

A computational framework for the longitudinal assessment of mechanoregulation of fracture healing

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Nicholas Ohs

M.Sc. Physics, RWTH Aachen

born on 08.01.1991

citizen of Germany

accepted on the recommendation of

Prof. Dr. Ralph Müller, examiner

Dr. Patrik Christen, co-examiner

Table of contents

Acknowledgements	i
Summary	ii
Zusammenfassung	v
Introduction	1
1.1 Thesis motivation	2
1.2 Specific aims	5
1.3 Thesis outline	5
Background	15
2.1 Validation of HR-pQCT against micro-CT for morphometric and biomechanical analyses: A review	16
Development of a computational framework for the assessment of bone fracture healing	67
3.1 A cross-platform and scalable Python framework for bone microstructure simulations and analysis	68
3.2 Towards <i>in silico</i> prognosis using big data	81
3.3 Upscaling of <i>in vivo</i> HR-pQCT images enables accurate simulations of human microstructural bone adaptation	92
3.4 Automated segmentation of fractured distal radii by 3D geodesic active contouring of <i>in vivo</i> HR-pQCT images	127
Longitudinal assessment of bone mechanoregulation of fracture healing	153
4.1 Mechanoregulation in distal radius fractures assessed <i>in vivo</i> by time-lapsed HR-pQCT images	154
Synthesis	175
Curriculum Vitae	193

Acknowledgements

The work presented in this thesis would not have been possible without the support of several individuals, whom I would like to thank.

First of all, I would like to thank Prof. Dr. Ralph Müller for giving me the opportunity to work in his group. I am particularly thankful for his commitment to provide ample opportunities for every PhD student, to participate, experience and learn from international conferences. I am also grateful that Prof. Müller gave me the freedom to pursue my own ideas, while providing guidance when necessary.

I would also like to thank Dr. Patrik Christen for agreeing to be my co-examiner. Furthermore, I am grateful for Dr. Christen's engagement in the DACH project that provided funding for and enabled my work. Moreover, I would like to thank him for choosing me to work on the DACH project.

I am grateful for all the scientific feedback, all the scientific discussions, the moral support and the fun in the office that Dr. Duncan Tourolle né Betts has provided over the entire duration of my PhD. His love for science was inspirational, as was his commitment to always perform work with all due diligence. I would like to thank Graeme Paul for countless discussions, his strategic perspective on how to be a successful PhD student and all the hours on the ski slopes that helped me to clear my mind after a stressful week. Furthermore, I would like to thank Dr. Felicitas Flohr and Dr. Ariane Scheuren, who were always helpful when questions about Biology needed answers.

Moreover, I would like to thank some fantastic students that I have worked with over the course of my PhD, in particular Francisco Correia Marques, Matthias Graß, Fabian Keller and Bryant Schroeder. Their work has helped me tremendously during my research.

I would like to thank my sister who has always been there for me in Zurich, if I needed support. Furthermore, I would also like to thank my parents who for over a quarter of a century have supported me with all their energy.

Finally, I would like to thank the Swiss National Science Foundation (320030L_170205) for providing funding and the Swiss National Supercomputing Centre for providing computational resources for all my projects.

Summary

In our ageing society, bone fractures are becoming increasingly common and have become a major socioeconomic burden. The fracture of the radius is one of the most common bone fractures and indicative of overall poor bone quality in a patient. As there is limited consensus on the optimal treatment of bone fractures among physicians, a better understanding of the bone fracture healing process is vital, to develop new treatment protocols and improve functional outcome. Bone is a remarkable material that constantly adapts and maintains itself and which can also repair itself without scar tissue. The key driver of these processes is mechanical loading of bone and the tissue surrounding it called mechanoregulation. Failure of bone tissue to react to applied loading can be indicative of disruptions of the underlying biological signalling pathways. With the introduction of high-resolution peripheral quantitative computed tomography (HR-pQCT), it is now for the first time possible, to image the bone microstructure in patients longitudinally. However, many of the technical challenges involved in quantifying bone tissue response to mechanical loading in HR-pQCT fracture images have not been solved. Furthermore, while bone adaptation simulations are currently being developed for animal studies to link observed changes in mechanoregulation to changes in biological signalling pathways, it is unclear if and how such approaches work with HR-pQCT images.

Therefore, in this thesis, we tackled these computational challenges via the following three aims: (i) develop an approach to run *in silico* bone adaptation simulations on HR-pQCT patient data, (ii) develop HR-pQCT fracture image pre-processing approaches necessary for large patient cohorts, (iii) assess mechanoregulation of fracture healing in a patient cohort. To achieve these aims, we showed how upscaling and regularisation can be used to run realistic bone adaptation simulations on HR-pQCT images. We developed an automatic contouring approach for HR-pQCT images of distal radius fractures and we compared the quantified mechanoregulation in fractured and intact distal radii that were scanned six times over the course of one year.

To address the first aim, we developed a computational framework to handle HR-pQCT and other CT data. The established method of mechanostat based bone adaptation simulation was implemented within this framework. A new regularisation approach for input data was developed which compared to traditional threshold-based methods created a more accurate

digital representation of the scanned bone structure *in silico*. Since these simulations have already been shown to yield realistic results using micro computed tomography (micro-CT) images, five micro-CT cadaver samples were used to generate reference simulations. Using downscaling, different lower resolution scanner resolutions (including HR-pQCT resolution) were obtained. Finally, simulations run on these lower resolutions, as well as on upscaled versions of these images were compared to the reference simulation. It was observed, that the new regularized inputs removed initialisations shocks that were previously visible in the global apparent stiffness of the simulated bone structure. Furthermore, we showed that upscaling increased the accuracy in morphometric parameters and mechanical properties at the same time compared to the micro-CT data, whereas without upscaling, only one of these features could be matched using a single *in silico* representation. Finally, we demonstrated that upscaling was also required to achieve accurate simulation results compared to the reference micro-CT simulations. Ultimately, our results indicate that HR-pQCT images appear to capture enough detail of the bone microstructure to enable bone adaptation simulations based on morphometry and mechanics in general.

For the second aim, we observed that while for image filtering, registration and finite element analysis, existing tools could be integrated into the computational framework; no existing tool was able to generate accurate contours of HR-pQCT images of fractured distal radii, a key step in determining volumes of interest. A new algorithm was developed that at its core uses 3 dimensional geodesic active contours (3D-GAC). Key parts of the algorithm are the reliable detection and removal of the ulna, and the identification of the optimal conversion function from a density (mg HA / cm³) image to an energy landscape for (3D-GAC). HR-pQCT images of fractured distal radii were hand contoured by trained operators to compare to the output of the algorithm. The scanner manufacturer's automatic procedure for intact radii and the newly developed algorithm were also run on HR-pQCT images of intact radii. For images of fractured distal radii, the algorithm yielded accurate contours within the inter-operator variability. For images of the contralateral side, no significant difference was found in terms of morphometric indices between the manufacturer's approach and the novel algorithm that was larger than the *in vivo* reproducibility of these parameters, suggesting that the novel approach could also be a viable alternative for the assessment of intact distal radii.

For the investigation of mechanoregulation in fracture patients, we first employed the mechanostat based bone remodelling simulation to obtain a dataset of fully mechanoregulated

bone adaptation. Using virtual experiments, in which the dataset was downsampled to various resolutions and noise levels (including those of HR-pQCT), we showed that the image quality of HR-pQCT is sufficient to utilize existing established methods of mechanoregulation quantification on HR-pQCT data. Selecting patients with only minimal amounts of movement artefacts, little fracture fragment movement, mechanoregulation was quantified for both, the fractured and the contralateral side. Mechanoregulation was found to govern the bone maintenance on the contralateral as well as the fracture repair on the fractured side. Mechanoregulation levels of the fractured side were similar to those found in animal studies. Furthermore, the late reparative phase showed similar levels of mechanoregulation to the bone maintenance observed in the contralateral side, indicating potentially shared mechanoregulation pathways between these two processes. Future studies should optimize their study protocol to achieve higher number of patients with images suitable for mechanoregulation analysis.

In summary, the developed computational framework allowed us to study fracture healing in longitudinal HR-pQCT images of fracture patients. This demonstrated, for the first time, the mechanoregulation of microstructural bone fracture healing in humans *in vivo*. Since the same methods for quantification was used as is typically used in animal studies, this opens up new possibilities to inform lab studies from knowledge gained clinically and vice versa. To efficiently quantify mechanoregulation in HR-pQCT images, we leveraged our newly developed contouring algorithm, which replaced the time-intensive task of manually contouring images of fractures, avoiding operator bias. Future studies could benefit from this, enabling larger sample numbers. Our simulation study indicated that HR-pQCT images are a viable source for *in silico* investigations using morphometry and mechanics based bone adaptation simulations. In the future, studies could employ the developed framework to address the study of bone fracture healing using clinical HR-pQCT data complemented with animal study data and investigate both data sources using *in silico* bone simulation approaches.

Zusammenfassung

In unserer alternden Gesellschaft sind Knochenbrüche ein immer häufiger auftretendes Problem, das zu einer großen sozioökonomischen Belastung geworden ist. Die Fraktur des Radius ist eine der am häufigsten auftretenden Knochenfrakturen und ein Anzeichen für eine insgesamt schlechte Knochenqualität bei Patienten. Da unter Ärzten zurzeit kein Konsens über die optimale Behandlung von Knochenbrüchen besteht, ist ein besseres Verständnis des Heilungsprozesses von Knochenbrüchen unerlässlich, um neue Behandlungsprotokolle zu entwickeln und funktionelle Ergebnisse zu verbessern. Knochen ist ein bemerkenswertes Material, das sich ständig anpasst und selbst erhält und das sich auch ohne Narbengewebe selbst reparieren kann. Die wichtigste Triebfeder dieser Prozesse ist die mechanische Belastung des Knochens und des ihn umgebenden Gewebes. Dieser Prozess wird auch als Mechanoregulation bezeichnet. Das Versagen des Knochengewebes, auf eine angewandte Belastung zu reagieren, kann ein Hinweis auf Störungen der zugrundeliegenden biologischen Signalwege sein. Mit der Einführung der hochauflösenden peripheren quantitativen Computertomographie (HR-pQCT) ist es nun zum ersten Mal möglich, die Knochenmikrostruktur bei Patienten longitudinal abzubilden. Viele der technischen Herausforderungen, die mit der Quantifizierung der Reaktion des Knochengewebes auf mechanische Belastung in HR-pQCT-Fraktur Bildern verbunden sind, sind jedoch noch immer ungelöst. Darüber hinaus werden zwar derzeit Knochenadaptationssimulationen für Tierstudien entwickelt, um beobachtete Veränderungen der Mechanoregulation mit Veränderungen der biologischen Signalwege in Verbindung zu bringen, doch ist noch unklar, ob und wie solche Ansätze mit HR-pQCT-Bildern funktionieren könnten.

Daher haben wir uns in dieser Arbeit diesen technischen Herausforderungen mit den folgenden drei Zielen gestellt: (i) Entwicklung eines Frameworks zur Durchführung von *In-silico*-Knochenadaptationssimulationen auf HR-pQCT-Patientendaten, (ii) Entwicklung von HR-pQCT-Fraktur bildvorverarbeitungsprozessen, die für große Patientenkohorten notwendig sind, (iii) Beurteilung der Mechanoregulation der Frakturheilung in einer Patientenkohorte. Um diese Ziele zu erreichen, haben wir gezeigt, wie Hochskalieren und Regularisieren genutzt werden können, um realistische Knochenadaptationssimulationen auf HR-pQCT-Bildern durchzuführen. Wir entwickelten einen automatischen Konturierungsprozess für HR-pQCT-

Bilder von distalen Radiusfrakturen und verglichen die quantifizierte Mechanoregulation in frakturierten und intakten distalen Radien, die im Laufe eines Jahres sechsmal gescannt wurden.

Um das erste Ziel zu erreichen, entwickelten wir ein Framework für die Verarbeitung von HR-pQCT- und anderen CT-Daten. Als Teil dieses Frameworks, wurde die etablierte Methode der Mechanostat-basierten Simulation der Knochenanpassung implementiert. Es wurde ein neuer Regularisierungsprozess für die Eingabedaten entwickelt, der im Vergleich zu herkömmlichen schwellenwertbasierten Methoden eine genauere digitale Darstellung der gescannten Knochenstruktur *in silico* ermöglicht. Da sich bereits gezeigt hat, dass diese Simulationen mit Hilfe von Mikro-Computertomographie (Mikro-CT)-Bildern realistische Ergebnisse liefern, wurden fünf Mikro-CT-Kadaverproben zur Erstellung von Referenzsimulationen verwendet. Durch Herunterskalieren wurden verschiedene Scannerauflösungen mit niedrigerer Auflösung (einschließlich HR-pQCT-Auflösung) generiert. Schließlich wurden die Simulationen, die mit diesen niedrigeren Auflösungen sowie mit hochskalierten Versionen dieser Bilder durchgeführt wurden, mit der Referenzsimulation verglichen. Es wurde beobachtet, dass die neuen regularisierten Eingaben Initialisierungsschocks entfernten, die zuvor in der globalen scheinbaren Steifigkeit der simulierten Knochenstruktur sichtbar waren. Darüber hinaus zeigten wir, dass das Hochskalieren im Vergleich zu den Mikro-CT-Daten gleichzeitig die Genauigkeit der morphometrischen Parameter und mechanischen Eigenschaften erhöhte, während ohne Hochskalieren nur eines dieser Merkmale mit einer einzigen *in-silico*-Darstellung erfasst werden konnte. Schließlich haben wir gezeigt, dass Hochskalieren auch erforderlich ist, um genaue Simulationsergebnisse im Vergleich zu den Referenz-Mikro-CT-Simulationen zu erzielen. Letztendlich deuten unsere Ergebnisse darauf hin, dass HR-pQCT-Bilder im Allgemeinen ausreichend Details der Knochenmikrostruktur zu erfassen scheinen, um Knochenanpassungs-Simulationen auf der Grundlage von Morphometrie und Mechanik zu ermöglichen.

Für das zweite Ziel stellten wir fest, dass für die Bildfilterung, Registrierung und Finite-Elemente-Analyse zwar vorhandene Werkzeuge in das Framework integriert werden konnten; kein vorhandenes Werkzeug jedoch in der Lage war, genaue Konturen von HR-pQCT-Bildern von frakturierten distalen Radien zu erzeugen, was ein wichtiger Schritt bei der Bestimmung der interessierenden Volumina ist. Es wurde ein neuer Algorithmus entwickelt, der auf dreidimensionale geodätisch aktive Konturen (3D-GAC) basiert. Weitere Schlüsselemente des Algorithmus sind die zuverlässige Erkennung und Entfernung der Elle und die

Identifizierung der optimalen Umwandlungsfunktion von einem Dichtebild ($\text{mg HA} / \text{cm}^3$) in eine Energielandschaft für (3D-GAC). HR-pQCT-Bilder von frakturierten distalen Radien wurden von geschulten Operatoren von Hand konturiert, um sie mit der Ausgabe des Algorithmus zu vergleichen. Das automatische Verfahren des Scanner-Herstellers für intakte Radien und der neu entwickelte Algorithmus wurden auch auf HR-pQCT-Bildern von intakten Radien angewendet. Bei Bildern von frakturierten distalen Radien lieferte der Algorithmus genaue Konturen innerhalb der Interoperator-Variabilität. Bei Bildern der kontralateralen Seite wurde kein signifikanter Unterschied in Bezug auf die morphometrischen Indizes zwischen dem Programm des Herstellers und dem neuartigen Algorithmus gefunden, der größer war als die *in vivo*-Reproduzierbarkeit dieser Parameter. Dies deutet darauf hin, dass der neuartige Algorithmus auch eine brauchbare Alternative für die Beurteilung intakter distaler Radien sein könnte.

Für die Untersuchung der Mechanoregulation bei Frakturpatienten verwendeten wir zunächst die Mechanostat-basierte Simulation der Knochenstrukturanpassung, um einen Datensatz zu erhalten, der vollständig mechanoregulierte Knochenanpassung enthält. Anhand virtueller Experimente, in denen der Datensatz auf verschiedene Auflösungen und Rauschpegel (einschliesslich derer von HR-pQCT) herunterskaliert wurde, zeigten wir, dass die Bildqualität von HR-pQCT ausreicht, um die bestehenden etablierten Methoden zur Quantifizierung der Mechanoregulation an HR-pQCT-Daten zu nutzen. Bei der Auswahl von Patienten mit nur minimalen Bewegungsartefakten und wenig Bewegung der Frakturfragmente wurde die Mechanoregulation sowohl für die frakturierte als auch für die kontralaterale Seite quantifiziert. Es zeigte sich, dass die Mechanoregulation sowohl den Knochenerhalt auf der kontralateralen als auch die Frakturheilung auf der frakturierten Seite steuert. Der Grad der Mechanoregulation auf der frakturierten Seite war ähnlich hoch wie in Tierstudien. Darüber hinaus zeigte die späte Reparationsphase ein ähnliches Niveau der Mechanoregulation wie der Knochenerhalt auf der kontralateralen Seite, was auf möglicherweise gemeinsame Mechanoregulationswege zwischen diesen beiden Prozessen hinweist. Zukünftige Studien sollten ihr Studienprotokoll optimieren, um eine höhere Anzahl von Patienten mit Bildern zu erreichen, die für die Analyse der Mechanoregulation geeignet sind.

Zusammenfassend lässt sich sagen, dass das entwickelte Framework es uns ermöglichte, die Frakturheilung in longitudinalen HR-pQCT-Bildern von Frakturpatienten zu untersuchen. Damit konnte zum ersten Mal die Mechanoregulation der mikrostrukturellen

Knochenbruchheilung beim Menschen *in vivo* nachgewiesen werden. Da die gleichen Methoden zur Quantifizierung verwendet wurden, wie sie typischerweise in Tierversuchen eingesetzt werden, eröffnet dies neue Möglichkeiten, Laborstudien basierend auf klinisch gewonnenen Erkenntnissen zu planen und umgekehrt. Um die Mechanoregulation in HR-pQCT-Bildern effizient zu quantifizieren, setzten wir unseren neu entwickelten Konturierungsalgorithmus ein, der die zeitintensive Aufgabe der manuellen Konturierung von Fraktur Bildern ohne Präzisionsverlust ersetzt, wodurch eine Verzerrung durch manuelle Arbeit vermieden wurde. Zukünftige Studien könnten von diesem Algorithmus profitieren und größere Probenzahlen ermöglichen. Unsere Simulationsstudie zeigte, dass HR-pQCT-Bilder eine brauchbare Quelle für *In-silico*-Untersuchungen mit morphometrischen und mechanikbasierten Knochenanpassungssimulationen sind. In Zukunft könnten Studien das entwickelte Framework nutzen, um die Forschung der Knochenbruchheilung unter Verwendung klinischer HR-pQCT-Daten, ergänzt durch Daten aus Tierstudien voran zu treiben. Zudem könnten dann beide Datenquellen mit *in silico*-Knochensimulationen untersucht werden.

Chapter 1

Introduction

1.1 Thesis motivation

With the massive growth of biomedical big data (Dimitrov, 2016; Murdoch and Detsky, 2013; Stratton et al., 2009), new opportunities to study and understand complex diseases and improve diagnoses arise (Kessel and Combs, 2016). Especially using medical images there is an enormous potential for data science applications (Luo et al., 2016). These applications could potentially improve our understanding and thereby the diagnosis and prognosis of complex diseases or conditions. Many diseases and complications are complex, characterised and influenced by multiply factors thus potentially benefitting from a more data-rich analysis (Ohs et al., 2016). However, making use of all the generated data presents huge challenges due to, e.g., the diversity of the data (different data types) which require different analysis approaches (Gietzelt et al., 2016) and the lack of data normalization as for example two different bone fracture patients will never have identical fracture patterns down to the micro-scale.

Having a look at biomedical data currently available, we notice that there is a broad range of different types available. Structural information can be obtained for example by high-resolution quantitative peripheral computed tomography (HR-pQCT) scanners, which even allow longitudinal data to be collected *in vivo* in patients (Boutroy et al., 2008, 2005; Kirmani et al., 2009; Nicks et al., 2012; Nishiyama et al., 2013; Nishiyama and Shane, 2013; Sornay-Rendu et al., 2017, 2007; Yu et al., 2014; Zhu et al., 2014). Blood, urine, saliva and other non-invasive tests have been a clinical standard for years now, allowing collection of information on all kinds of molecular and cellular factors. Current technological advancements pave the way to do blood tests with just a single drop of blood (Qin et al., 2015). Of course, the developments of gene sequencing technologies add to the data one can obtain from blood. Mobile monitoring is already happening with data like “kilometres walked per day“ being used to offer personal health insurance fees by some health insurance companies. It is estimated that “by 2020, 40% IoT-related (Internet of Things) technology will be health-related” (Dimitrov, 2016).

