Bottom: comparison of reference to supersampled simulation. Blue: more bone in reference simulation. Orange: less bone in reference simulation. For the supersampled simulation, very small structures were still lost, due to very thin trabeculae that cannot be captured in a 61 µm image, but the major part of the bone structure remodelled identical to the reference simulation for supersampled images.

Figure 7


Experiment C:
Static Parameters During the Course of an *In Silico* Bone Adpatation
Low-Resolution vs Supersampled
Low-Resolution
Supersampled
BV/TV [%]
Tb.N [1/mm]
Tb.Sp [mm]
Tb.Th [mm]
SMI
simulated time [months]
25 µm
40 µm
61 µm
82 µm

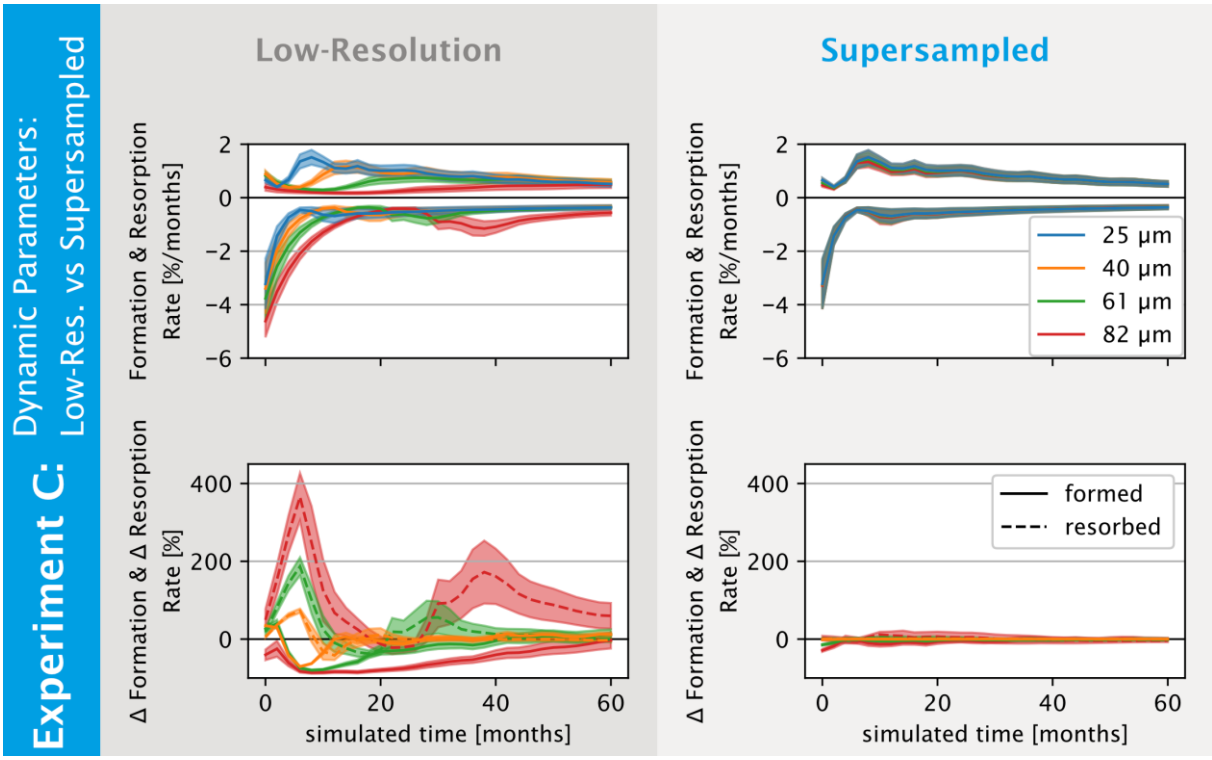
Static parameter results for experiment C. Static parameters over the course of the simulation for low-resolution (left) and supersampled regularized input models (right) are shown. Computed static parameters are: Bone volume fraction (BV/TV), trabecular number (Tb.N), trabecular spacing (Tb.Sp), trabecular thickness (Tb.Th), and structure model index (SMI). BV/TV was matched for the initial model, which is why initially we got perfect agreement between the resolutions. Overall we see for the different samples that supersampling improves the agreement with the reference (25 µm) simulations.

Figure 8


Experiment C:
Static Parameter Deviation from Reference Simulation
Low-Resolution vs Supersampled
Low-Resolution
Supersampled
Δ BV/TV / BV/TV [%]
Δ Tb.N / Tb.N [%]
Δ Tb.Sp / Tb.Sp [%]
Δ Tb.Th / Tb.Th [%]
Δ SMI / SMIi [%]
200
100
0
0
20
40
60
simulated time [months]
40 µm
61 µm
82 µm