However, the current clinical reality is different, with many clinical protocols not yet making use of the available data. One such example are fracture treatment protocols, which have been developed before much of the current technology such as for example HR-pQCT has been available. The case of fracture treatment is of particular interest, as in our ageing society (United Nations, 2020), bone fracture are increasingly common and are becoming a major socioeconomic burden (Johnell and Kanis, 2006). The most common fracture type is the distal

radius fracture which is also indicative of reduce bone quality (Court-Brown and Caesar, 2006). However, despite the critical role of radius fracture in our society, there is currently a limited consensus on the optimal treatment protocol for such fractures (Ng and McQueen, 2011), warranting further research of fracture healing to improve protocols and ultimately functional outcome for patients.

We already know, that bone is a multiscale (Gardner et al., 2006; Isaksson et al., 2009; Moustafa et al., 2012; Robling and Turner, 2009; Röntgen et al., 2010; Webster et al., 2015, 2008) and hierarchical tissue that provides structural support to the human body (Rho et al., 1998). Bone has the remarkable ability to constantly adapt and renew itself in a process referred to as bone remodelling (Hill, 1998), based on a complex interaction of multiple tissues, cell types and various signalling molecules across all scales. However, within all this complexity mechanical loading has been identified as a key signal to understand structural changes in bone longitudinally, as healthy bone tries to optimize itself to be able to withstand mechanical loading (Huiskes et al., 2000). Our key access point to understanding mechanoregulation is the bone microstructure which can be observed using high resolution CT. Observing this microstructure in healthy humans using HR-pQCT, it has been found out that mechanical loading alone can already explain the majority of structural changes occurring (Christen et al., 2014). To identify changes in mechanoregulation which could indicate the presence of a disease reducing bone's ability to adapt and repair, Schulte and colleagues have established a method to assess mechanoregulation in animals (Schulte et al., 2011), with which they could already demonstrate that up to 80% of all remodelling events can be linked to mechanics in mice (Schulte et al., 2013a). This process of mechanoregulation and adaptation was also found to be the key driver of fracture healing (Augat et al., 2005, 1996; Boerckel et al., 2012; Carter, 1987; Carter et al., 1998; Claes et al., 1998; Giannoudis et al., 2007; Klein et al., 2003), posing the question how this knowledge can be translated into clinically relevant insights to inform doctors and shape future treatment protocols.

One issue with the analysis of mechanoregulation is that while it identifies differences between groups, it does not answer the question what the underlying causes are and what the consequences of these found differences are. To bridge this gap between observed differences in bone structure over time and our knowledge of bone adaptation, different *in silico* models have been proposed. The model by Schulte and colleagues (Schulte et al., 2013b) uses an abstract mechanostat concept (Frost, 1987) which only requires a few tuneable parameters,

providing a prediction of structural changes over several years. Using *ex vivo* data, realistic osteoporotic structures have already been simulated (Badilatti et al., 2016). The model by Tourolle né Betts (Tourolle né Betts, 2019) on the other hand employs a first-principles many-parameter model that, once tuned, allows to trace back observed structural changes to e.g. differences in hormonal levels. Using this model, they could already simulate realistic fracture healing in mice (Tourolle né Betts, 2019). These models could profoundly impact our way of studying bone as their computational nature allows for rapid hypothesis testing, e.g. testing the effect of increased levels of certain hormones such as oestrogen, which is known to affect bone remodelling (Wehrle et al., 2019, 2015).

With all this success in *in silico* and animal models, one should not forget that differences between e.g. mice and human (Lambers et al., 2015) always require studies using patient data to verify our hypotheses from animal studies. While it may seem surprising that there are no studies so far using human data, a closer look reveals many challenges that hinder a trivial application of previously mentioned methods in a clinical setting. For one, the difference in imaging resolution between micro computed-tomography (micro-CT) devices used for animal and *ex vivo* studies is a known stumbling block in the translation of computational methods (Christen et al., 2016; Manske et al., 2015; Tjong et al., 2012). On the other hand, the lack of homogeneity among patients increases the need for larger sample numbers, thus increasing the need for automated imaging pipelines while at the same time increasing the difficulty of developing such pipelines.

To unify the way we work with human and animal data, we propose to develop a computational framework that is designed to study fracture healing in patients and thus verify the results already obtained in animal studies. For this, we will investigate implementations to accurately run *in silico* simulations on human HR-pQCT data. Furthermore, developing the necessary image processing pipelines will then enable the investigation of fracture healing using patient HR-pQCT fracture images.

1.2 Specific aims

The goal of this thesis was to develop a computation framework to unify the longitudinal assessment of mechanoregulation during fracture healing using micro-CT and HR-pQCT. Specifically, the following three aims were defined for this framework:

Aim 1: Development of an approach to run *in silico* bone adaptation simulations on HR-pQCT patient data.

Aim 2: Development of HR-pQCT fracture image pre-processing approaches necessary for large cohort studies of fracture healing.

Aim 3: Longitudinal assessment of mechanoregulation of fracture healing in a patient cohort.

1.3 Thesis outline

The thesis consists of five chapters. The current chapter outlines the motivation and specific aims of this thesis. The following four chapters have the following content:

Chapter 2 discusses existing efforts to validate existing computational models to analyse mouse micro-CT data to work on HR-pQCT patient data. Furthermore, it discusses limitations of current validation techniques for the validation of methods such as the analysis of mechanoregulation.

Chapter 3 describes the developed computational framework. In the first part, the structure of the computational framework and techniques used to ensure scalability and cross-platform compatibility are explained. The second part shows a proof-of-concept of *in silico* simulations on HR-pQCT can include various clinical biomarkers. In a third part, it is shown that by using upscaling, the established *in silico* bone adaptation approach by Schulte and colleagues can be adapted to yield accurate results when run on HR-pQCT patient data. The final part discusses a novel approach used to contour HR-pQCT images of fractured distal radii to enable scalability of HR-pQCT fracture studies.

Chapter 4 shows the application of the developed framework from the previous chapter to verify for the first time the local microstructural mechanoregulation of bone fracture healing in patients. For this, the adapted simulation by Schulte and colleagues from the previous chapter

is first used to generate ground truth for the validation of the ability of the mechanoregulation analysis to recover mechanoregulation in HR-pQCT images. The developed contour approach is used to automatically process longitudinal *in vivo* HR-pQCT fracture images of patients. Lastly, mechanoregulation is assessed for the different stages of fracture healing and compared to the intact contralateral site of each patient.

Chapter 5 is the synthesis of this thesis. It summarized the major findings, the limitations of the presented work, and provides an outlook for future research.

References

- Augat, P., Merk, J., Ignatius, A., Margevicius, K., Bauer, G., Rosenbaum, D., Claes, L., 1996. Early, Full Weightbearing With Flexible Fixation Delays Fracture Healing. *Clin. Orthop. Relat. Res.* 328, 194–202. <https://doi.org/10.1097/00003086-199607000-00031>
- Augat, P., Simon, U., Liedert, A., Claes, L., 2005. Mechanics and mechano-biology of fracture healing in normal and osteoporotic bone. *Osteoporos. Int.* 16, 36–43. <https://doi.org/10.1007/s00198-004-1728-9>
- Badilatti, S.D., Christen, P., Parkinson, I., Müller, R., 2016. Load-adaptive bone remodeling simulations reveal osteoporotic microstructural and mechanical changes in whole human vertebrae. *J. Biomech.* 49, 3770–3779. <https://doi.org/10.1016/j.jbiomech.2016.10.002>
- Boerckel, J.D., Kolambkar, Y.M., Stevens, H.Y., Lin, A.S.P., Dupont, K.M., Guldberg, R.E., 2012. Effects of in vivo mechanical loading on large bone defect regeneration. *J. Orthop. Res.* 30, 1067–1075. <https://doi.org/10.1002/jor.22042>
- Boutroy, S., Buxsein, M.L., Munoz, F., Delmas, P.D., 2005. In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. *J. Clin. Endocrinol. Metab.* 90, 6508–6515. <https://doi.org/10.1210/jc.2005-1258>
- Boutroy, S., Van Rietbergen, B., Sornay-Rendu, E., Munoz, F., Buxsein, M.L., Delmas, P.D., 2008. Finite element analysis based on in vivo HR-pQCT images of the distal radius is associated with wrist fracture in postmenopausal women. *J. Bone Miner. Res.* 23, 392–399. <https://doi.org/10.1359/jbmr.071108>
- Carter, D.R., 1987. Mechanical loading history and skeletal biology. *J. Biomech.* 20, 1095–1109. [https://doi.org/10.1016/0021-9290\(87\)90027-3](https://doi.org/10.1016/0021-9290(87)90027-3)

- Carter, D.R., Beaupré, G.S., Giori, N.J., Helms, J.A., 1998. Mechanobiology of Skeletal Regeneration. *Clin. Orthop. Relat. Res.* 355S, S41–S55. <https://doi.org/10.1097/00003086-199810001-00006>
- Christen, P., Ito, K., Ellouz, R., Boutroy, S., Sornay-Rendu, E., Chapurlat, R.D., Van Rietbergen, B., 2014. Bone remodelling in humans is load-driven but not lazy. *Nat. Commun.* 5. <https://doi.org/10.1038/ncomms5855>
- Christen, P., Schulte, F.A., Zwahlen, A., van Rietbergen, B., Boutroy, S., Melton, L.J., Amin, S., Khosla, S., Goldhahn, J., Müller, R., 2016. Voxel size dependency, reproducibility and sensitivity of an in vivo bone loading estimation algorithm. *J. R. Soc. Interface* 13, 20150991. <https://doi.org/10.1098/rsif.2015.0991>
- Claes, L.E., Heigele, C.A., Neidlinger-Wilke, C., Kaspar, D., Seidl, W., Margevicius, K.J., Augat, P., 1998. Effects of Mechanical Factors on the Fracture Healing Process. *Clin. Orthop. Relat. Res.* 355S, S132–S147. <https://doi.org/10.1097/00003086-199810001-00015>
- Court-Brown, C.M., Caesar, B., 2006. Epidemiology of adult fractures: A review. *Injury* 37, 691–697. <https://doi.org/10.1016/j.injury.2006.04.130>
- Dimitrov, D. V., 2016. Medical internet of things and big data in healthcare. *Healthc. Inform. Res.* 22, 156–163. <https://doi.org/10.4258/hir.2016.22.3.156>
- Frost, H.M., 1987. The mechanostat: a proposed pathogenic mechanism of osteoporoses and the bone mass effects of mechanical and nonmechanical agents. *Bone Miner.* 2, 73–85.
- Gardner, M.J., van der Meulen, M.C.H., Demetrakopoulos, D., Wright, T.M., Myers, E.R., Bostrom, M.P., 2006. In vivo cyclic axial compression affects bone healing in the mouse tibia. *J. Orthop. Res.* 24, 1679–1686. <https://doi.org/10.1002/jor.20230>
- Giannoudis, P. V., Einhorn, T.A., Marsh, D., 2007. Fracture healing: The diamond concept. *Injury* 38, S3–S6. [https://doi.org/10.1016/S0020-1383\(08\)70003-2](https://doi.org/10.1016/S0020-1383(08)70003-2)

-
- Gietzelt, M., Löpprich, M., Karmen, C., Ganzinger, M., 2016. Models and Data Sources Used in Systems Medicine. *Methods Inf. Med.* 55, 107–113. <https://doi.org/10.3414/ME15-01-0151>
- Hill, P.A., 1998. Bone remodelling. *Br. J. Orthod.* 25, 101–107. <https://doi.org/10.1093/ortho/25.2.101>
- Huiskes, R., Rulmerman, R., Van Lenthe, G.H., Janssen, J.D., 2000. Effects of mechanical forces on maintenance and adaptation of form in trabecular bone. *Nature* 405, 704–706. <https://doi.org/10.1038/35015116>
- Isaksson, H., Gröngröft, I., Wilson, W., Van Donkelaar, C.C., Ven Rietbergen, B., Tami, A., Huiskes, R., Ito, K., 2009. Remodeling of fracture callus in mice is consistent with mechanical loading and bone remodeling theory. *J. Orthop. Res.* 27, 664–672. <https://doi.org/10.1002/jor.20725>
- Johnell, O., Kanis, J.A., 2006. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos. Int.* 17, 1726–1733. <https://doi.org/10.1007/s00198-006-0172-4>
- Kessel, K.A., Combs, S.E., 2016. Review of developments in electronic, clinical data collection, and documentation systems over the last decade - are we ready for big data in routine health care? *Front. Oncol.* 6, 1–6. <https://doi.org/10.3389/fonc.2016.00075>
- Kirmani, S., Christen, D., Van Lenthe, G.H., Fischer, P.R., Bouxsein, M.L., McCready, L.K., Melton, L.J., Riggs, B.L., Amin, S., Müller, R., Khosla, S., 2009. Bone structure at the distal radius during adolescent growth. *J. Bone Miner. Res.* 24, 1033–1042. <https://doi.org/10.1359/jbmr.081255>
- Klein, P., Schell, H., Streitparth, F., Heller, M., Kassi, J.P., Kandziora, F., Bragulla, H., Haas, N.P., Duda, G.N., 2003. The initial phase of fracture healing is specifically sensitive to mechanical conditions. *J. Orthop. Res.* 21, 662–669. [https://doi.org/10.1016/S0736-0266\(02\)00259-0](https://doi.org/10.1016/S0736-0266(02)00259-0)

- Lambers, F.M., Kuhn, G., Weigt, C., Koch, K.M., Schulte, F.A., Müller, R., 2015. Bone adaptation to cyclic loading in murine caudal vertebrae is maintained with age and directly correlated to the local micromechanical environment. *J. Biomech.* 48, 1179–1187. <https://doi.org/10.1016/j.jbiomech.2014.11.020>
- Luo, J., Wu, M., Gopukumar, D., Zhao, Y., 2016. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed. Inform. Insights* 8, BII.S31559. <https://doi.org/10.4137/bii.s31559>
- Manske, S.L., Zhu, Y., Sandino, C., Boyd, S.K., 2015. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone* 79, 213–221. <https://doi.org/10.1016/j.bone.2015.06.006>
- Moustafa, A., Sugiyama, T., Prasad, J., Zaman, G., Gross, T.S., Lanyon, L.E., Price, J.S., 2012. Mechanical loading-related changes in osteocyte sclerostin expression in mice are more closely associated with the subsequent osteogenic response than the peak strains engendered. *Osteoporos. Int.* 23, 1225–1234. <https://doi.org/10.1007/s00198-011-1656-4>
- Murdoch, T.B., Detsky, A.S., 2013. The inevitable application of big data to health care. *JAMA - J. Am. Med. Assoc.* 309, 1351–1352. <https://doi.org/10.1001/jama.2013.393>
- Ng, C.Y., McQueen, M.M., 2011. What are the radiological predictors of functional outcome following fractures of the distal radius? *J. Bone Jt. Surg. - Ser. B* 93 B, 145–150. <https://doi.org/10.1302/0301-620X.93B2.25631>
- Nicks, K.M., Amin, S., Atkinson, E.J., Riggs, B.L., Melton, L.J., Khosla, S., 2012. Relationship of age to bone microstructure independent of areal bone mineral density. *J. Bone Miner. Res.* 27, 637–644. <https://doi.org/10.1002/jbmr.1468>
- Nishiyama, K.K., Macdonald, H.M., Hanley, D.A., Boyd, S.K., 2013. Women with previous fragility fractures can be classified based on bone microarchitecture and finite element

-
- analysis measured with HR-pQCT. *Osteoporos. Int.* 24, 1733–1740. <https://doi.org/10.1007/s00198-012-2160-1>
- Nishiyama, K.K., Shane, E., 2013. Clinical imaging of bone microarchitecture with HR-pQCT. *Curr. Osteoporos. Rep.* <https://doi.org/10.1007/s11914-013-0142-7>
- Ohs, N., Keller, F., Blank, O., Wayne Lee, Y.W., Jack Cheng, C.Y., Arbenz, P., Müller, R., Christen, P., 2016. Towards in silico prognosis using big data. *Curr. Dir. Biomed. Eng.* 2, 57–60. <https://doi.org/10.1515/cdbme-2016-0016>
- Qin, C., Tao, L., Phang, Y.H., Zhang, C., Chen, S.Y., Zhang, P., Tan, Y., Jiang, Y.Y., Chen, Y.Z., 2015. The Assessment of the Readiness of Molecular Biomarker-Based Mobile Health Technologies for Healthcare Applications. *Sci. Rep.* 5, 1–14. <https://doi.org/10.1038/srep17854>
- Rho, J.Y., Kuhn-Spearing, L., Zioupos, P., 1998. Mechanical properties and the hierarchical structure of bone. *Med. Eng. Phys.* 20, 92–102. [https://doi.org/10.1016/S1350-4533\(98\)00007-1](https://doi.org/10.1016/S1350-4533(98)00007-1)
- Robling, A.G., Turner, C.H., 2009. Mechanical signaling for bone modeling and remodeling. *Crit. Rev. Eukaryot. Gene Expr.* 19, 319–338. <https://doi.org/10.1615/CritRevEukarGeneExpr.v19.i4.50>
- Röntgen, V., Blakytyn, R., Matthys, R., Landauer, M., Wehner, T., Göckelmann, M., Jermendy, P., Amling, M., Schinke, T., Claes, L., Ignatius, A., 2010. Fracture healing in mice under controlled rigid and flexible conditions using an adjustable external fixator. *J. Orthop. Res.* 28, 1456–1462. <https://doi.org/10.1002/jor.21148>
- Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2011. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. *Bone* 48, 433–442. <https://doi.org/10.1016/j.bone.2010.10.007>

- Schulte, F.A., Ruffoni, D., Lambers, F.M., Christen, D., Webster, D.J., Kuhn, G., Müller, R., 2013a. Local Mechanical Stimuli Regulate Bone Formation and Resorption in Mice at the Tissue Level. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0062172>
- Schulte, F.A., Zwahlen, A., Lambers, F.M., Kuhn, G., Ruffoni, D., Betts, D., Webster, D.J., Müller, R., 2013b. Strain-adaptive in silico modeling of bone adaptation - A computer simulation validated by in vivo micro-computed tomography data. *Bone* 52, 485–492. <https://doi.org/10.1016/j.bone.2012.09.008>
- Sornay-Rendu, E., Boutroy, S., Duboeuf, F., Chapurlat, R.D., 2017. Bone Microarchitecture Assessed by HR-pQCT as Predictor of Fracture Risk in Postmenopausal Women: The OFELY Study. *J. Bone Miner. Res.* 32, 1243–1251. <https://doi.org/10.1002/jbmr.3105>
- Sornay-Rendu, E., Boutroy, S., Munoz, F., Delmas, P.D., 2007. Alterations of cortical and trabecular architecture are associated with fractures in postmenopausal women, partially independent of decreased BMD measured by DXA: The OFELY study. *J. Bone Miner. Res.* 22, 425–433. <https://doi.org/10.1359/jbmr.061206>
- Stratton, M.R., Campbell, P.J., Futreal, P.A., 2009. The cancer genome. *Nature* 458, 719–724. <https://doi.org/10.1038/nature07943>
- Tjong, W., Kazakia, G.J., Burghardt, A.J., Majumdar, S., 2012. The effect of voxel size on high-resolution peripheral computed tomography measurements of trabecular and cortical bone microstructure. *Med. Phys.* 39, 1893–1903. <https://doi.org/10.1118/1.3689813>
- Tourolle né Betts, D.C., 2019. A micro-scale multiphysics framework for fracture healing and bone remodelling. <https://doi.org/10.3929/ethz-b-000364637>
- United Nations, 2020. World Population Ageing 2019 (ST/ESA/SER.A/444) [WWW Document]. Dep. Econ. Soc. Aff. Popul. Div. URL www.unpopulation.org

-
- Webster, D., Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2015. Strain energy density gradients in bone marrow predict osteoblast and osteoclast activity: A finite element study. *J. Biomech.* 48, 866–874. <https://doi.org/10.1016/j.jbiomech.2014.12.009>
- Webster, D.J., Morley, P.L., van Lenthe, H.H., Müller, R., 2008. A novel in vivo mouse model for mechanically stimulated bone adaptation - A combined experimental and computational validation study. *Comput. Methods Biomech. Biomed. Engin.* 11, 435–441. <https://doi.org/10.1080/10255840802078014>
- Wehrle, E., Liedert, A., Heilmann, A., Wehner, T., Bindl, R., Fischer, L., Haffner-Luntzer, M., Jakob, F., Schinke, T., Amling, M., Ignatius, A., 2015. The impact of low-magnitude high-frequency vibration on Fracture healing is profoundly influenced by the oestrogen status in mice. *DMM Dis. Model. Mech.* 8, 93–104. <https://doi.org/10.1242/dmm.018622>
- Wehrle, E., Tourolle né Betts, D.C., Kuhn, G.A., Scheuren, A.C., Hofmann, S., Müller, R., 2019. Evaluation of longitudinal time-lapsed in vivo micro-CT for monitoring fracture healing in mouse femur defect models. *Sci. Rep.* 9, 1–12. <https://doi.org/10.1038/s41598-019-53822-x>
- Yu, W.S., Chan, K.Y., Yu, F.W.P., Ng, B.K.W., Lee, K.M., Qin, L., Lam, T.P., Cheng, J.C.Y., 2014. Bone structural and mechanical indices in Adolescent Idiopathic Scoliosis evaluated by high-resolution peripheral quantitative computed tomography (HR-pQCT). *Bone* 61, 109–115. <https://doi.org/10.1016/j.bone.2013.12.033>
- Zhu, T.Y., Griffith, J.F., Qin, L., Hung, V.W., Fong, T.N., Au, S.K., Li, M., Lam, Y.Y.O., Wong, C.K., Kwok, A.W., Leung, P.C., Li, E.K., Tam, L.S., 2014. Alterations of bone density, microstructure, and strength of the distal radius in male patients with rheumatoid arthritis: A case-control study with HR-pQCT. *J. Bone Miner. Res.* 29, 2118–2129. <https://doi.org/10.1002/jbmr.2221>

Chapter 2

Background

2.1 Validation of HR-pQCT against micro-CT for morphometric and biomechanical analyses: A review

Nicholas Ohs¹, Caitlyn Collins¹, Penny Atkins¹

¹Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

Published in:

Bone Reports 13 (2020)

Postprint version according to publisher copyright policy

Abstract

High-resolution peripheral quantitative computed-tomography (HR-pQCT) has the potential to become a powerful clinical assessment and diagnostic tool. Given the recent improvements in image resolution, from 82 to 61 μm , this technology may be used to accurately quantify *in vivo* bone microarchitecture, a key biomarker of degenerative bone diseases. However, computational methods to assess bone microarchitecture were developed for micro computed tomography (micro-CT), a higher-resolution technology only available for *ex vivo* studies, and validation of these computational analysis techniques against the gold-standard micro-CT has been inconsistent and incomplete. Herein, we review methods for segmentation of bone compartments and microstructure, quantification of bone morphology, and estimation of mechanical strength using finite-element analysis, highlighting the need throughout for improved standardization across the field.

Studies have relied on homogenous datasets for validation, which does not allow for robust comparisons between methods. Consequently, the adaptation and validation of novel segmentation approaches has been slow to non-existent, with most studies still using the manufacturer's segmentation for morphometric analysis despite the existence of better performing alternative approaches. The promising accuracy of HR-pQCT for capturing morphometric indices is overshadowed by considerable variability in outcomes between studies. For finite element analysis (FEA) methods, the use of disparate material models and

FEA tools has led to a fragmented ability to assess mechanical bone strength with HR-pQCT. Further, the scarcity of studies comparing 62 μm HR-pQCT to the gold standard micro-CT leaves the validation of this imaging modality incomplete.

This review revealed that without standardization, the capabilities of HR-pQCT cannot be adequately assessed. The need for a public, extendable, heterogeneous dataset of HR-pQCT and corresponding gold-standard micro-CT images, which would allow HR-pQCT users to benchmark existing and novel methods and select optimal methods depending on the scientific question and data at hand, is now evident. With more recent advancements in HR-pQCT, the community must learn from its past and provide properly validated technologies to ensure that HR-pQCT can truly provide value in patient diagnosis and care.

Keywords:

HR-pQCT; micro-CT; patient evaluation; bone morphometrics; finite element analysis; segmentation

2.1.1 Introduction

Micro computed tomography (micro-CT) was first used to analyse the trabecular structure of bone in three dimensions in 1989 (Feldkamp et al., 1989). Shortly thereafter, the first commercial micro-CT device was made available (Rüegsegger et al., 1996); this quickly led to widespread adoption of micro-CT as a standard research tool for bone tissue analysis at the micrometre (μm) scale. In contrast to two-dimensional (2D) methods, the three-dimensional (3D) acquisition capabilities and high resolution (up to $1\ \mu\text{m}$) allowed for direct assessment of 3D structures. By 2008, the plethora of commonly used micro-CT technologies and bone morphology analysis techniques with translational applications warranted a comprehensive book chapter (Stauber and Müller, 2008). More recently, micro-CT-based finite element analysis (FEA) has provided an experimentally validated method to assess bone mechanical strength and failure non-invasively (Chen et al., 2017; Hambli, 2013).

In parallel with the laboratory development of micro-CT, peripheral quantitative computed tomography (pQCT) emerged as a potential method for identifying risk factors of disease, such as osteoporosis (Müller et al., 1989). Given that existing clinical tools, such as dual energy x-ray absorptiometry (DXA), were unable to adequately identify patient at risk of fracture, pQCT provided an enhanced method to evaluate bone clinically (Bolotin and Sievänen, 2001; Järvinen et al., 2008; Nelson et al., 2002). While the resolution of pQCT is lower than micro-CT ($170\ \mu\text{m}$ voxels), it enabled patient imaging and non-invasive, time-lapse patient studies. Comparisons with micro-CT measurements were, however, limited due to the order of magnitude difference in image resolution. Moreover, the resolution of pQCT prohibited accurate assessment of bone microarchitecture for patients with degenerative bone diseases, such as osteoporosis, where the average thickness of individual trabeculae is $200\ \mu\text{m}$ (Borah et al., 2006, 2004). To combat this, high-resolution pQCT (HR-pQCT) devices were introduced as improved successors to pQCT devices. These devices (XtremeCT I and II, Scanco Medical) have isotropic voxel sizes of 82 and $61\ \mu\text{m}$, respectively, which has allowed for direct comparison to micro-CT (Figure 2.1.1). Unfortunately, the conclusions of these studies have been inconsistent, potentially due to the varied methods of calculation for the parameters used for comparison of HR-pQCT images to micro-CT images.

Since HR-pQCT devices are now available and in use in many clinical centres, the establishment of image processing and analysis standards are crucial for comparisons between

studies and research groups. Therefore, we aim to provide a comprehensive review of existing computational processing and analysis tools which have been validated against the micro-CT gold standard for HR-pQCT and highlight the need for improved standardization across the field.

2.1.2 Data

For this purpose, the keywords HR-pQCT and micro-CT were searched on PubMed. Due to automatic term mapping, this included an extensive list of closely related terms. Of the 68 papers identified, 31 provided data comparing HR-pQCT and micro-CT (Table 2.1.1). The most common exclusion criterion was the lack of micro-CT data, despite the use of the keyword in the paper. Of the selected studies, 25 evaluated only 82 μm (first generation) HR-pQCT, and one study evaluated both 82 μm and 61 μm HR-pQCT (Figure 2.1.1). The remaining five

Table 2.1.1 HR-pQCT image modalities, anatomical sites, and reported disease status or other classification of samples used in validation studies.

Study	82 μm	61 μm	Anatomical Site(s)	Sample Classification (as applicable)
(Buie et al., 2007)	Yes	No	Radius/ Tibia	-
(Burghardt et al., 2007)	Yes	No	Femoral Head	Degenerative Joint Disease
(MacNeil and Boyd, 2007)	Yes	No	Radius	-
(Varga and Zysset, 2009)	Yes	No	Radius	-
(Cohen et al., 2010)	Yes	No	Radius/ Tibia/ Iliac Crest ^a	Idiopathic Osteoporosis/ Hypoparathyroidism
(Liu et al., 2010)	Yes	No	Tibia	-
(Nishiyama et al., 2010)	Yes	No	Radius	Osteopenia
(Liu et al., 2011)	Yes	Emulated ^b	Tibia	-
(Pahr et al., 2012)	Yes	No	Vertebrae	Fracture

(Tjong et al., 2012)	Yes	Emulated ^c	Radius	-
(Liu et al., 2013)	Yes	No	Tibia	Fracture
(Zebaze et al., 2013)	Yes	No	Radius/ Tibia	-
(Krause et al., 2014)	Yes	No	Radius/ Tibia	Osteoporosis
(Ostertag et al., 2014)	Yes	No	Tibia	-
(Jorgenson et al., 2015)	Yes	No	Tibia	-
(Manske et al., 2015)	Yes/ Emulated ^d	Yes	Radius	-
(Christen et al., 2016)	Emulated ^b	Emulated ^b	Radius	-
(Klintström et al., 2016)	Yes	No	Radius	-
(Ostertag et al., 2016)	Yes	No	Tibia/ Femur	-
(Scharmga et al., 2016)	Yes	No	Finger	-
(Zhou et al., 2016)	Yes	No	Radius/ Tibia	-
(Hosseini et al., 2017)	Emulated ^c	Yes	Radius	-
(Peters et al., 2017)	Yes	No	Joints of Index Finger	-
(Werner et al., 2017)	Yes	No	Finger	Rheumatoid Arthritis
(Alsayednoor et al., 2018)	Yes	No	Calcaneus	-
(Chiang et al., 2018)	Yes	No	Radius	Menopause
(Metcalf et al., 2018)	Yes	No	Calcaneus	-
(Ang et al., 2020)	Yes/ No	No/ Yes	Radius/ Tibia	-
(Mys et al., 2019)	Yes	Yes	Trapezium	Severe Arthritis
(Soltan et al., 2019)	Yes	No	Radius	-
(Wang et al., 2019)	Yes	No	Radius/ Tibia	-

^aScanned with micro-CT only

^bDownscaled micro-CT images

^cScanned at 41 μm using a 82 μm HR-pQCT device in non-patient mode

^dComparison to micro-CT only with emulated data

^eScanned using a 61 μm HR-pQCT device

studies looked at both HR-pQCT device resolutions, but emulated one or both resolutions by acquiring lower-resolution images on a higher-resolution device or by scaling down higher-

20

resolution images in post-processing (Table 2.1.1). Studies that utilized 82 μm images, either emulated or from XtremeCT I, are referred to as 82 μm HR-pQCT studies, while studies that utilized 61 μm images, either emulated or from XtremeCT II, are referred to as 61 μm HR-pQCT studies.

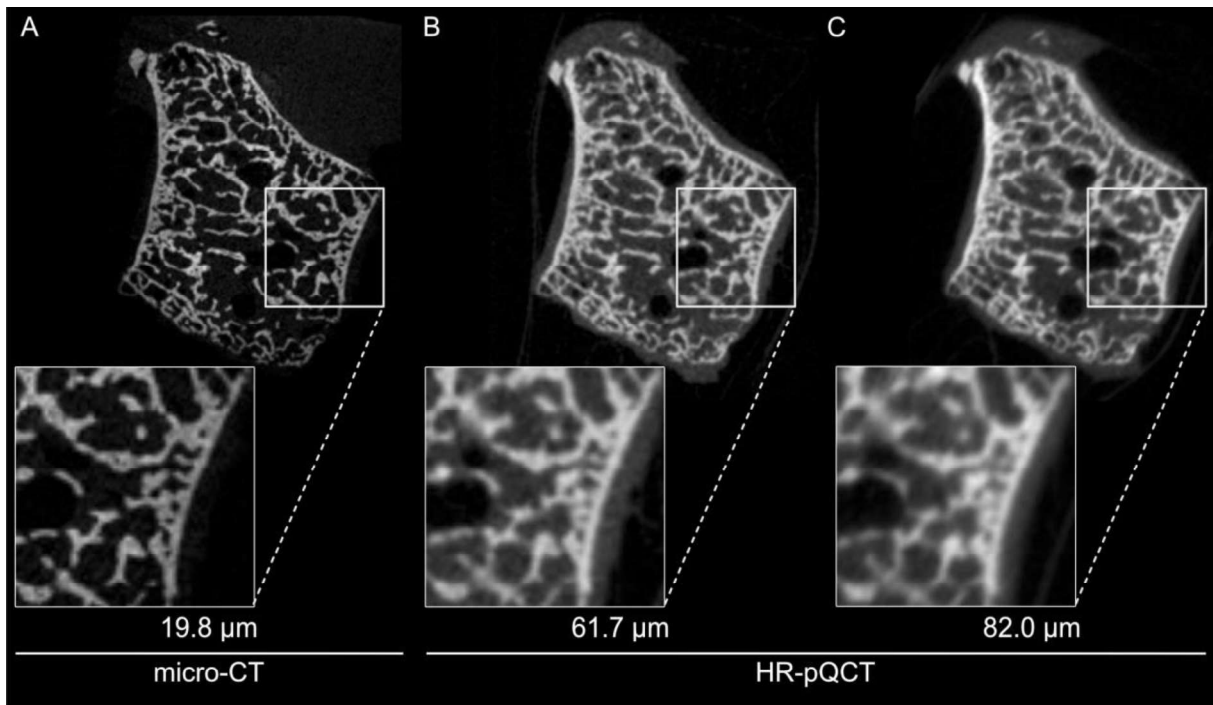


Fig. 2.1.1 With increasing voxel size, pores and edges of smaller features become difficult to capture using computed tomography techniques. In the reconstructed (A) micro-CT, (B) 61 μm HR-pQCT, and (C) 82 μm HR-pQCT images of an *ex vivo* trapezium, extracted from the wrist of a patient with severe arthritis, pores near the cortex become hard to distinguish with decreasing image resolution. Adapted from (Mys et al., 2019).

All selected studies utilized human bone samples. While many studies evaluated a single anatomical site, eight studies analysed more than one anatomical site. In total, the radius was evaluated in 17 studies, the tibia in 11, the femur in two, various locations in the hand (trapezium or finger joints) in four, the calcaneus in two, the vertebrae in one, and the iliac crest in one (Table 2.1.1). Studies included data from healthy patients, as well as patients with idiopathic osteoporosis, idiopathic hypoparathyroidism, osteopenia, osteoporosis, previous fracture, severe arthritis, and degenerative joint diseases.

2.1.3 Segmentation

Most micro-CT image analysis requires segmentation, which is the identification and partitioning of objects and boundaries of interest, as a pre-processing step. Only seven studies used a common segmentation procedure, while the remaining studies used different procedures for each imaging modality. Seventeen studies used segmentation to separate the cortical shell and the trabecular compartment (Figure 2.1.2A) and 20 studies separated the individual trabeculae from the background (Figure 2.1.2B). Despite the lack of an industry standard for micro-CT segmentation and the increased noise and blurring of lower resolution images (Figure 2.1.1C), micro-CT segmentation methods were commonly used to develop and validate approaches for HR-pQCT.

2.1.3.1 Segmenting Bone Compartments

Trabecular and cortical compartments were segmented using the manufacturer's approach for HR-pQCT, the gold-standard approach micro-CT, and alternative approaches to segment both types of images.

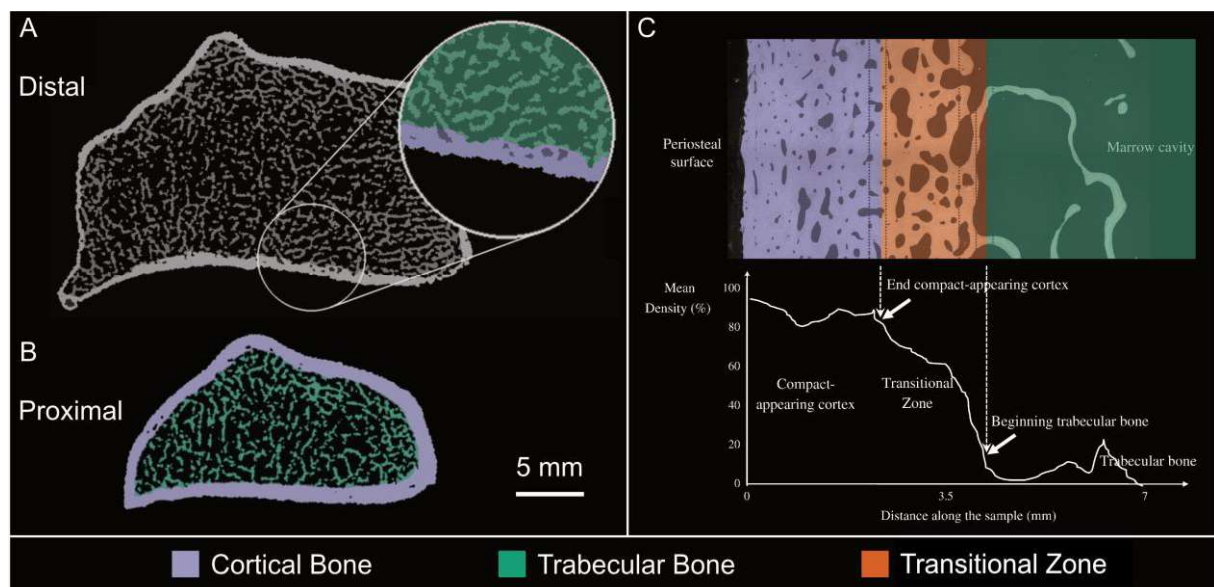


Fig. 2.1.2 Micro-CT and HR-pQCT image segmentation algorithms predominantly target separation of the cortex and trabecular compartment. (A) The standard manufacturer approach for segmentation of HR-pQCT data was developed to identify the periosteal and endosteal margins of the distal radius (pictured) or tibia and separate the bone into two compartments: cortical (purple) and trabecular (green); (B) Quantification of bone microstructure in these

compartments, for example the morphological assessment of individual trabeculae (highlighted in green) or cortical pores (black spots within the purple, cortical segmentation), can then be performed. Adapted from (Burghardt et al., 2010). (C) One alternative segmentation approach identifies an additional compartment, the transitional zone (orange), which is representative of the junction between cortical and trabecular compartments; here, intracortical remodelling can blur the line between cortical and trabecular compartments. Images in A and B were captured with 82 μm HR-pQCT, while the image in C was captured with 2.5 μm scanning electron microscopy. Adapted from (Zebaze et al., 2013).

2.1.3.1.1 HR-pQCT – Manufacturer’s Approach

The manufacturer segmentation procedures for HR-pQCT are widely used in HR-pQCT studies (Whittier et al., 2020). The approach for 82 μm HR-pQCT begins with a semi-automated hand-drawn contour of the periosteal surface, generated using the scanner manufacturer’s software. From here, the trabecular bone compartment is separated from the cortical bone compartment. Herein, the voxel-based intensity data is converted to a physical density of hydroxyapatite (HA) ($\text{mg HA}/\text{cm}^3$) and blurred into compartments of low- and high-density, generating a Gaussian-weighted mean image. A threshold of one-third of the apparent cortical bone density is then used to isolate the cortical and trabecular compartments. This segmentation procedure was implemented in a majority of the studies without modification.

The 61 μm HR-pQCT manufacturer segmentation procedure is based on a dual threshold technique and a cortical pore analysis to separate the cortex from the trabecular compartment (Figure 2.1.2A) and was developed to avoid the introduction of operator error associated with hand-drawn contours. The dual threshold technique reduces the inclusion of noise by applying a higher value threshold to segment the higher-density cortex. Closing and connectivity operations are applied to generate the bone outer segmentation. This method works well provided there are no large gaps in the cortex, e.g. large Volkmann’s canals, which disturb the connectivity operations. Large-radius blurring provides a smooth transition from the cortex into the trabecular compartment and a lower value threshold isolates the trabecular compartment. Before implementation into the manufacturer’s software, the algorithm was applied to 82 μm HR-pQCT and micro-CT images and validated against the respective hand-drawn contours (Buie et al., 2007). The cortical pore analysis adds voids that are not connected to the background to the cortical compartment, as pores. The approach was validated against hand-drawn contours (Burghardt et al., 2010).

2.1.3.1.2 Gold-Standard Micro-CT Segmentation Method

For micro-CT data, the cortical and trabecular compartments are separated using hand-drawn contours. However, semi-automatic procedures, including interpolating between two hand-drawn contours or snapping periosteal hand-drawn contours to nearby edges in an image, have been included in segmentation software packages to speed up the process of segmentation.

2.1.3.1.3 Common Segmentation Using the Software Fiji

Soltan and colleagues used Fiji (Schindelin et al., 2012) to generate a periosteal surface segmentation for cadaveric radii scanned with both 82 μm HR-pQCT and synchrotron radiation micro-CT (17.7 μm voxels) (Soltan et al., 2019). For both image modalities, the segmentation was generated using a threshold of 400 mg HA/cm³ and an optimized number of erosion iterations, such that the automatically generated HR-pQCT cortical masks most closely resembled the hand-drawn micro-CT contours.