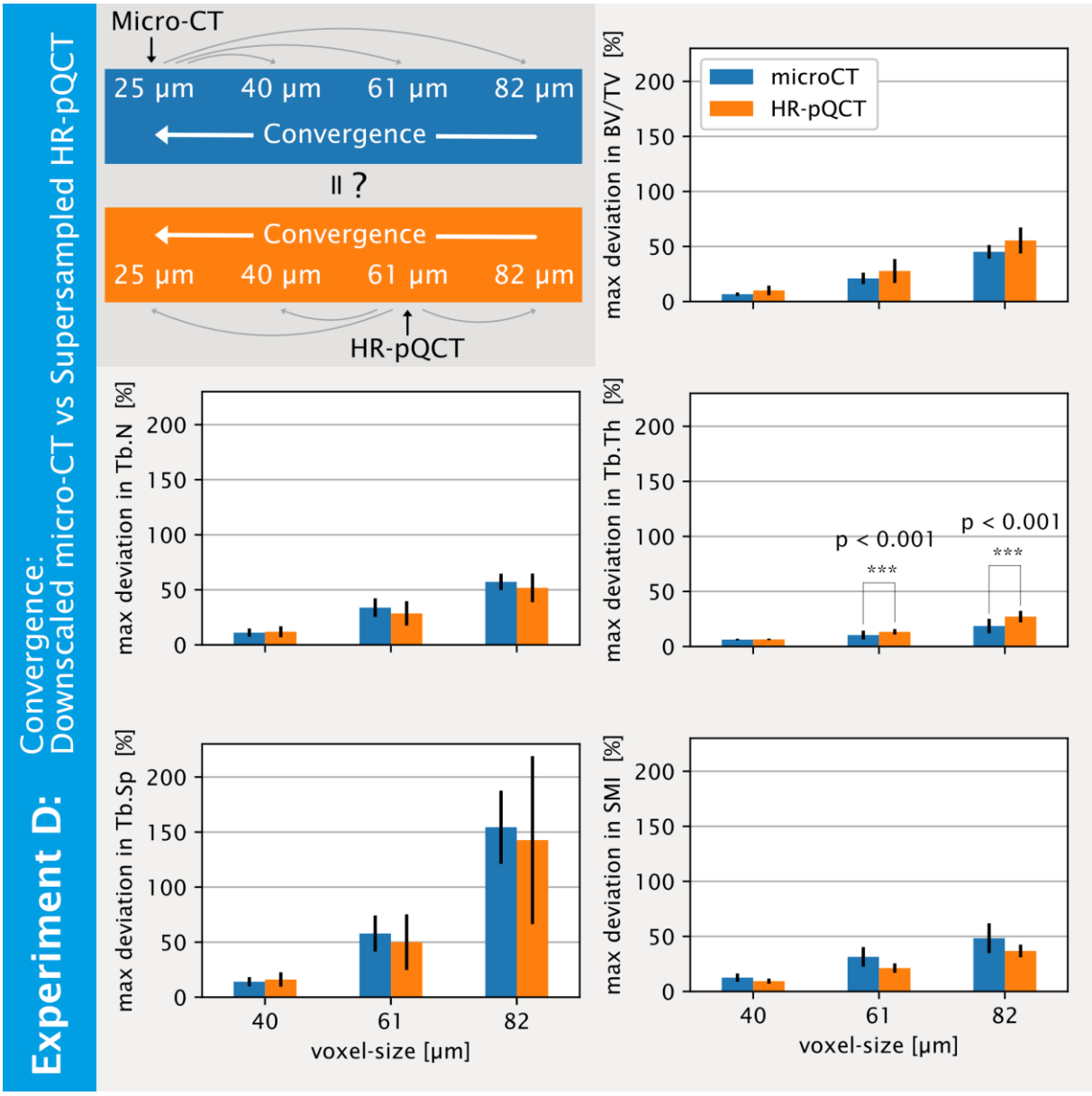
Static parameter results for experiment C. Deviations in static parameters from the reference simulation (25 µm) over the course of the simulation for low-resolution (left) and supersampled regularized input models (right) are shown. Computed static parameters are: Bone volume fraction (BV/TV), trabecular number (Tb.N), trabecular spacing (Tb.Sp), trabecular thickness (Tb.Th), and structure model index (SMI). BV/TV was matched for the initial model, resulting in perfect agreement between the resolutions. Generally, for the different samples, supersampling improves the agreement with the reference (25 µm) simulations. Deviations were more predictable after supersampling of the image.

Figure 9



Dynamic parameter results of experiment C. Dynamic parameter results of simulation run on low-resolution (left) and supersampled regularized input models (right) are shown. Bone formation and resorption over time (top), and deviations of these parameters from the reference simulation (25 µm) (bottom). Reference simulation results were aligned with the use of supersampled regularized input models with respect to the amount and time-point of formation and resorption events. Deviations were an order of magnitude smaller for the simulations run on supersampled regularized input models.

Figure 10



Experiment D. Convergence behaviour of low-resolution simulations from experiment C and from rescaled HR-pQCT image simulations. Maximum mean deviations from the reference (25 µm) simulations and corresponding standard deviations are shown for all computed static parameters. Significant differences were observed for trabecular thickness (Tb.Th); the mean maximum deviations of all other parameters are not significantly different, indicating that voxel-size was the dominating factor on the simulation outcomes.