2.1.3.1.4 Common Segmentation of the Transitional Zone

Zebaze and colleagues proposed a three-compartment segmentation method which produced cortical and trabecular compartments, as well as a transitional zone (Figure 2.1.2C) (Zebaze et al., 2013). Through comparison to expert-generated hand-drawn contours of scanning electron microscopy images, their implementation produced accurate and reproducible cross-sectional areas of the three compartments for 82 μm HR-pQCT and micro-CT images. However, cortical interruptions greater than a few voxels, as would be expected for porous bone samples, resulted in method failure. For this reason, their analysis was limited to only 40 diaphyseal image slices, due to the reduced frequency of cortical interruptions in this region. Further, this segmentation method was implemented on 2D images, not on a 3D image volume, thus the application to consecutive slices would likely not result in a consistent and smooth surface, longitudinally. Importantly, Zebaze and colleagues both performed the validation study and co-wrote the patent behind the commercial software of this proposed method, StrAx1.0.

2.1.3.1.5 Common Segmentation with Thickness-Based Separation

Due to the lack of clarity in the transition between the cortical and trabecular compartments in the epiphysis, Ang and colleagues modified the dual-threshold segmentation algorithm (Buie et al., 2007) to consistently segment images of both the diaphysis and the epiphyses (Ang et al., 2020). Here, a blurred distance measure of all slice-based pixels touching the outer surface segmentation was cropped at the outer surface to generate the cortical compartment segmentation. This algorithm was applied to relatively low-resolution micro-CT (50 μm) and

HR-pQCT (both 61 and 82 μm) images and compared against the respective hand-drawn contours generated for each image.

2.1.3.2 Segmenting Bone Microstructure

For assessment of bone microstructure, the trabecular compartment is further segmented into bone and background. Since individual trabeculae can be only a few voxels thick for HR-pQCT images (Figure 2.1.1), obtaining an accurate representation of the scanned bone structure can be challenging.

2.1.3.2.1 HR-pQCT – Manufacturer’s Approach

Laib and colleagues introduced a two-step procedure on pQCT image (165 μm), which has since been implemented in the manufacturer’s software (Laib et al., 1998; Laib and R uegsegger, 1999a). The image is first filtered and then a threshold is applied to isolate the trabecular structure. For 82 μm HR-pQCT image, an edge-enhancing Laplace-Hamming filter is used in combination with a threshold of 400 after normalization of the image to 1000 Hounsfield units. For 61 μm HR-pQCT image, a noise-reducing and structure enhancing Gaussian filter is used (sigma 0.8, filter 3x3x3) in combination with a 320 mg HA/cm³ threshold (Hosseini et al., 2017; Manske et al., 2017, 2015; Mys et al., 2019).

2.1.3.2.2 Micro-CT Segmentation Methods

Micro-CT data is processed similarly to the 61 μm HR-pQCT images, however Gaussian settings and threshold values are varied. Gaussian sigma values ranged from 0.5 to 1.2 and filter sizes ranged from 3x3x3 to 5x5x5 (Table 2.1.2). With regards to threshold, some studies used specimen-specific thresholds given in percent of the maximum intensity value, while other studies used fixed thresholds given in mg HA/cm³ for all specimens. In more recent studies, starting in 2015, a fixed threshold in terms of mg HA/cm³ was more commonly used (Christen et al., 2016; Hosseini et al., 2017; Jorgenson et al., 2015; Manske et al., 2015; Soltan et al., 2019). Interestingly, none of the studies selected for this review used the same threshold, even when identical micro-CT devices were used (Table 2.1.2). When provided, the reason for choosing a particular threshold was either visual assessment (Burghardt et al., 2007; Manske et al., 2015; Metcalf et al., 2018), matching of morphometric parameters to the HR-pQCT scans (Peters et al., 2017), or being consistent with previous studies conducted using that dataset (Christen et al., 2016).

Table 2.1.2 Gaussian filter and threshold values used on gold standard micro-CT images.

Study	Morphometric Analysis	CT Device	Gaussian Settings		Voxel Size [μm]	Threshold
			Sigma	Support		
(Cohen et al., 2010)	Trabecular, Cortical	μCT 40	1.2	3x3x3	18	34% max
(MacNeil and Boyd, 2007)	Trabecular, Cortical	vivaCT 40	1.2	5x5x5	38	11.2% max
(Metcalf et al., 2018)	Trabecular	SkyScan 1172	1.2	5x5x5	17	33 % max ^a
(Hosseini et al., 2017)	Trabecular	μCT 100	0.8	3x3x3	16	320 mg HA/cm ³
(Burghardt et al., 2007)	Trabecular	μCT 40	-	-	16	25% max
(Manske et al., 2015)	Trabecular	XtremeCT II	0.8	3x3x3	30.3	390 mg HA/cm ³
(Ostertag et al., 2014)	Cortical	SR Grenoble	-	-	7.5	53.9% max
(Ostertag et al., 2016)	Cortical	Skyscan 1172	-	-	7.5	15.6% max
(Chiang et al., 2018)	Cortical	vivaCT 40	-	-	19	960 mg HA/cm ³
(Soltan et al., 2019)	Cortical	SR Canadian Light Source	-	-	17.7	400 mg HA/cm ³
(Nishiyama et al., 2010)	Cortical	vivaCT 40	1.2	5x5x5	19	18.4% max
(Peters et al., 2017)	Cortical	μCT 80	0.8	3x3x3	18	24.7% max
(Alsayednoor et al., 2018)	Trabecular	SkyScan 1172	0.5	5x5x5	17.4	90 ^b
(Christen et al., 2016)	Trabecular	vivaCT 80	1.2	3x3x3	25	450 mg HA/cm ³
(Jorgenson et al., 2015)	Cortical	μCT 100	-	-	9	750 mg HA/cm ³

^aFor select cases, the threshold was visually determined to be 35% of the maximum

^bUnknown units

Other terms: HA, hydroxyapatite; SR, synchrotron radiation.

2.1.3.2.3 Common Segmentation Methods for HR-pQCT and Micro-CT

A histogram approach (Ridler and Calvard, 1978) was used on both micro-CT and 82 μm HR-pQCT images to separate bone from background using an image-specific, histogram-based threshold (Varga and Zysset, 2009). Alternatively, Hosseini and colleagues used the manufacturer based filtering for each image modality followed by a fixed global threshold of 320 mg HA/cm³ for all images (Hosseini et al., 2017).

2.1.3.2.4 Alternative Segmentation Methods for HR-pQCT

Several additional segmentation methods have been developed specifically for either HR-pQCT or micro-CT images. One such strategy is that of local adaptive thresholds, which uses gradient edge detection to separate trabeculae from background in 82 μm HR-pQCT images (Burghardt et al., 2007). The algorithm analyses voxels at the edges of a scanned bone structure to preserve all trabeculae without filling small pores.

Another approach is to use an automated 3D region growing algorithm (ARG) (Revol-Muller et al., 2002) which has been compared to the manufacturer's default segmentation for 82 μm HR-pQCT images (Klintström et al., 2016). The region of interest is isolated using a very high threshold and then grown over multiple iterations using decreasing thresholds. The iteration that performed best with regards to an unspecified assessment function is chosen as the final segmentation. Importantly, this study used an alternative, Otsu histogram-based segmentation approach for their gold standard micro-CT images (Otsu, 1979), thus it is difficult to conclude whether the approach is accurate. Alsayednoor and colleagues (Alsayednoor et al., 2018) developed a method to segment 82 μm HR-pQCT images that preserved the underlying geometry of the bone structure according to fractal theory (Alberich-Bayarri et al., 2010). Therein, a geometry-preserving threshold that best matched the fractal dimension curves generated for each HR-pQCT and corresponding micro-CT image was chosen for each specimen, which is important for mechanical analysis (Alsayednoor et al., 2018). Notably, for morphometric analysis a threshold preserving the ratio of bone volume to compartment volume can be used instead.

2.1.3.3 Discussion

A variety of approaches have been shown to successfully segment aspects of the trabecular structure, the cortex, and the transitional zone. However, only one study compared the results of their proposed segmentation method on 61 μm HR-pQCT images directly to those of micro-

CT (Ang et al., 2020). Importantly, the quality of segmentation was nearly always validated relative to morphometric measurements or finite element analysis results, instead of directly, which may not provide adequate validation. Furthermore, none of the studies included in this analysis attempted to validate their segmentation techniques on fractured bone or degenerative disease cases, such as osteoporosis or osteoarthritis.

Since none of the reviewed segmentation approaches were capable of segmenting cortices with larger gaps, an alternative approach, such as 3D active contours which can be tuned parametrically (Caselles et al., 1997; Hafri et al., 2016a, 2016b; Kass et al., 1988; Marquez-Neila et al., 2014), might be necessary to provide accurate automatic segmentation for bones with reduced cortical connectivity. Due to the scope of this review, methods not yet validated against gold standard micro-CT, e.g. (Treece et al., 2012, 2010), have not been mentioned but might prove useful for the analysis of HR-pQCT images after validation.

Due to the limited validation of most approaches, a single method cannot be recommended without reservation. One of the greatest problems in identifying an optimal segmentation approach is the lack of available datasets for comparison. Currently, each study relies on a single, often homogenous, dataset for evaluation, while the availability of an extendable, publicly available, and diverse dataset of micro-CT and corresponding HR-pQCT images of various anatomical sites and disease states would allow for the direct comparison of the performance of different approaches. This would, in turn, allow for a clear recommendation as to which segmentation approach should be used for each type of data.

2.1.4 Morphometrics

Degenerative bone diseases not only alter the overall bone mass, but also the underlying microstructure of the bone (Brandi, 2009; Zhang et al., 2010). Hence, methods to quantitatively characterize bone microstructure (static parameters) and microstructural change over time (dynamic parameters) were developed for micro-CT (Goulet et al., 1994; Hildebrand et al., 1999; Hildebrand and Rüegsegger, 1997; Odgaard and Gundersen, 1993; Rüegsegger et al., 1996; Schulte et al., 2011; Wachter et al., 2001; Whitehouse, 1974). With the advent of HR-pQCT, these existing methods have been applied to HR-pQCT images and working alternatives have been investigated. In total, 25 of the selected studies evaluated the accuracy of morphometric parameters derived from HR-pQCT images relative to those derived from micro-

CT images. Several of these parameters have been implemented in the manufacturer's device software, increasing their frequency within the literature.

Anatomical sites used for these morphometric comparisons included the radius (15 studies), tibia (10 studies), femoral head (one study), calcaneus (one study), and trapezium (one study). Generally, the median prediction accuracy was in the R^2 range of 0.50 to 0.91 across all morphometric parameters. However, one study comparing 82 μm HR-pQCT images of the radius and tibia to micro-CT images of iliac crest biopsies reported lower morphometric prediction accuracy than all other evaluated studies (R^2 of 0.02 to 0.27) (Cohen et al., 2010). The noticeable drop in prediction accuracy demonstrates the importance of anatomical site matching when performing such comparisons. Due to the difficulty in interpreting these results, the study by Cohen and colleagues has not been included in this Section. For studies including several anatomical sites, R^2 values have been reported separately for each anatomical site.

2.1.4.1 Analysis Methods

The compartmental segmentation gave rise to two separate classes of morphometric parameters: trabecular and cortical parameters. Parameters have been abbreviated based on standardized nomenclature (Parfitt et al., 2009).

2.1.4.1.1 Regression Analysis

Data are presented in the format of (median R^2 value for 82 μm results, median R^2 value for 61 μm results), unless otherwise noted. R^2 values are described as good ($R^2 \geq 0.80$), moderate ($0.60 \leq R^2 < 0.80$), slight ($0.40 \leq R^2 < 0.60$), or poor ($R^2 < 0.4$). When available, regression analysis values (slope and intercept) were extracted and converted to the form $\text{slope} * \text{value}_{\text{HR-pQCT}} + \text{intercept} = \text{value}_{\text{micro-CT}}$, as necessary. Linear regressions were only reported when data from at least five studies was provided. The ideal regression parameters (slope=1, intercept=0) indicate perfect agreement between HR-pQCT and micro-CT results. Regression parameters deviating from this optimum indicate that the HR-pQCT results have to be calibrated to match those of micro-CT. Greater regression parameters (slope>1, intercept>0) indicate an underestimation of morphometric values by HR-pQCT, whereas lesser parameters (slope<1, intercept<0) indicate overestimation. For mixed parameters, over- and underestimation depend on the morphometric value.

2.1.4.1.2 Indirect and Direct Methods

The analysis of morphometrics has been split into direct and indirect comparisons. Direct refers to the direct application of micro-CT algorithms, while indirect refers to morphometrics calculated using the recommended manufacturer's approach. While both methods were used for 82 μm HR-pQCT, only direct methods were used for 61 μm HR-pQCT. Similar to segmentation, the manufacturer's approach for morphometric parameter calculation is based on studies performed using pQCT devices. Since any modifications to these original methods by the manufacturer have not been published and can be made at any time without notice, a comprehensive overview of these methods can only be provided by the manufacturer.

2.1.4.2 Trabecular Parameters

Since HR-pQCT images have increased noise and only resolve single trabeculae with a few voxels (Figure 2.1.1B, 2.1.1C), a thorough validation of analysis methods against a high quality gold standard such as micro-CT is necessary.

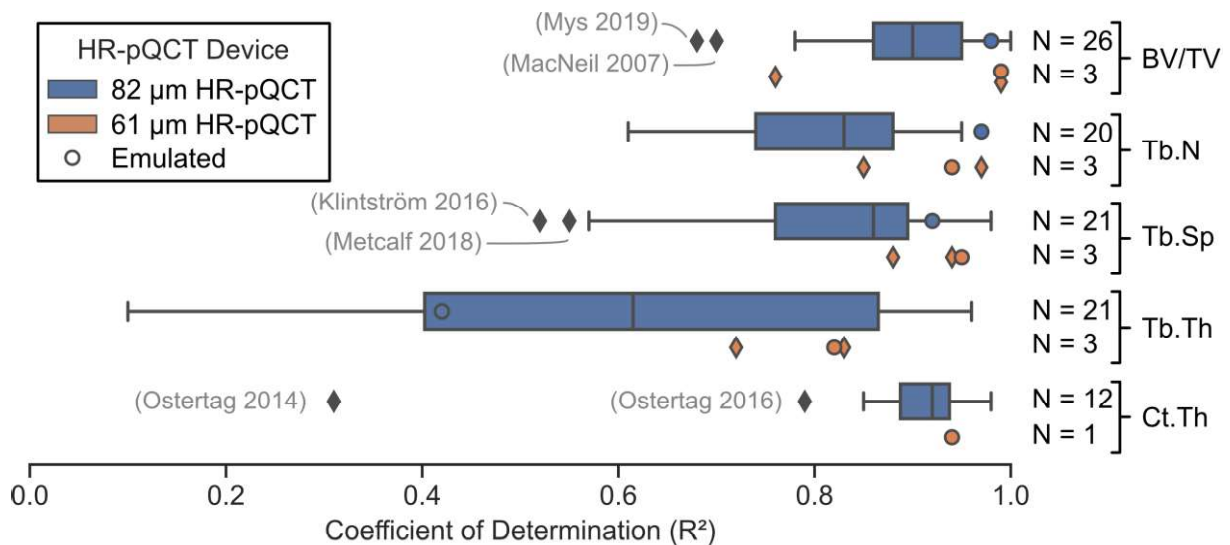


Fig. 2.1.3 Coefficient of determination for the most commonly reported morphometric parameters, which were reported in more than ten studies: bone volume fraction (BV/TV), trabecular number (Tb.N), trabecular separation (Tb.Sp), trabecular thickness (Tb.Th), and cortical thickness (Ct.Th). Emulated image data, such as 82 μm HR-pQCT data obtained from downsampled micro-CT or 61 μm HR-pQCT devices, is included both in the box plot and additionally highlighted by circles to identify any differences to parameters derived from 61 μm or 82 μm HR-pQCT images using the default manufacturer settings. In general, we observe that BV/TV can be captured well with both HR-pQCT resolutions while results for other parameters, such as Tb.Th, vary drastically across studies for 82 μm images and should be interpreted with caution.

2.1.4.2.1 Bone Volume Fraction

Bone volume fraction (BV/TV), the most commonly reported trabecular parameter, was captured well by both 82 μm and 61 μm HR-pQCT devices when compared to micro-CT ($R^2=0.91, 0.99$) (Figure 2.1.3). The lowest agreement with micro-CT was found for 82 μm HR-pQCT scans of the trapezium (Mys et al., 2019) ($R^2=0.68$) (Figure 2.1.3). Two approaches have been used to compute BV/TV for 82 μm HR-pQCT images; to date, the indirect density approach (Laib et al., 1998) has performed slightly better ($R^2=0.92, N=12$) than the direct segmentation-based approach ($R^2=0.88, N=14$) when comparing to BV/TV measured from the gold standard micro-CT. While both approaches agree well, neither is unassailable. The density-based approach suffers from inaccurate grey-values due to beam hardening effects, which disproportionately affect voxel intensities near the centre of the image. The segmentation-based approach suffers from partial volume effects, thus it is very sensitive to the chosen threshold (Varga and Zysset, 2009). The use of different threshold approaches on BV/TV yielded a range of R^2 values from 0.88 to 0.95 when images of the same samples were compared with micro-CT (Varga and Zysset, 2009). The highest agreement was found for microstructural segmentation using a Laplace-Hamming filter followed by a fixed threshold of 40% of the maximum grey value. These settings resulted in the most visually similar bone architecture and connectivity relative to gold standard micro-CT. The local adaptive and the fixed threshold segmentation approaches showed equal agreement when comparing 82 μm HR-pQCT with gold standard micro-CT ($R^2=0.97$) (Burghardt et al., 2007). Conversely, compared to gold standard micro-CT segmentation, the ARG segmentation method had a lower agreement than the manufacturer's default segmentation method ($R^2=0.86$ vs $R^2=0.94$) (Klintström et al., 2016).

Regression analysis between HR-pQCT and micro-CT BV/TV varied widely, with slopes ranging from 0.40 to 1.71 and intercepts ranging from -0.08 to 0.05 (Figure 2.1.4A). There was a trend for underestimation of BV/TV for studies using the indirect method, while studies using the direct method showed a trend for overestimation (Figure 2.1.4A, 2.1.4B). The two 61 μm HR-pQCT studies had a narrower range in slope and intercept than the 82 μm studies without clear over- or underestimation of BV/TV (Figure 2.1.4A). Overall, due to the even spread of slope and intercept values, no standardized calibration for BV/TV could be identified.

2.1.4.2.2 Trabecular Number

Trabecular number (Tb.N) agreed well with the values generated from micro-CT for 82 μm and 61 μm HR-pQCT ($R^2=0.83, 0.94$) (Figure 2.1.3). The strength of this agreement for 82 μm devices was affected by disease state and anatomical location. Lower agreement ($R^2=0.67$ to 0.81, depending on the computational method used) was observed for osteoporotic bone samples, which was attributed to an increased number of thin trabeculae that could not be assessed accurately with 82 μm HR-pQCT (Krause et al., 2014). An even lower agreement between HR-pQCT and micro-CT derived Tb.N ($R^2=0.61$) was observed in the calcaneus (Metcalf et al., 2018). The local adaptive threshold strategy improved direct agreements with gold standard micro-CT Tb.N for 82 μm HR-pQCT compared to the indirect manufacturer's approach ($R^2=0.85$ vs $R^2=0.70$) for femoral head samples (Burghardt et al., 2007). While the manufacturer's approach relies on estimated parameters such as an assumed compact tissue density (Laib and R uegsegger, 1999b), which might be patient specific and disease dependent, similar levels of agreement have been reported for radii using the manufacturer's indirect approach ($R^2=0.82$) (MacNeil and Boyd, 2007). Relative to gold standard micro-CT, the ARG segmentation based direct method showed lower agreement than the manufacturer's default approach for the radius ($R^2=0.66$ vs $R^2=0.81$) (Klintstr m et al., 2016). In total, these studies suggest that the selection of an appropriate segmentation approach may be heavily dependent on the studied anatomical site.

Regression analysis performed by several studies showed a range of slopes of 0.35 to 1.61 with intercepts ranging from -0.56 to 0.81 mm^{-1} (Figure 2.1.4C). Most studies showed a value-dependent over- and underestimation with the turning point being slightly above 1 mm^{-1} (Figure 2.1.4D). No clear agreement was observed for slope and intercept between the different studies, however, reduced variability was observed for the 61 μm compared to 82 μm studies (Figure 2.1.4C).

2.1.4.2.3 Trabecular Separation

Agreements similar to those reported for Tb.N were found for trabecular separation (Tb.Sp) ($R^2=0.87, 0.94$) (Figure 2.1.3). The range of agreement in Tb.Sp was also similar to that of Tb.N (Figure 2.1.3), as is expected since they are reciprocals. Interestingly, the trapezium study found Tb.Sp values from 82 μm images agreed better with micro-CT than Tb.Sp values from 61 μm images ($R^2=0.93$ vs $R^2=0.88$) (Mys et al., 2019). Some studies found better agreement for the indirect, manufacturer recommended method (Krause et al., 2014; MacNeil and Boyd, 2007),

which is based on grey-scale BV/TV and indirect Tb.N (Laib et al., 1998). Others found no difference between Tb.Sp measured using the indirect and direct segmentation based methods (Liu et al., 2010; Zhou et al., 2016). The direct local adaptive threshold method resulted in higher agreement than the indirect method ($R^2=0.93$ vs $R^2=0.85$) (Burghardt et al., 2007). One of the studies reporting higher agreement for indirect methods evaluated osteoporotic samples (Krause et al., 2014), however no osteoporotic samples were analysed by Burghardt and colleagues. Therefore, it is unclear whether the observed differences between direct and indirect methods are due to the segmentation method or the bone quality of the samples. The direct ARG segmentation method resulted in lower agreement than the indirect method ($R^2=0.52$ vs $R^2=0.64$) (Klintström et al., 2016) and is not recommended for computing Tb.Sp. Overall, the accuracy with which Tb.Sp can be assessed and the preferred method of calculation seems dependent on the size of the trabeculae relative to the voxel size.

The regression results from all studies resulted in slopes in the range of 0.28 to 3.64 with intercepts from -1.90 to 0.80 mm (Figure 2.1.4E). As with Tb.N, a turning point was observed for Tb.Sp slightly below 1 mm, which is roughly the inverse of the turning point for Tb.N (Figure 2.1.4F). Overall, no clear agreement was observable for the regression parameters, while reduced ranges for the regression parameters were observed for 61 μm images compared to 82 μm images (Figure 2.1.4E).

Interestingly, Tb.Sp was inconsistently defined as either trabecular spacing or trabecular separation. By definition, trabecular spacing refers to the distance between midlines of trabeculae, while trabecular separation describes the distance between edges of trabeculae. The source data for Tb.Sp, whether defined as separation or spacing, has not been modified herein.

2.1.4.2.4 Trabecular Thickness

For the commonly reported trabecular metrics, trabecular thickness (Tb.Th) performed the poorest ($R^2=0.59, 0.82$) (Figure 2.1.3). The direct measure of Tb.Th for 82 μm HR-pQCT images, either using fixed or local adaptive threshold segmentation, agreed better than or equally as well as the indirect method of the manufacturer ($R^2=0.80$ vs $R^2=0.55$), which is based on grey-scale BV/TV and indirect Tb.N (Laib et al., 1998). As an exception, one study reported better performance of the manufacturer's indirect method (MacNeil and Boyd, 2007). While they argued that the poor agreement found for the direct method from 82 μm HR-pQCT

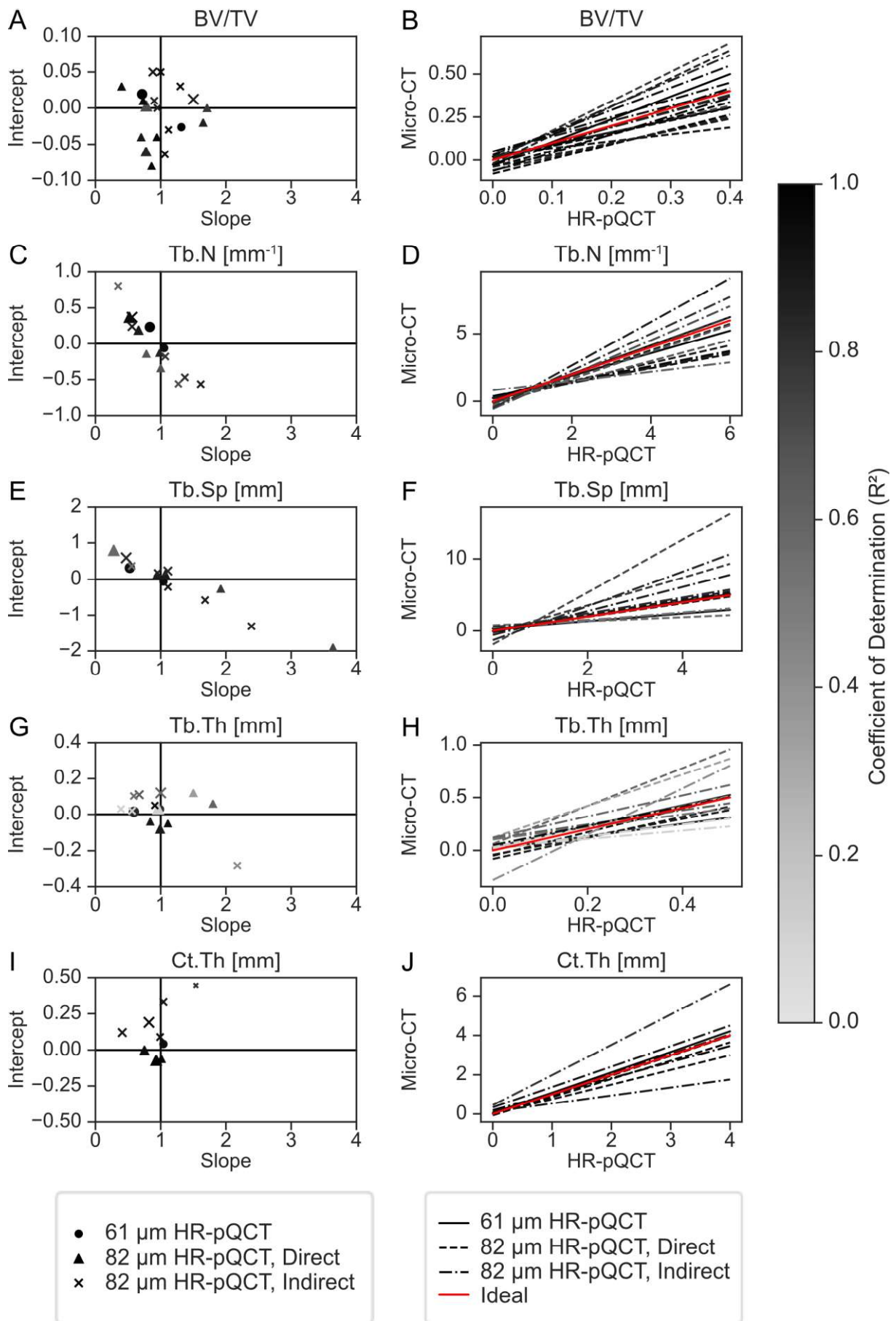


Fig. 2.1.4 Linear regression analysis shows large variation across studies in the most commonly reported morphometric parameters, which were reported in more than ten studies. (A, C, E, G, I) Scatter plots of slope vs intercept for 61 μm HR-pQCT and 82 μm HR-pQCT validated against micro-CT show variability between parameters. Results from 82 μm HR-pQCT studies are further divided into direct and indirect approaches for all parameters. Grey-levels indicate the corresponding coefficient of determination (R^2), while marker size indicates the voxel size of the micro-CT data used for comparison. (B, D, F, H, J) Linear regression lines from the same regression analyses are shown. The red, ideal line, indicates a perfect match between HR-pQCT and micro-CT and has slope of one and intercept of zero. In general, we observe that the parameters that have higher average coefficient of determination (e.g. BV/TV, Ct.Th) exhibit either under- or overestimation throughout the entire parameter space. In contrast, the parameter with the lowest coefficient of determination (e.g. Tb.Th) does not show a clear trend across studies. Furthermore, Tb.N and Tb.Sp have a consistent turning point between over- and underestimation across all studies.

($R^2=0.08$) could be a result of the inherent low resolution of the images, this did not seem to hinder the other studies assessed herein. This apparent contradiction may be a result of variation across studies in pre-processing of the image data for the direct method. Hence, future studies looking at Tb.Th should also look at the sensitivity to different pre-processing protocols.

Slopes ranged from 0.39 to 2.17 and intercepts ranged from -0.28 to 0.12 mm (Figure 2.1.4G). Since the indirect Tb.Th is a derived parameter, often represented as Tb.Th* or Tb.Th^d, dependent on grey-scale BV/TV and indirect Tb.N, which showed no clear trend in terms of over- or underestimation, the linear regressions found for indirect Tb.Th also did not show a clear over- or underestimation trend (Figure 2.1.4H). As with other parameters, 61 μm images showed reduced variability compared to the 82 μm images (Figure 2.1.4H).

2.1.4.2.5 Bone Surface to Bone Volume Ratio

Bone surface to bone volume ratio (BS/BV) was only included in three studies (Liu et al., 2010; MacNeil and Boyd, 2007; Zhou et al., 2016), all of which used 82 μm HR-pQCT and scans of radii or tibiae. Even though, all studies used the same method to compute BS/BV (Müller et al., 1994), two studies found good agreement ($R^2=0.83$) (Liu et al., 2010; Zhou et al., 2016) and one study found poor agreement ($R^2=0.18$) with BS/BV measured using micro-CT (MacNeil and Boyd, 2007). No alternative segmentation method was tested nor were any distinct patient populations included in these studies.

2.1.4.2.6 Structural Model Index

Structural model index (SMI) provides a metric for how plate- or rod-like a trabecular architecture is. However, one study cautioned against this interpretation of SMI, as a loss in BV/TV has been shown to shift SMI to indicate a more rod-like bone structure (Salmon et al., 2015). All five studies reporting SMI utilized 82 μm HR-pQCT images. The same study that found poor agreement for BS/BV found poor agreement for SMI ($R^2=0.08$) (MacNeil and Boyd, 2007). However, the other studies found moderate agreement ($R^2=0.78$) (Burghardt et al., 2007; Krause et al., 2014; Liu et al., 2010; Zhou et al., 2016) with the highest agreement found for the local adaptive threshold ($R^2=0.92$) (Burghardt et al., 2007). Since the study with osteoporotic bone samples found lower agreement between SMI from micro-CT and 82 μm HR-pQCT data (Krause et al., 2014), the accuracy for SMI may depend on the disease status or bone quality of the scanned patient. Although the local adaptive threshold procedure produced results most similar to micro-CT SMI, this pre-processing method was only tested on non-osteoporotic bone samples. Further validation should be performed before this method can be recommended unconditionally as a pre-processing step for the computation of SMI.

2.1.4.2.7 Connectivity Density and the Degree of Anisotropy

Connectivity density (Conn.D) is a computational measure of the inter-connectivity among trabeculae and can be an indicator of the mechanical strength of the trabecular architecture (Odgaard and Gundersen, 1993). Degree of anisotropy (DA) is a measure that describes the degree to which trabeculae are oriented along a common axis (Harrigan and Mann, 1985). For Conn.D and DA only slight and moderate agreement was found for 82 μm HR-pQCT, respectively ($R^2=0.50$ for Conn.D, $R^2=0.62$ for DA) (Burghardt et al., 2007; Krause et al., 2014; Liu et al., 2010; MacNeil and Boyd, 2007; Zhou et al., 2016). No comparisons between 61 μm HR-pQCT and micro-CT have been performed for Conn.D and DA.

2.1.4.2.8 Less Common Parameters

Trabecular nodes (Tb.Nd) is the count of trabecular intersections. Trabecular termini (Tb.Tm) is the number of free ends in the trabecular structure. Only one study evaluated Tb.Nd and Tb.Tm and found slight to moderate agreement ($R^2=0.49$ to 0.74) using the ARG segmentation algorithm for 82 μm HR-pQCT (Klintström et al., 2016). The standard manufacturer's segmentation resulted in poor agreement for these parameters ($R^2=0.07$).

Mean intercept length (MIL), a measure of micro-architectural anisotropy (fabric), agreed well between 82 μm HR-pQCT and micro-CT, even when different filtering techniques for pre-

processing were applied ($R^2=0.97$ vs $R^2=0.99$ for Laplace Hamming and Gaussian filtering, respectively) (Varga and Zysset, 2009). Using 82 μm HR-pQCT, Hosseini and colleagues showed that the manufacturer's function to evaluate DA (TRI (Laib et al., 2000)) and their own open source implementation, mean surface length (MSL), both agreed well with MIL computed on micro-CT ($R^2>0.98$ for all) (Hosseini et al., 2017). Here, the same threshold (320 mg HA/cm³) was used for micro-CT and HR-pQCT, demonstrating that fabric allows for more direct comparison between the two imaging modalities. However, while other studies looking at DA assessed the entire trabecular compartment, Hosseini and colleagues used three hand-selected cubic regions of interest (ROIs) (edge length 6mm) per sample. This ROI-dependent precision ($R^2=0.99$ vs $R^2=0.62$) shows the importance of standardizing ROI selection to ensure cross-study comparability.

Individual trabecula segmentation (ITS) is another approach, implemented in commercial software, to extract morphometric information from a CT image that differentiates between plate- and rod-like structures (Liu et al., 2011). Agreement between plate parameters was generally higher than for rod parameters. The declared conflict of interest for ITS by Liu and colleagues calls for independent research groups to substantiate these results.

2.1.4.2.9 Effects of Segmentation on Trabecular Parameters

The effect of custom segmentation methods on trabecular parameters remains unclear. The local adaptive threshold segmentation method performed equivalently or better for the analysis of morphological parameters of trabecular bone, e.g. Tb.N, Tb.Sp, etc., when compared to the global threshold segmentation method (Burghardt et al., 2007). Similarly, MSL was extracted accurately with a fixed threshold segmentation (Hosseini et al., 2017). However, others observed that the scanner manufacturer's compartmental segmentation and indirect grey-scale method yielded better estimates of BV/TV and most morphometric indices, respectively, compared to micro-CT based or their proposed segmentation approaches (Klintström et al., 2016; Varga and Zysset, 2009). Simultaneously, their alternative segmentation approaches outperformed the manufacturer's approach when assessing less common parameters, such as Tb.Tm or MIL.

2.1.4.3 Cortical Parameters

Due to the low resolution of HR-pQCT images and their increased noise, segmenting the cortex accurately, especially identifying pores and the correct endosteal surface, can be challenging.

Consequently, parameters derived from these segmentations require a thorough validation against a high-quality gold standard such as micro-CT.

2.1.4.3.1 Cortical Thickness

Agreement between micro-CT and HR-pQCT measured cortical thickness (Ct.Th), the most commonly reported cortical parameter, was higher than most other cortical parameters for both device resolutions ($R^2=0.92, 0.94$) (Liu et al., 2010; MacNeil and Boyd, 2007; Nishiyama et al., 2010; Ostertag et al., 2016; Tjong et al., 2012; Zhou et al., 2016) (Figure 2.1.3). The dual threshold segmentation algorithm improved agreement in Ct.Th for 82 μm images in two studies compared to Ct.Th derived from images using the indirect method in the scanner manufacturer's software. In one study, both had good agreement ($R^2=0.98$ vs $R^2=0.90$) (Nishiyama et al., 2010), while in the other study, good agreement was observed in directly computed Ct.Th derived from dual threshold segmented images and only moderate agreement in Ct.Th derived from images using the indirect method in the scanner manufacturer's software ($R^2=0.85$ vs $R^2=0.79$) (Ostertag et al., 2016). The manufacturer's indirect approach for Ct.Th uses an assumed relation of cortical volume over outer bone surface (MacNeil and Boyd, 2007), which could explain the lower agreement observed for the indirect compared to the direct method.

Regression analyses found that Ct.Th computed from HR-pQCT matched Ct.Th from gold standard micro-CT much better if the direct segmentation-based method was used (Figure 2.1.4I, 2.1.4J). Slopes ranged from 0.41 to 1.54 and intercepts ranged from -0.07 to 0.45 mm (Figure 2.1.4I). Only one 61 μm study assessed Ct.Th (Tjong et al., 2012) and reported a good match with micro-CT (Figure 2.1.4I), but it should be noted that Tjong and colleagues used an 82 μm HR-pQCT device in non-patient mode with a resolution of 41 μm . Since this resolution is higher than the resolution of 61 μm HR-pQCT devices, further studies should clarify the effect this difference in resolution has on the accuracy of Ct.Th measured using 61 μm HR-pQCT.

2.1.4.3.2 Cortical Bone Mineral Density

One study reported slight agreement ($R^2=0.59$) with cortical bone mineral density (Ct.BMD) for an 82 μm HR-pQCT device running in non-patient mode (41 μm resolution) and a weaker agreement for the normal 82 μm HR-pQCT resolution ($R^2=0.44$) (Tjong et al., 2012). No direct comparison between a 61 μm HR-pQCT device and micro-CT has been performed thus far.

Given that only one 82 μm HR-pQCT study has evaluated Ct.BMD relative to micro-CT, the accuracy with which HR-pQCT can measure this parameter remains unclear.

2.1.4.3.3 Cortical Porosity

Cortical porosity (Ct.Po) agreed well for both HR-pQCT resolutions ($R^2=0.85, 0.84$) (Jorgenson et al., 2015; Nishiyama et al., 2010; Ostertag et al., 2016, 2014; Soltan et al., 2019; Tjong et al., 2012). For 82 μm HR-pQCT, the direct 3D evaluation yielded better agreement than the indirect density derived evaluation ($R^2=0.85$ vs $R^2=0.75$). Using StrAx1.0, Ct.Po measurements agreed well ($R^2=0.87$ to 0.98) (Zebaze et al., 2013), while with the direct approach, the number of detected pores agreed poorly ($R^2=0.27$) (Nishiyama et al., 2010). This discrepancy could be due to the inability of HR-pQCT to detect the smallest pores, thus reducing the pore count while not greatly affecting the measured porosity.

2.1.4.3.4 Less Common Parameters

StrAx1.0 assesses cortical area using three parameters, total cross-sectional area, the compact-appearing cortex area, and the transitional zone area (Figure 2.1.1C). These parameters, measured using 82 μm HR-pQCT data, agreed well with their micro-CT measured equivalents ($R^2=0.99, 0.98, 0.95$, respectively). However, agreement with the scanner manufacturer's software was poor ($R^2=0.32$) (Zebaze et al., 2013).

The parameter of matrix mineral density (MMD) was introduced to address that Ct.BMD does not differentiate between a reduced mineralized bone matrix volume and reduced mineralisation of the matrix. Moderate agreement was found between MMD computed with StrAx1.0 and with the manufacturer's software for micro-CT data ($R^2=0.76$) (Chiang et al., 2018).

While cortical interruptions can be used as a predictor of progressing rheumatoid arthritis (Scott, 2003), only moderate inter-rater reliability was found for the visual detection of cortical interruptions with diameters of less than 0.5 mm using 82 μm HR-pQCT and micro-CT, despite the extensive training and prior experience of the operators (Scharmga et al., 2016). To avoid inter-rater variability, an algorithm was proposed for detecting cortical interruptions (Peters et al., 2017), which modified the default segmentation of the manufacturer. This algorithm identified similar cortical interruptions for both HR-pQCT and micro-CT images and performed best for interruptions with a diameter of at least 0.16 mm for 82 μm HR-pQCT (inter-class correlation coefficient=0.91).

2.1.4.3.5 Effects of Segmentation on Cortical Parameters

Only a few studies have looked at the effects of segmentation on cortical parameters. The thickness-based separation segmentation approach achieved sub-voxel precision on some of the cortical morphometric parameters (Ang et al., 2020), while the use of manually-corrected segmentations resulted in inter-observer reproducibility errors of 3.7% for Ct.Th, 5.3% for Ct.BMD, and 6.2% for cortical porosity (Ostertag et al., 2014). Further studies should investigate the replacement of manual hand-drawn contours and manually-corrected segmentations with automatic approaches to eliminate such uncertainties and enable more direct cross-study and cross-centre comparisons.

2.1.4.4 Discussion

Few 61 μm HR-pQCT studies have validated morphometrics against micro-CT and the accuracy of 61 μm HR-pQCT cannot be assessed from comparisons of 82 μm and 61 μm HR-pQCT without a proper gold standard (Agarwal et al., 2016; Manske et al., 2017, 2015), especially for parameters which do not agree well between 82 μm HR-pQCT and micro-CT. Importantly, a clear difference in morphological parameter agreement relative to analysis type was found between 82 and 41 μm HR-pQCT (Tjong et al., 2012); here, indirect methods resulted in a better agreement with the parameters derived from micro-CT for 82 μm HR-pQCT, while direct methods resulted in better agreement for 41 μm HR-pQCT. Given that 61 μm HR-pQCT is between these two resolutions, it is not obvious which method best analyses 61 μm HR-pQCT images.

There are a number of issues which may hinder the development of a generalized validation of HR-pQCT relative to micro-CT. Perhaps the largest issue with the comparison is the lack of a standardized approach for pre-processing micro-CT data. The plethora of pre-processing steps described for micro-CT might explain the systematic deviations between studies, which are clearly visible in the regression results for all morphometric parameters (Figure 2.1.4).

Micro-CT resolution was also not standardized across studies, which is especially relevant for trabecular indices, such as Tb.Th, which have a strong dependence on image resolution (Kim et al., 2004; Kothari et al., 1998; Müller et al., 1996; Sode et al., 2008). Müller reported that resolutions of less than 10 μm should be used to obtain accurate results (Müller, 2003). Yet, we observed a wide variety of micro-CT resolutions, with some exceeding 30 μm (Table 2.1.2). Since the accuracy of HR-pQCT is not expected to match that of micro-CT, the improved

accuracy of using $< 10 \mu\text{m}$ micro-CT for comparison is unlikely to affect the validation of HR-pQCT. Importantly, we did not observe dependence on scanner resolution for morphologic parameters (Figure 2.1.3). However, studies looking at very thin, i.e. osteoporotic, micro-architectures used resolutions below $20 \mu\text{m}$ for validation. The low resolutions of HR-pQCT devices may require the use of various segmentation methods to yield accurate results for all morphological parameters (Varga and Zysset, 2009).

The accuracy for trabecular parameters varied by anatomical site, with the calcaneus having less accurate results with HR-pQCT compared to the radius or the tibia. This discrepancy may be a result of the morphometric analysis methods being fine-tuned for the distal radius and tibia, necessitating the development of a more universal or site-specific evaluation function.

Different implementations have been introduced to compute the various morphometric indices; unfortunately, only the groups that proposed these implementations have reported on their accuracy. The lack of validation by independent research groups limits the amount of data, relative to anatomical site, disease status, etc., that is available for cross-comparison. This lack of comparative data, especially when existing datasets are overly homogenous, can lead to conflicts of interest when only internally-produced data is used for the validation of proprietary algorithms or derived commercial products. The introduction and use of a publicly available database of diverse validation data could help to avoid such issues in the future.

Finally, only static cortical and trabecular parameters have been evaluated; dynamic measures for HR-pQCT, such as bone formation and resorption parameters, have yet to be investigated (Schulte et al., 2011). Since longitudinal micro-CT data cannot be generated for patients, other validation methods are required to assess these dynamic parameters. *In silico* models can now provide both realistic bone structures and simulated time-lapse data. By coupling these models with simulated image artefacts, such as noise, a wide variety of input data could be generated to evaluate dynamic parameters that extend what is possible with a single micro-CT or HR-pQCT scan.

2.1.5 Finite Element Analysis (FEA)

Shortly after the introduction of micro-CT, researchers began investigating the use of imaging to not only quantify bone structure but also evaluate bone strength through FEA (Rüegsegger

et al., 1996; van Rietbergen et al., 1995). Importantly, non-destructive micro-CT FEA methods have since been validated against experimentally derived measures of strength and failure (Chen et al., 2017; Hambli, 2013). With the introduction of HR-pQCT for patient imaging in the clinic, it became evident that measures of strength and failure prediction from HR-pQCT FEA may aid in patient diagnosis and treatment. However, the application of FEA methods to lower resolution images requires thorough validation before these benefits can be realized.

To date, 11 studies have utilized FEA in their analysis of 82 μm HR-pQCT relative to micro-CT; 61 μm HR-pQCT FEA has yet to be evaluated relative to micro-CT FEA. Of these 11 studies, eight included a comparison of mechanical measures between HR-pQCT FEA and micro-CT FEA, three of which also included comparisons of mechanical measures to experimentally derived measures, three utilized micro-CT FEA to evaluate the ability of HR-pQCT image morphometrics to predict mechanical parameters, and one varied the voxel size of the input micro-CT data to isolate its effect on the FEA (Table 2.1.3). Seven studies used hexahedral elements and an isotropic linear-elastic material model, three used hexahedral elements and an elastic-plastic model with 5% or 50% reduction in elastic modulus after 0.33% tensile strain or 0.81% compressive strain, and one study also used quadratic pentahedral and tetrahedral elements in a homogenised FEA with an orthotropic elasticity tensor (Table 2.1.3). Similar to the Section 4, data from previous studies is presented as $\text{slope} * \text{value}_{\text{HR-pQCT}} + \text{intercept} = \text{value}_{\text{micro-CT}}$. Due to the lack of consistent reporting, all R^2 values are specific to a single study.

2.1.5.1 Bulk Mechanical Properties

In studies comparing the mechanical response of FEA models from HR-pQCT to those from micro-CT, increased bone strength and stiffness estimations were observed with increasing voxel size (Alsayednoor et al., 2018; Cohen et al., 2010; Liu et al., 2013, 2011, 2010; MacNeil and Boyd, 2007; Wang et al., 2019; Zhou et al., 2019). Specifically, one study found reaction force was overestimated by HR-pQCT ($y = -5 + 0.42x$ N; $R^2 = 0.73$) (MacNeil and Boyd, 2007). However, another study found that both reaction force and failure load estimates depend on the chosen threshold, such that a BV/TV-matched threshold led to significant underestimations, while a geometry-preserving threshold led to general overestimations (Alsayednoor et al., 2018). Due to the small sample size ($N=5$), no regressions were performed, and the variability between specimens was notably large (Alsayednoor et al., 2018). Regarding analysis type,

strength was underestimated using linear HR-pQCT FEA for the radius ($y=-0.92+1.7x$ kN; $R^2=0.92$) and for the tibia ($y=-0.39+1.5x$ kN; $R^2=0.96$) and overestimated using nonlinear HR-pQCT FEA for the radius ($y=-0.31+0.84x$ kN; $R^2=0.91$) and for the tibia ($y=0.77+0.76x$ kN; $R^2=0.95$) (Zhou et al., 2016).

Stiffness was also overestimated through HR-pQCT FEA. This was observed in the evaluation of trabecular bone stiffness ($y=-16.6+0.69x$ kN/mm; $R^2=0.90$ (Liu et al., 2010)) and whole bone stiffness ($y=-24.6+0.87x$ kN/mm; $R^2=0.96$ (Liu et al., 2010) and $y=4.3+0.86x$ kN/mm; $R^2=0.94$ (Zhou et al., 2016)) for the tibia. Similar results were found for stiffness of the radius (Zhou et al., 2016), but are not reported herein. While the slopes for whole bone stiffness of the tibia were similar between the two studies, the intercept

Table 2.1.3 Parameters of HR-pQCT and micro-CT finite element analyses used to validate the use of HR-pQCT.

Study	Material Model	Element Type (Voxel Size)	Micro-CT Geometry ^a	HR-pQCT Geometry ^a
(MacNeil and Boyd, 2007)	Isotropic, Linear Elastic (E=10 GPa, $\nu=0.3$)	Hexahedral (19 $\mu\text{m}/82 \mu\text{m}$)	164x164x164 (Radius, 3.12mm)	38x38x38 (Radius, 3.12 mm)
(Liu et al., 2010)	Isotropic, Linear Elastic (E=15 GPa, $\nu=0.3$)	Hexahedral (40 $\mu\text{m}/82 \mu\text{m}$)	143x143x143 (Tibia, 5.72 mm) 10 mm CS (Tibia)	70x70x70 (Tibia, 5.74 mm) 9.02 mm CS (Tibia)
(Alsayednoor et al., 2018)	Isotropic, Linear Elastic ^b (E=10 GPa, $\nu=0.33$)	Hexahedral (17.41 $\mu\text{m}/82 \mu\text{m}$)	301x301x301 (Calcaneus, 5.24 mm)	65x65x65 ^c (Calcaneus, 5.33mm)
(Cohen et al., 2010)	Isotropic, Linear Elastic (E=15 GPa, $\nu=0.3$)	Hexahedral (8 $\mu\text{m}/82 \mu\text{m}$)	640x640x300 (Iliac Crest, 2.40mm)	70x70x70 (Radius, 5.74 mm) 110x110x110 (Tibia, 9.02 mm)

Study	Material Model	Element Type (Voxel Size)	Micro-CT Geometry ^a	HR-pQCT Geometry ^a
(Liu et al., 2013)	Isotropic, Elastic-Plastic (E _{HRR-pQCT} =16.59 (5%) GPa, E _{μCT} =10.43 (5%) GPa, ν=0.3)	Hexahedral (25 μm/ 82 μm)	230x230x230 (Radius/ Tibia, 5.75 mm)	70x70x70 (Radius, 5.74 mm) 110x110x110 (Tibia, 9.02 mm)
	(E=39.62 (5%) GPa, ν=0.3)	PR (--/ 82 μm)	--	9.02 mm CS (Radius/ Tibia)
(Zhou et al., 2016)	Isotropic, Elastic-Plastic (E=15 (50%) GPa, ν=0.3)	Hexahedral (37 μm/ 82 μm)	Approx. 9 mm CS (Radius/ Tibia)	9.02 mm CS (Radius/ Tibia)
(Wang et al., 2019)	Isotropic, Elastic-Plastic (E=15 (50%) GPa, G=7 GPa ^e , ν=0.3)	Hexahedral (37 μm/ 82 μm)	9.02 mm CS (Radius/ Tibia)	9.02 mm CS (Radius/ Tibia)
		PR with Hexahedral Cortex (--/ 82 μm)	--	9.02 mm CS (Radius/ Tibia)
(Pahr et al., 2012)	Isotropic, Linear Elastic ^b (E _{BMD} =21.96(BV/TV) ^{1.7} GPa or E _{SEG} =8.78 GPa, ν=0.3 ^b)	Hexahedral (--/ 82 μm)	--	19.33± 2.15 mm CS (Vertebra) ^b
(Pahr et al., 2012)	Homogenized, Orthotropic Elastic (E=12 GPa, G=3.913 GPa, ν=0.249)	Pentahedral Cortex/ Tetrahedral Trabecular with Fabric (--/ 82 μm)	--	19.33± 2.15 mm CS (Vertebra) ^b
(Christen et al., 2016)	Isotropic, Linear Elastic ^b (E=6.8 GPa, ν=0.3)	Hexahedral (25 μm/ --) ^d	9.02 mm CS (Radius)	--
(Liu et al., 2011)	Isotropic, Linear Elastic ^b (E=15 GPa, ν=0.3)	Hexahedral (40 μm/ --)	143x143x143 (Tibia, 5.72 mm)	--

2.1 Validation of HR-pQCT against micro-CT for morphometric and biomechanical analyses

Study	Material Model	Element Type (Voxel Size)	Micro-CT Geometry ^a	HR-pQCT Geometry ^a	
(Klintström et al., 2016)	Isotropic, Elastic ^b (E=12 GPa, v=0.3)	Linear (20µm/ --)	Hexahedral (20µm/ --)	260x260x260 (Radius, 5.20mm)	--

Study	Loading Conditions	FE Solver	Study Measures	Purpose
(MacNeil and Boyd, 2007)	1.0% Comp	Custom (Su et al., 2007)	Reaction Force, Stress, SED	Validation (Micro-CT)
(Liu et al., 2010)	SV: Combi; CS: 1.0% Comp (van Rietbergen, 1996)	Olympus (Adams, 2002; Adams et al., 2004; Laib, 1997)	SV: Elastic and Shear Moduli; CS: Trabecular and Total Stiffness	Validation (Micro-CT)
(Alsayednoor et al., 2018)	0.13% Comp	Abaqus	Reaction Force, von Mises Stress, Failure Load	Validation (Micro-CT)
(Cohen et al., 2010)	Combi (Hollister et al., 1994; van Rietbergen, 1996)	Custom (Arbenz et al., 2008)	Elastic Moduli	Comparison (Micro-CT)
(Liu et al., 2013)	1.0% Comp ^a	Olympus (Adams et al., 2004)	Elastic Moduli, Yield Strength	Validation (Micro-CT), Fracture Risk Prediction
	1.0% Comp ^a	Abaqus		

Study	Loading Conditions	FE Solver	Study Measures	Purpose
(Zhou et al., 2016)	1.2% Comp	FEAP (Adams et al., 2004)	Total Stiffness, Yield Strength, Morph	Validation (Exp, Micro-CT)
(Wang et al., 2019)	1.2% Comp 1.2% Comp	FEAP (Adams et al., 2004) Abaqus	Total Stiffness, Total Yield Strength	Validation (Exp, Micro-CT), Fracture Risk Prediction
(Pahr et al., 2012)	Exp, (Dall'Ara et al., 2010) Comp	ParFE (Arbenz et al., 2008)	Elastic Modulus	Validation (Exp)
(Pahr et al., 2012)	1.2% Comp	Abaqus	Orthotropic Stiffness	Validation (Exp)
(Christen et al., 2016)	Unit Comp (3 uniaxial directions)	ParOSol (Flaig, 2012)	Load Estimate Error	Voxel Size Dependency
(Liu et al., 2011)	Combi (van Rietbergen, 1996)	FEAP (Adams et al., 2004)	Elastic Modulus, Morph	Micro-FEA Evaluation of Morph
(Klintström et al., 2016)	Combi (Chevalier et al., 2008)	Abaqus	Stiffness, Shear Moduli, Morph	Micro-FEA Evaluation of Morph

^aSubvolume sample geometry is presented as voxel dimensions (bone source, subvolume height) and bone cross-section sample geometry is presented as cross-section height (bone source)

^bAssumed value, not available in text

^cData was evaluated using three methods: resampling to micro-CT resolution, thresholded to match BV/TV of micro-CT images, and thresholded to match fractal structure of micro-CT images

^dData was resampled to 50, 61, 75, 82, 100, 125, and 150 μm for analysis at multiple voxel sizes

^eShear modulus, G, only used in PR model

Other terms: SV, sub-volume of bone isolated from the trabecular compartment; PR, plate-rod geometry consisting of 2-node rods and 3-node shell plates; CS, bone cross-section; Combi, a combination of six micro-FEA analyses including three uniaxial compressions and three uniaxial shear; Comp, uniaxial compression along the long axis of the bone; Exp, experimental; FEAP, Finite Element Analysis Program; SED, strain energy density; Morph, morphometrics derived from HR-pQCT images.

varied in both sign and magnitude, possibly due to the use of different material models (linear elastic vs elastic-plastic, respectively).

With respect to tissue-level properties, strain energy density was overestimated by HR-pQCT ($y=0.004+0.42x$ J/mm³; $R^2=0.50$) (MacNeil and Boyd, 2007). Two studies presented conflicting results in their comparison of average von Mises stress between micro-CT and HR-pQCT FEA. In one study, von Mises stress was overestimated by HR-pQCT ($y=2.5+0.46x$ MPa; $R^2=0.51$) (MacNeil and Boyd, 2007), while another study found von Mises stress from HR-pQCT-based models was either equivalent or underestimated depending on the threshold used; however, no regression or quantification was performed in this second study (Alsayednoor et al., 2018).

The study which explicitly evaluated voxel size using micro-CT images found increased error in estimated loading with increased voxel size (Christen et al., 2016). Similarly, the use of a BV/TV-matched threshold on downsampled micro-CT images resulted in slight underestimations of failure load and significant underestimations of apparent stiffness and von Mises stresses in comparison to native resolution (Alsayednoor et al., 2018). In contrast, a third study noted that continuum parameters, e.g. reaction force and von Mises stress, correlated better with results from micro-CT FEA than tissue-level parameters, e.g. strain energy density (MacNeil and Boyd, 2007). The inconsistencies indicate that further research is warranted to elucidate the true effect of both voxel size and imaging modality on FEA-derived mechanical properties.

2.1.5.2 Tissue Mechanical Properties

Several studies utilized both HR-pQCT and micro-CT FEA to assess isotropic or orthotropic elastic moduli of a cubic trabecular bone section; herein, elastic moduli were overestimated

when HR-pQCT images were utilized (Cohen et al., 2010; Liu et al., 2013, 2010). This evaluation method was first used to compare images acquired from different anatomical locations and found HR-pQCT-derived apparent Young's moduli to be poor predictors ($R^2 < 0.19$) of micro-CT values (Cohen et al., 2010). Specifically, when compared to micro-CT values from the iliac crest, mean apparent Young's moduli in the longitudinal direction was overestimated by HR-pQCT by an average of 46% (305 MPa) and 54% (358 MPa) for the tibia and radius, respectively. Due to the use of varied anatomical sites, including both weight-bearing and non-weight-bearing samples, it was difficult to draw conclusions about the specific effect of using micro-CT vs HR-pQCT (Cohen et al., 2010). The second study to evaluate the apparent elastic moduli from HR-pQCT and micro-CT FEA found an overestimation of elastic moduli (longitudinal, $y = -403 + 0.83x$ MPa; $R^2 = 0.92$), but also measured average BV/TV to be 0.25 for HR-pQCT and 0.14 for micro-CT which may help to explain these differences (Liu et al., 2010).

The use of varied material models and properties for micro-CT and HR-pQCT FEA models eliminated significant differences in apparent elastic moduli (relationship not quantified) (Liu et al., 2013; Pahr et al., 2012). For one such study, material properties derived on a separate cadaveric cohort were applied to a cohort of 60 patients, enabling the differentiation between patients with and without previous vertebral fractures (Liu et al., 2013). Interestingly, experimentally derived apparent stiffness was relatively well predicted using both the voxel-based BMD ($y = -0.02 + 1.03x$ GPa, $R^2 = 0.86$) and binarised ($y = 0.07 + 0.88x$ GPa, $R^2 = 0.84$) FEA models (Pahr et al., 2012).

2.1.5.3 Non-voxel-based FEA

In an effort to reduce the computational time associated with micro-FEA, two alternative methods were introduced for HR-pQCT FEA and validated against micro-CT FEA and experimental findings (Liu et al., 2013; Pahr et al., 2012; Wang et al., 2019). The first of these methods included a fabric-based homogenization method which was applied to vertebral samples, calibrated using micro-CT FEA, and evaluated against experimental measures (Pahr et al., 2012). This method was more computationally efficient and was able to predict apparent stiffness relatively well in comparison to experimental measures using BMD-based ($y = 0.06 + 1.01x$ GPa, $R^2 = 0.75$) and binarised ($y = -0.04 + 1.07x$ GPa, $R^2 = 0.86$) model inputs (Pahr et al., 2012). The second method was a result of a geometry simplification process, i.e. ITS,

which resulted in a plate-rod (PR) geometry consisting of 2-node rods and 3-node shell plates (Liu et al., 2011). PR FEA resulted in similar predictions of elastic modulus ($y=2.42+0.79x$ MPa; $R^2=0.91$) and yield strength after initial adjustment of material properties ($y=2.36+0.72x$ MPa; $R^2=0.86$) relative to voxel-based micro-CT FEA (Liu et al., 2013). A follow-on study utilized this method in the evaluation of stiffness and yield strength predictions against those from both micro-CT FEA and mechanical testing (Wang et al., 2019). Here, stiffness from HR-pQCT PR FEA was a good estimate for voxel-based micro-CT FEA ($y=-6.63+1.04x$ kN/mm; $R^2=0.94$) and values from mechanical testing ($y=-26.5+1.06x$ kN/mm; $R^2=0.88$) in a pooled dataset from the radius and tibia. Yield strength from HR-pQCT PR FEA was also a good estimate of voxel-based micro-CT FEA ($y=-1.38+1.16x$ kN; $R^2=0.95$), but overestimated yield strength ($y=-1.17+1.30x$ kN; $R^2=0.94$) for the same pooled dataset.

2.1.5.4 Morphological Assessments

BV/TV was found to be an independent predictor of whole bone stiffness and apparent elastic moduli ($R^2=0.49$ to 0.74 for HR-pQCT and $R^2=0.65$ to 0.82 for micro-CT) (Liu et al., 2010). Interestingly, studies evaluating HR-pQCT morphometrics relative to micro-CT FEA also found strong agreement of micro-CT FEA mechanical parameters with both BV/TV and Tb.Tm (Klintström et al., 2016); with total bone area (Tt.Ar), BMD, Ct.Th, and Tb.Th (Zhou et al., 2016); and with ITS-based metrics of trabecular plates, orientation, and structure (Liu et al., 2011). Therefore, while BV/TV has an inherent influence on the mechanical results from FEA due to its characterization of the anatomy, several other morphologic parameters appear to also be relevant to the prediction of mechanical properties for HR-pQCT imaged anatomy.

2.1.5.5 Discussion

Generally, the use of material models and material properties has been inconsistent across HR-pQCT FEA investigations. This holds true even for the gold standard of micro-CT FEA. Additionally, the mechanical properties and metrics measured by various study groups is inconsistent. Combined, these factors resulted in difficulty in categorizing and comparing results between studies and investigating trends in observed differences. Specifically, the most commonly reported parameter across studies was orthotropic elastic moduli, however these values have only been reported in three studies by the same research group, including one study which compared different anatomical sites.

HR-pQCT FEA models clearly tend to overestimate bone strength and elastic modulus when applied without additional calibration or material property adjustments. While the results from 61 μm HR-pQCT FEA have yet to be compared to micro-CT FEA, the overestimation of reaction force and failure load observed with 82 μm HR-pQCT FEA seems to be reduced when directly comparing 61 μm to 82 μm HR-pQCT FEA (Whittier et al., 2018). This difference in mechanical properties may be resultant of inherent differences in BV/TV and trabecular structure preservation between the two imaging modalities (Liu et al., 2010). However, one study investigated the effect of using either BV/TV-matched or geometry-preserving thresholds and found that neither method resulted in similar values to those obtained from micro-CT FEA for any of the mechanical measures evaluated (Alsayednoor et al., 2018). Despite this, mechanical properties from HR-pQCT FEA are still highly correlated with those from both micro-CT FEA and experimental findings, indicating that a correction may be possible.

Given the current state of the literature, HR-pQCT FEA may not be a feasible solution for accurately quantifying mechanical properties, but instead an excellent option for use in comparative studies. Accordingly, two studies have shown that HR-pQCT FEA models were able to distinguish between patients with and without previous radius fractures, indicating a potential clinical application of these models in the future (Liu et al., 2013; Wang et al., 2019). While comparative studies are sufficient for many clinical applications, the availability of diverse datasets, which include experimentally derived mechanical properties and both micro-CT and HR-pQCT images, would provide the necessary basis for future validation of the quantification of mechanical properties using HR-pQCT, leading to patient-specific applications of HR-pQCT FEA.

2.1.6 Discussion

The aim of this review was to provide an overview of the current status of computational method validation for HR-pQCT with respect to gold standard micro-CT and of the limited standardization across the field. We found that there is good agreement between HR-pQCT and micro-CT for a variety of morphometric or mechanical parameters. Notably, of the most commonly used morphometric parameters, BV/TV and Ct.Th had the highest agreement. While most segmentation methods are calibrated to yield accurate BV/TV values, Ct.Th is insensitive to partial volume effects, noise, and other imaging artefacts, such as movement artefacts, since

it has a lower ratio of surface voxels to cortex thickness. On the contrary, trabecular parameters, such as Tb.Th, showed weaker agreement with micro-CT, likely since trabeculae are only a few voxels thick in HR-pQCT images. We also observed moderate to good agreement for samples representing diseased, e.g. osteoporotic, bone.

However, no segmentation method performed equally well for all computational applications. While BV/TV was most precisely measured without microstructural segmentation (i.e. grey-scale BV/TV) for 82 μm HR-pQCT, trabecular parameters, such as Tb.N and Tb.Sp, showed higher precision when using a fixed or local adaptive threshold approach. This was especially true for anatomical sites other than the radius and tibia. In contrast, using a threshold which matched morphological parameters resulted in incorrect predictions of mechanical measures such as stress, apparent stiffness, and failure load criteria (Alsayednoor et al., 2018). Despite the evidence for a more careful selection of segmentation method, most studies used the manufacturer's undocumented approaches for segmentation and morphometric parameter computation. The lack of standardization across studies is likely a major factor as to why newly developed tools are rarely adopted by other research groups. We identified four major issues of standardization that affect our ability to validate new and existing methods and limit their use across research groups.

2.1.6.1 Inconsistent Pre-Processing

While micro-CT was consistently used as the gold standard, the pre-processing methods of these images were not standardized. This lack of consistency resulted in noticeable systematic deviations in the regression of morphological parameters, making it impossible to conclude whether these parameters are under- or overestimated by HR-pQCT. Hence, future studies should aim to define standards for pre-processing of both micro-CT and HR-pQCT images.

2.1.6.2 Limited Comparisons to Similar Methods

Studies that proposed new analysis methods did not always directly compare the results of their approach to those from already existing methods. This was particularly noticeable for FEA studies, where every group used different material models, material properties, geometries, and boundary conditions. The use of independent datasets further complicated the differentiation between method- and dataset-based biases in the results.

2.1.6.3 Limited Validation on Heterogeneous Datasets

Due to the observed dependency of analysis methods on anatomical site and patient cohort, the results from single studies with homogenous datasets cannot directly be compared to one another and the lack of dataset diversity prohibits sufficient assessment of method robustness. Importantly, if these proposed methods are incorporated into proprietary commercial software, opportunities for external validation are then severely limited. Future method development and validation must therefore be based on more diverse datasets to allow for the differentiation between dataset- and pre-processing effects, as this is currently unclear.

2.1.6.4 Lack of 61 μm HR-pQCT Validation

Based on the limited number of 61 μm HR-pQCT validation studies, morphological agreement with micro-CT was improved compared to 82 μm HR-pQCT. While this improved accuracy in morphometric parameters shows promise for mechanical analysis using HR-pQCT FEA, these analyses have yet to be validated against micro-CT FEA. Unfortunately, the validation of morphological parameters is also incomplete, as several morphological parameters have not yet been validated for 61 μm HR-pQCT. As 61 μm HR-pQCT is expected to outperform other clinical imaging modalities, additional validation studies are urgently needed to ensure widespread clinical use of this advanced imaging technology.

2.1.6.5 Conclusion

HR-pQCT is a promising technology with a variety of potential clinical applications. The improved resolution of 61 μm HR-pQCT provides superior quantification of bone morphometrics when compared to 82 μm HR-pQCT. However, the clear lack of standardization is prohibitive for widespread clinical use. Despite the number of studies using HR-pQCT, there is little standardization or agreement as to how to calculate many morphometric indices or perform FEA. Results of HR-pQCT studies are difficult to compare due to varied pre-processing methods. While many groups use the manufacturer's recommended pre-processing, studies have already shown that there are more accurate methods for some morphometric parameters (e.g. Tb.N). However, due to the lack of a publicly available, heterogeneous, and comprehensive dataset for methods validation, studies proposing new analysis methods often draw an incomplete picture of their methods' capabilities. As such, adoption and cross-validation of new methods by other groups is slow and often non-existent. With the availability

of 61 μm HR-pQCT, the community must learn from its mistakes with the introduction of 82 μm HR-pQCT and establish an open access software and data repository, instead of referencing the conclusions of studies using older generation devices with completely different physical properties. Using this data repository, new methods can be benchmarked against the same comprehensive dataset to allow for straightforward comparison. Optimal analysis methods can be selected for each specific application, and updates to methods and their effects on parameters can be immediately accessed through the online repository. With the ability to properly validate new technologies, we can ensure that the clinical use of HR-pQCT can truly provide value in patient diagnosis and care.

Acknowledgements

The authors acknowledge the Swiss National Science Foundation (320030L_170205), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement 841316), and the ETH Postdoctoral Fellowship for financial support.

Conflict of Interest

The authors declare that they have no competing interests.

References

- Adams, M., 2002. Evaluation of three unstructured multigrid methods on 3D finite element problems in solid mechanics. *Int. J. Numer. Methods Eng.* 55, 519–534. <https://doi.org/10.1002/nme.506>
- Adams, M.F., Bayraktar, H.H., Keaveny, T.M., Papadopoulos, P., 2004. Ultrascaleable Implicit Finite Element Analyses in Solid Mechanics with over a Half a Billion Degrees of Freedom, in: *Proceedings of the ACM/IEEE SC 2004 Conference: Bridging Communities*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/SC.2004.62>
- Agarwal, S., Rosete, F., Zhang, C., McMahon, D.J., Guo, X.E., Shane, E., Nishiyama, K.K., 2016. In vivo assessment of bone structure and estimated bone strength by first- and second-generation HR-pQCT. *Osteoporos. Int.* 27, 2955–2966. <https://doi.org/10.1007/s00198-016-3621-8>
- Alberich-Bayarri, A., Marti-Bonmati, L., Pérez, M.A., Sanz-Requena, R., Lerma-Garrido, J.J., García-Martí, G., Moratal, D., 2010. Assessment of 2D and 3D fractal dimension measurements of trabecular bone from high-spatial resolution magnetic resonance images at 3 T. *Med. Phys.* 37, 4930–4937. <https://doi.org/10.1118/1.3481509>
- 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. <https://doi.org/10.1007/s10237-018-1051-6>
- Ang, I.C., Fox, M., Polk, J.D., Kersh, M.E., 2020. An Algorithm for Automated Separation of Trabecular Bone from Variably Thick Cortices in High-Resolution Computed Tomography Data. *IEEE Trans. Biomed. Eng.* 1–1. <https://doi.org/10.1109/TBME.2019.2924398>

- Arbenz, P., van Lenthe, G.H., Mennel, U., Müller, R., Sala, M., 2008. A scalable multi-level preconditioner for matrix-free μ -finite element analysis of human bone structures. *Int. J. Numer. Methods Eng.* 73, 927–947. <https://doi.org/10.1002/nme.2101>
- Bolotin, H.H., Sievänen, H., 2001. Inaccuracies inherent in dual-energy x-ray absorptiometry in vivo bone mineral density can seriously mislead diagnostic/prognostic interpretations of patient-specific bone fragility. *J. Bone Miner. Res.* 16, 799–805. <https://doi.org/10.1359/jbmr.2001.16.5.799>
- Borah, B., Dufresne, T.E., Chmielewski, P.A., Johnson, T.D., Chines, A., Manhart, M.D., 2004. Risedronate preserves bone architecture in postmenopausal women with osteoporosis as measured by three-dimensional microcomputed tomography. *Bone* 34, 736–746. <https://doi.org/10.1016/j.bone.2003.12.013>
- Borah, B., Dufresne, T.E., Ritman, E.L., Jorgensen, S.M., Liu, S., Chmielewski, P.A., Phipps, R.J., Zhou, X., Sibonga, J.D., Turner, R.T., 2006. Long-term risedronate treatment normalizes mineralization and continues to preserve trabecular architecture: Sequential triple biopsy studies with micro-computed tomography. *Bone* 39, 345–352. <https://doi.org/10.1016/j.bone.2006.01.161>
- Brandi, M.L., 2009. Microarchitecture, the key to bone quality. *Rheumatology* 48, iv3–iv8. <https://doi.org/10.1093/rheumatology/kep273>
- Buie, H.R., Campbell, G.M., Klinck, R.J., MacNeil, J.A., Boyd, S.K., 2007. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* 41, 505–515. <https://doi.org/10.1016/j.bone.2007.07.007>
- Burghardt, A.J., Buie, H.R., Laib, A., Majumdar, S., Boyd, S.K., 2010. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. *Bone* 47, 519–528. <https://doi.org/10.1016/j.bone.2010.05.034>

- Burghardt, A.J., Kazakia, G.J., Majumdar, S., 2007. A local adaptive threshold strategy for high resolution peripheral quantitative computed tomography of trabecular bone. *Ann. Biomed. Eng.* 35, 1678–1686. <https://doi.org/10.1007/s10439-007-9344-4>
- Caselles, V., Kimmel, R., Sapiro, G., 1997. Geodesic Active Contours. *Int. J. Comput. Vis.* 22, 61–79. <https://doi.org/10.1023/A:1007979827043>
- Chen, Y., Dall'Ara, E., Sales, E., Manda, K., Wallace, R., Pankaj, P., Viceconti, M., 2017. Micro-CT based finite element models of cancellous bone predict accurately displacement once the boundary condition is well replicated: A validation study. *J. Mech. Behav. Biomed. Mater.* 65, 644–651. <https://doi.org/10.1016/j.jmbbm.2016.09.014>
- Chevalier, Y., Charlebois, M., Pahr, D., Varga, P., Heini, P., Schneider, E., Zysset, P., 2008. A patient-specific finite element methodology to predict damage accumulation in vertebral bodies under axial compression, sagittal flexion and combined loads. *Comput. Methods Biomech. Biomed. Engin.* 11, 477–487. <https://doi.org/10.1080/10255840802078022>
- Chiang, C.Y., Zebaze, R., Wang, X.F., Ghasem-Zadeh, A., Zajac, J.D., Seeman, E., 2018. Cortical Matrix Mineral Density Measured Noninvasively in Pre- and Postmenopausal Women and a Woman With Vitamin D-Dependent Rickets. *J. Bone Miner. Res.* 33, 1312–1317. <https://doi.org/10.1002/jbmr.3415>
- Christen, P., Schulte, F.A., Zwahlen, A., van Rietbergen, B., Boutroy, S., Melton, L.J., Amin, S., Khosla, S., Goldhahn, J., Müller, R., 2016. Voxel size dependency, reproducibility and sensitivity of an in vivo bone loading estimation algorithm. *J. R. Soc. Interface* 13, 20150991. <https://doi.org/10.1098/rsif.2015.0991>
- Cohen, A., Dempster, D.W., Müller, R., Guo, X.E., Nickolas, T.L., Liu, X.S., Zhang, X.H., Wirth, A.J., van Lenthe, G.H., Kohler, T., McMahon, D.J., Zhou, H., Rubin, M.R., Bilezikian, J.P., Lappe, J.M., Recker, R.R., Shane, E., 2010. Assessment of trabecular and cortical architecture and mechanical competence of bone by high-resolution

- peripheral computed tomography: Comparison with transiliac bone biopsy. *Osteoporos. Int.* 21, 263–273. <https://doi.org/10.1007/s00198-009-0945-7>
- Dall'Ara, E., Schmidt, R., Pahr, D., Varga, P., Chevalier, Y., Patsch, J., Kainberger, F., Zysset, P., 2010. A nonlinear finite element model validation study based on a novel experimental technique for inducing anterior wedge-shape fractures in human vertebral bodies in vitro. *J. Biomech.* 43, 2374–2380. <https://doi.org/10.1016/j.jbiomech.2010.04.023>
- Feldkamp, L.A., Goldstein, S.A., Parfitt, M.A., Jesion, G., Kleerekoper, M., 1989. The direct examination of three-dimensional bone architecture in vitro by computed tomography. *J. Bone Miner. Res.* 4, 3–11. <https://doi.org/10.1002/jbmr.5650040103>
- Flaig, C., 2012. A highly scalable memory efficient multigrid solver for μ -finite element analyses. ETH Zurich. <https://doi.org/10.3929/ethz-a-007613965>
- Goulet, R.W., Goldstein, S.A., Ciarelli, M.J., Kuhn, J.L., Brown, M.B., Feldkamp, L.A., 1994. The relationship between the structural and orthogonal compressive properties of trabecular bone. *J. Biomech.* 27. [https://doi.org/10.1016/0021-9290\(94\)90014-0](https://doi.org/10.1016/0021-9290(94)90014-0)
- Hafri, M., Jennane, R., Lespessailles, E., Toumi, H., 2016a. Dual active contours model for HR-pQCT cortical bone segmentation, in: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, pp. 2270–2275. <https://doi.org/10.1109/ICPR.2016.7899974>
- Hafri, M., Toumi, H., Boutroy, S., Chapurlat, R.D., Lespessailles, E., Jennane, R., 2016b. Fuzzy energy based active contours model for HR-PQCT cortical bone segmentation, in: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 4334–4338. <https://doi.org/10.1109/ICIP.2016.7533178>
- Hambli, R., 2013. Micro-CT finite element model and experimental validation of trabecular bone damage and fracture. *Bone* 56, 363–374. <https://doi.org/10.1016/j.bone.2013.06.028>

- Harrigan, T.P., Mann, R.W., 1985. Characterization of microstructural anisotropy in cancellous bone using a second rank tensor. *J. Mater. Sci.* 19, 761–767.
- Hildebrand, T., Laib, A., Müller, R., Dequeker, J., Rüegegger, P., 1999. Direct Three-Dimensional Morphometric Analysis of Human Cancellous Bone: Microstructural Data from Spine, Femur, Iliac Crest, and Calcaneus. *J. Bone Miner. Res.* 14, 1167–1174. <https://doi.org/10.1359/jbmr.1999.14.7.1167>
- Hildebrand, T., Rüegegger, P., 1997. A new method for the model-independent assessment of thickness in three-dimensional images. *J. Microsc.* 185, 67–75. <https://doi.org/10.1046/j.1365-2818.1997.1340694.x>
- Hollister, S.J., Brennan, J.M., Kikuchi, N., 1994. A homogenization sampling procedure for calculating trabecular bone effective stiffness and tissue level stress. *J. Biomech.* 27, 433–444. [https://doi.org/10.1016/0021-9290\(94\)90019-1](https://doi.org/10.1016/0021-9290(94)90019-1)
- Hosseini, H.S., Maquer, G., Zysset, P.K., 2017. μ CT-based trabecular anisotropy can be reproducibly computed from HR-pQCT scans using the triangulated bone surface. *Bone* 97, 114–120. <https://doi.org/10.1016/j.bone.2017.01.016>
- Järvinen, T.L.N., Sievänen, H., Khan, K.M., Heinonen, A., Kannus, P., 2008. Shifting the focus in fracture prevention from osteoporosis to falls. *Bmj* 336, 124–126. <https://doi.org/10.1136/bmj.39428.470752.ad>
- Jorgenson, B.L., Buie, H.R., McErlain, D.D., Sandino, C., Boyd, S.K., 2015. A comparison of methods for in vivo assessment of cortical porosity in the human appendicular skeleton. *Bone* 73, 167–175. <https://doi.org/10.1016/j.bone.2014.11.023>
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *Int. J. Comput. Vis.* 1, 321–331. <https://doi.org/10.1007/BF00133570>
- Kim, D.G., Christopherson, G.T., Dong, X.N., Fyhrie, D.P., Yeni, Y.N., 2004. The effect of microcomputed tomography scanning and reconstruction voxel size on the accuracy of

- stereological measurements in human cancellous bone. *Bone* 35, 1375–1382. <https://doi.org/10.1016/j.bone.2004.09.007>
- Klintström, E., Klintström, B., Moreno, R., Brismar, T.B., Pahr, D.H., Smedby, Ö., 2016. Predicting trabecular bone stiffness from clinical cone-beam CT and HR-pQCT Data; an In vitro study using finite element analysis. *PLoS One* 11, 1–19. <https://doi.org/10.1371/journal.pone.0161101>
- Kothari, M., Keaveny, T.M., Lin, J.C., Newitt, D.C., Genant, H.K., Majumdar, S., 1998. Impact of spatial resolution on the prediction of trabecular architecture parameters. *Bone* 22, 437–443. [https://doi.org/10.1016/S8756-3282\(98\)00031-3](https://doi.org/10.1016/S8756-3282(98)00031-3)
- Krause, M., Museyko, O., Breer, S., Wulff, B., Duckstein, C., Vettorazzi, E., Glueer, C., Püschel, K., Engelke, K., Amling, M., 2014. Accuracy of trabecular structure by HR-pQCT compared to gold standard μ CT in the radius and tibia of patients with osteoporosis and long-term bisphosphonate therapy. *Osteoporos. Int.* 25, 1595–1606. <https://doi.org/10.1007/s00198-014-2650-4>
- Laib, A., 1997. Ridge number density: A new parameter for in vivo bone structure analysis. *Bone* 21, 541–546. [https://doi.org/10.1016/S8756-3282\(97\)00205-6](https://doi.org/10.1016/S8756-3282(97)00205-6)
- Laib, A., Barou, O., Vico, L., Lafage-Proust, M.H., Alexandre, C., Rügsegger, P., 2000. 3D micro-computed tomography of trabecular and cortical bone architecture with application to a rat model of immobilisation osteoporosis. *Med. Biol. Eng. Comput.* 38, 326–332. <https://doi.org/10.1007/BF02347054>
- Laib, A., Häuselmann, H.J., Rügsegger, P., 1998. In vivo high resolution 3D-QCT of the human forearm. *Technol. Heal. Care* 6, 329–337. <https://doi.org/10.3233/THC-1998-65-606>
- Laib, A., Rügsegger, P., 1999a. Comparison of structure extraction methods for in vivo trabecular bone measurements. *Comput. Med. Imaging Graph.* 23, 69–74. [https://doi.org/10.1016/S0895-6111\(98\)00071-8](https://doi.org/10.1016/S0895-6111(98)00071-8)

- Laib, A., Rüegsegger, P., 1999b. Calibration of trabecular bone structure measurements of in vivo three-dimensional peripheral quantitative computed tomography with 28- μ m-resolution microcomputed tomography. *Bone* 24, 35–39. [https://doi.org/10.1016/S8756-3282\(98\)00159-8](https://doi.org/10.1016/S8756-3282(98)00159-8)
- Liu, X.S., Shane, E., McMahon, D.J., Guo, X.E., 2011. Individual trabecula segmentation (ITS)-based morphological analysis of microscale images of human tibial trabecular bone at limited spatial resolution. *J. Bone Miner. Res.* 26, 2184–2193. <https://doi.org/10.1002/jbmr.420>
- Liu, X.S., Wang, J., Zhou, B., Stein, E., Shi, X., Adams, M., Shane, E., Guo, X.E., 2013. Fast trabecular bone strength predictions of HR-pQCT and individual trabeculae segmentation-based plate and rod finite element model discriminate postmenopausal vertebral fractures. *J. Bone Miner. Res.* 28, 1666–1678. <https://doi.org/10.1002/jbmr.1919>
- Liu, X.S., Zhang, X.H., Sekhon, K.K., Adams, M.F., McMahon, D.J., Bilezikian, J.P., Shane, E., Guo, X.E., 2010. High-resolution peripheral quantitative computed tomography can assess microstructural and mechanical properties of human distal tibial bone. *J. Bone Miner. Res.* 25, 746–756. <https://doi.org/10.1359/jbmr.090822>
- MacNeil, J.A., Boyd, S.K., 2007. Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med. Eng. Phys.* 29, 1096–1105. <https://doi.org/10.1016/j.medengphy.2006.11.002>
- Manske, S.L., Davison, E.M., Burt, L.A., Raymond, D.A., Boyd, S.K., 2017. The Estimation of Second-Generation HR-pQCT From First-Generation HR-pQCT Using In Vivo Cross-Calibration. *J. Bone Miner. Res.* 32, 1514–1524. <https://doi.org/10.1002/jbmr.3128>
- Manske, S.L., Zhu, Y., Sandino, C., Boyd, S.K., 2015. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone* 79, 213–221. <https://doi.org/10.1016/j.bone.2015.06.006>

- Marquez-Neila, P., Baumela, L., Alvarez, L., 2014. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 2–17. <https://doi.org/10.1109/TPAMI.2013.106>
- Metcalf, L.M., Dall’Ara, E., Paggiosi, M.A., Rochester, J.R., Vilayphiou, N., Kemp, G.J., McCloskey, E. V., 2018. Validation of calcaneus trabecular microstructure measurements by HR-pQCT. *Bone* 106, 69–77. <https://doi.org/10.1016/j.bone.2017.09.013>
- Müller, A., Ruegsegger, E., Ruegsegger, P., 1989. Peripheral QCT: A low-risk procedure to identify women predisposed to osteoporosis. *Phys. Med. Biol.* 34, 741–749. <https://doi.org/10.1088/0031-9155/34/6/009>
- Müller, R., 2003. Bone microarchitecture assessment: current and future trends. *Osteoporos. Int.* 14 Suppl 5, 89–99. <https://doi.org/10.1007/s00198-003-1479-z>
- Müller, R., Hildebrand, T., Ruegsegger, P., 1994. Non-invasive bone biopsy: A new method to analyse and display the three-dimensional structure of trabecular bone. *Phys. Med. Biol.* 39, 145–164. <https://doi.org/10.1088/0031-9155/39/1/009>
- Müller, 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.
- Mys, K., Varga, P., Gueorguiev, B., Hemmatian, H., Stockmans, F., van Lenthe, G.H., 2019. Correlation Between Cone-Beam Computed Tomography and High-Resolution Peripheral Computed Tomography for Assessment of Wrist Bone Microstructure. *J. Bone Miner. Res.* 34, 867–874. <https://doi.org/10.1002/jbmr.3673>
- Nelson, H.D., Helfand, M., Woolf, S.H., Allan, J.D., 2002. Screening for Postmenopausal Osteoporosis: A Review of the Evidence for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 137, 529. <https://doi.org/10.7326/0003-4819-137-6-200209170-00015>

- Nishiyama, K.K., Macdonald, H.M., Buie, H.R., Hanley, D.A., Boyd, S.K., 2010. Postmenopausal women with osteopenia have higher cortical porosity and thinner cortices at the distal radius and tibia than women with normal aBMD: An in vivo HR-pQCT study. *J. Bone Miner. Res.* 25, 882–890. <https://doi.org/10.1359/jbmr.091020>
- Odgaard, A., Gundersen, H.J.G., 1993. Quantification of connectivity with special emphasis on 3D reconstructions. *Bone* 14, 173–182.
- Ostertag, A., Peyrin, F., Fernandez, S., Laredo, J.D., de Vernejoul, M.C., Chappard, C., 2014. Cortical measurements of the tibia from high resolution peripheral quantitative computed tomography images: A comparison with synchrotron radiation micro-computed tomography. *Bone* 63, 7–14. <https://doi.org/10.1016/j.bone.2014.02.009>
- Ostertag, A., Peyrin, F., Gouttenoire, P.J., Laredo, J.D., Devernejoul, M.C., Cohen Solal, M., Chappard, C., 2016. Multiscale and multimodality computed tomography for cortical bone analysis. *Phys. Med. Biol.* 61, 8553–8576. <https://doi.org/10.1088/0031-9155/61/24/8553>
- Otsu, N., 1979. Threshold Selection Method From Gray-Level Histograms. *IEEE Trans Syst Man Cybern SMC-9*, 62–66. <https://doi.org/10.1109/tsmc.1979.4310076>
- Pahr, D.H., Dall’Ara, E., Varga, P., Zysset, P.K., 2012. HR-pQCT-based homogenised finite element models provide quantitative predictions of experimental vertebral body stiffness and strength with the same accuracy as μ FE models. *Comput. Methods Biomech. Biomed. Engin.* 15, 711–720. <https://doi.org/10.1080/10255842.2011.556627>
- Parfitt, A.M., Drezner, M.K., Glorieux, F.H., Kanis, J.A., Malluche, H., Meunier, P.J., Ott, S.M., Recker, R.R., 2009. Bone histomorphometry: Standardization of nomenclature, symbols, and units: Report of the asbmr histomorphometry nomenclature committee. *J. Bone Miner. Res.* 2, 595–610. <https://doi.org/10.1002/jbmr.5650020617>
- Peters, M., Scharmga, A., van Tubergen, A., Arts, J., Loeffen, D., Weijers, R., van Rietbergen, B., Geusens, P., van den Bergh, J.P., 2017. The Reliability of a Semi-automated

- Algorithm for Detection of Cortical Interruptions in Finger Joints on High Resolution CT Compared to MicroCT. *Calcif. Tissue Int.* 101, 132–140. <https://doi.org/10.1007/s00223-017-0264-5>
- Revol-Muller, C., Peyrin, F., Carrillon, Y., Odet, C., 2002. Automated 3D region growing algorithm based on an assessment function. *Pattern Recognit. Lett.* 23, 137–150. [https://doi.org/10.1016/S0167-8655\(01\)00116-7](https://doi.org/10.1016/S0167-8655(01)00116-7)
- Ridler, T.W., Calvard, S., 1978. Picture Thresholding Using. *IEEE Trans. Syst. Man Cybern.* smc-8, 630–632.
- Rüegsegger, P., Koller, B., Müller, R., 1996. A Microtomographic System for the Nondestructive Evaluation of Bone Architecture. *Calcif. Tissue Int.* 58, 24–29. <https://doi.org/10.1007/s002239900006>
- Salmon, P.L., Ohlsson, C., Shefelbine, S.J., Doube, M., 2015. Structure model index does not measure rods and plates in trabecular bone. *Front. Endocrinol. (Lausanne)*. 6, 1–10. <https://doi.org/10.3389/fendo.2015.00162>
- Scharmga, A., Peters, M., van Tubergen, A., van Den Bergh, J., De Jong, J., Loeffen, D., van Rietbergen, B., Weijers, R., Geusens, P., 2016. Visual detection of cortical breaks in hand joints: Reliability and validity of high-resolution peripheral quantitative CT compared to microCT. *BMC Musculoskelet. Disord.* 17, 1–7. <https://doi.org/10.1186/s12891-016-1148-y>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2011. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone

- resorption parameters using time-lapsed imaging. *Bone* 48, 433–442.
<https://doi.org/10.1016/j.bone.2010.10.007>
- Scott, D.L., 2003. Joint damage and disability in RA. *Clin. Exp. Biol.* 21, S23–S27.
- Sode, M., Burghardt, A.J., Nissenon, R.A., Majumdar, S., 2008. Resolution Dependence of the Non-metric Trabecular Structure Indices. *Bone* 42, 728–736.
<https://doi.org/10.1016/j.bone.2007.12.004>
- Soltan, N., Kawalilak, C.E., Cooper, D.M., Kontulainen, S.A., Johnston, J.D., 2019. Cortical porosity assessment in the distal radius: A comparison of HR-pQCT measures with Synchrotron-Radiation micro-CT-based measures. *Bone* 120, 439–445.
<https://doi.org/10.1016/j.bone.2018.12.008>
- Stauber, M., Müller, R., 2008. Micro-Computed Tomography: A Method for the Non-Destructive Evaluation of the Three-Dimensional Structure of Biological Specimens, in: Westendorf, J.J. (Ed.), *Osteoporosis. Methods in Molecular Biology*. Humana Press, pp. 273–292. https://doi.org/10.1007/978-1-59745-104-8_19
- Su, R., Campbell, G.M., Boyd, S.K., 2007. Establishment of an architecture-specific experimental validation approach for finite element modeling of bone by rapid prototyping and high resolution computed tomography. *Med. Eng. Phys.* 29, 480–490.
<https://doi.org/10.1016/j.medengphy.2006.06.010>
- Tjong, W., Kazakia, G.J., Burghardt, A.J., Majumdar, S., 2012. The effect of voxel size on high-resolution peripheral computed tomography measurements of trabecular and cortical bone microstructure. *Med. Phys.* 39, 1893–1903.
<https://doi.org/10.1118/1.3689813>
- Treece, G.M., Gee, A.H., Mayhew, P.M., Poole, K.E.S., 2010. High resolution cortical bone thickness measurement from clinical CT data. *Med. Image Anal.* 14, 276–290.
<https://doi.org/10.1016/j.media.2010.01.003>

- Treece, G.M., Poole, K.E.S., Gee, A.H., 2012. Imaging the femoral cortex: Thickness, density and mass from clinical CT. *Med. Image Anal.* 16, 952–965. <https://doi.org/10.1016/j.media.2012.02.008>
- van Rietbergen, B., 1996. Direct mechanics assessment of elastic symmetries and properties of trabecular bone architecture. *J. Biomech.* 29, 1653–1657. [https://doi.org/10.1016/0021-9290\(96\)00093-0](https://doi.org/10.1016/0021-9290(96)00093-0)
- van Rietbergen, B., Weinans, H., Huiskes, R., Odgaard, A., 1995. A new method to determine trabecular bone elastic properties and loading using micromechanical finite-element models. *J. Biomech.* 28, 69–81. [https://doi.org/10.1016/0021-9290\(95\)80008-5](https://doi.org/10.1016/0021-9290(95)80008-5)
- Varga, P., Zysset, P.K., 2009. Assessment of volume fraction and fabric in the distal radius using HR-pQCT. *Bone* 45, 909–917. <https://doi.org/10.1016/j.bone.2009.07.001>
- Wachter, N.J., Augat, P., Krischak, G.D., Mentzel, M., Kinzl, L., Claes, L., 2001. Prediction of cortical bone porosity in vitro by microcomputed tomography. *Calcif. Tissue Int.* 68, 38–42. <https://doi.org/10.1007/s002230001182>
- Wang, J., Zhou, B., Jenny Hu, Y., Zhang, Z., Eric Yu, Y., Nawathe, S., Nishiyama, K.K., Keaveny, T.M., Shane, E., Guo, X.E., 2019. Accurate and efficient plate and rod microfinite element models for whole bone segments based on high-resolution peripheral computed tomography. *J. Biomech. Eng.* 141. <https://doi.org/10.1115/1.4042680>
- Werner, D., Simon, D., Englbrecht, M., Stemmler, F., Simon, C., Berlin, A., Haschka, J., Renner, N., Buder, T., Engelke, K., Hueber, A.J., Rech, J., Schett, G., Kleyer, A., 2017. Early Changes of the Cortical Micro-Channel System in the Bare Area of the Joints of Patients With Rheumatoid Arthritis. *Arthritis Rheumatol.* 69, 1580–1587. <https://doi.org/10.1002/art.40148>
- Whitehouse, W.J., 1974. The quantitative morphology of anisotropic trabecular bone. *J. Microsc.* 101, 153–168. <https://doi.org/10.1111/j.1365-2818.1974.tb03878.x>

- Whittier, D.E., Boyd, S.K., Burghardt, A.J., Paccou, J., Ghasem-Zadeh, A., Chapurlat, R., Engelke, K., Bouxsein, M.L., 2020. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. *Osteoporos. Int.* <https://doi.org/10.1007/s00198-020-05438-5>
- Whittier, D.E., Manske, S.L., Kiel, D.P., Bouxsein, M., Boyd, S.K., 2018. Harmonizing finite element modelling for non-invasive strength estimation by high-resolution peripheral quantitative computed tomography. *J. Biomech.* 80, 63–71. <https://doi.org/10.1016/j.jbiomech.2018.08.030>
- Zebaze, R., Ghasem-Zadeh, A., Mbala, A., Seeman, E., 2013. A new method of segmentation of compact-appearing, Transitional and trabecular compartments and quantification of cortical porosity from high resolution peripheral quantitative computed tomographic images. *Bone* 54, 8–20. <https://doi.org/10.1016/j.bone.2013.01.007>
- Zhang, Z.M., Li, Z.C., Jiang, L.S., Jiang, S.D., Dai, L.Y., 2010. Micro-CT and mechanical evaluation of subchondral trabecular bone structure between postmenopausal women with osteoarthritis and osteoporosis. *Osteoporos. Int.* 21, 1383–1390. <https://doi.org/10.1007/s00198-009-1071-2>
- Zhou, B., Wang, J., Yu, Y.E., Zhang, Z., Nawathe, S., Nishiyama, K.K., Rosete, F.R., Keaveny, T.M., Shane, E., Guo, X.E., 2016. High-resolution peripheral quantitative computed tomography (HR-pQCT) can assess microstructural and biomechanical properties of both human distal radius and tibia: Ex vivo computational and experimental validations. *Bone* 86, 58–67. <https://doi.org/10.1016/j.bone.2016.02.016>
- Zhou, B., Zhang, Z., Hu, Y.J., Wang, J., Yu, Y.E., Nawathe, S., Nishiyama, K.K., Keaveny, T.M., Shane, E., Guo, X.E., 2019. Regional Variations of HR-pQCT Morphological and Biomechanical Measurements of Bone Segments and Their Associations With Whole Distal Radius and Tibia Mechanical Properties. *J. Biomech. Eng.* 141. <https://doi.org/10.1115/1.4044175>

Chapter 3

Development of a computational framework for the assessment of bone fracture healing

3.1 A cross-platform and scalable Python framework for bone microstructure simulations and analysis

Nicholas Ohs¹, Duncan Tourolle né Betts¹, Daniele Boaretti¹, Matthias Graß¹, Sheel Nidhan¹, Francisco Correia Marques¹, Ralph Müller¹

¹ Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

Abstract

Computational tools have shaped science in the 21st century. With the ever increasing power of modern computer hardware, new ways to explore data, new ways to make predictions based on data and new ways to study the world using simulations have emerged. In the bone field specifically, the use of micro-computed tomography images quickly led to high demand of computational resources and software capable of handling large 3D images. Over the last decades, many research groups have tried to develop software to aid them in their research. Though, due to several reasons such as closed-source black box software or limited interest in sharing software in the community, the current computational state of the bone field is fragmented. This is particularly visible between research groups focusing on animal studies on the one and human studies on the other hand. This limits comparability between animal and human studies for example when studying bone fractures, thus breaking the critical feedback cycle between clinical and laboratory research.

To tackle this issue, we propose a new computational framework to provide the basis for bone research of both, clinical and laboratory work. It is based on Python, easily one of the most popular open source programming languages available today, with thousands of freely accessible libraries. Common functionality such as finite element analysis, morphological analysis and image processing routines are provided. The choice of finite element solver guarantees decent scalability from laptop to supercomputer with the Python programming language ensuring cross-platform compatibility. By providing both, tools for human and for animal studies within one framework, researchers can easily switch between them to compare and/or decide which approach to use.

Keywords:

Framework; Python; Bone; Bone Adaptation; Morphological Analysis; Finite Element Analysis

3.1.1 Introduction

Computers have come a long way with current super-computers reaching up to 514,000 TFlop/s (that is 514 thousand trillion floating point operations per second) (<https://top500.org/>). Nowadays, even regular laptops are faster than super-computers only a few decades ago. In the bone field, scientists have used computers for various applications to gain deeper insight into how bone remodels, especially since the availability of micro-computed tomography (micro-CT) which provides high resolution (up to 1 μm) of the bone microstructure. This allowed first of all, to quantify the morphology of biopsy data without the need for approximating 2D histology approaches (Müller et al., 1998) and also allowed for longitudinal *in vivo* mouse studies, which wasn't possible with other techniques (Schulte et al., 2011) and which has also been applied to study bone fractures (Tourolle né Betts et al., 2020). Since there was a clear connection between observed morphology and mechanical loading, methods were developed to compute mechanical properties such as stiffness, instead of invasive measurements (Flaig, 2012; Mennel and Sala, 2006).

However, the morphological analysis software, called IPL, was only available to customers of the micro-CT scanners, leading to e.g. parallel developments of morphological analysis software based on the popular imaging software ImageJ called BoneJ (Doube et al., 2010). While BoneJ was free and open-access, results were not always identical to previously used software. Furthermore, version incompatibilities between BoneJ and ImageJ lead to a slow adaptation of the software. Furthermore, IPL lacks many of the features one might expect from a modern programming language. For finite element analysis, the fragmentation of the community was even stronger with most groups developing their own approaches and only limited efforts to share (Ohs et al., 2020). To make matters worse, with the introduction of high-resolution peripheral computed tomography (HR-pQCT) and the resulting availability of 3D microstructural bone information of patients, scanner manufacturer's provided alternative morphological analyses methods without proper validation (Ohs et al., 2020). To not fall into the same trap, when studying bone fractures care should be taken to stay compatible with what has already been developed in the lab.

Moving forward, the initial step must be to bring researchers together on a common computational platform. This will enable easy exchange of ideas and pave the way for a unified set of tools to analyse bone. In the following, we propose a solution that does not exclude

existing standards, while clearly pointing the way forward to the next generation of computational bone tools.

3.1.2 Python

Python is an open-source high-level programming language developed by Guido van Rossum (Python Software Foundation, 2020). It is listed as the most popular programming language today by the PYPL PopularitY index (<https://pypl.github.io>). The language combines several programming paradigms like object-oriented and functional programming which combined with its simple white-space based syntax makes it easy to learn for most. While Python itself is slower than languages like C and C++ due to it being an interpreted language, Python can easily be extended by modules written in C and C++ making it as fast as these languages, while providing an easier to read and clearer code compared to those languages.

One major benefit of Python is that it runs on various platforms. It ships by default with Linux, can be installed on Mac OSX with the popular homebrew package manager and is bundled with a large amount of datascience and numerics packages for Windows in what is called the Anaconda distribution (Anaconda, 2020). The available packages combined with the Python package manager pip are another benefit of the Python programming language. From the numerics library numpy (Van Der Walt et al., 2011) and scipy (Virtanen et al., 2019) to the plotting library matplotlib (Hunter, 2007) and machine learning packages such as Scikit-Learn (Pedregosa et al., 2011) and TensorFlow (Mart'ın et al., 2015) etc., Python already provides many default utilities. However, the package repository of Python actually contains over 260.000 projects that are freely available. The ease of use, the cross-platform compatibility, the maturity Python has nowadays and the available packages make it an ideal basis for science today.

3.1.2 Basics of the framework

Modern code can easily reach dimensions that are beyond anyone's ability to oversee and comprehend fully. When code suddenly fails, reverting changes can become challenging, as basic file information like date of change do not really link back to a valid version of the software necessarily. When working with several developers on the same code base, another

question that quickly arises is how to combine changes made by different developers. For all these issues, version control software has been developed. The framework we propose is based on the most popular open-source version control software called git. In short, git is a decentralised version control/tracking software that provides features such as bundling several edits made to files into so called “commits”, splitting different feature developments into separated “branches”, assisted merging of several feature branches and more.

Next to being able to have a clear history of ones code, it is also essential to test it. To automate this, GitLab, a devops lifecycle tool based on git was setup. It provides a web-interface to report code issues, request features. It also provides a visual interface for merging branches. Most importantly, we setup GitLab with what is known as GitLab runners. These runners run a separate software that waits for new commits being detected by the GitLab server. Once the GitLab server reports new commits, the runners will then automatically start to run all written tests for the code repository.

As was mentioned previously, performance sensitive code can be written in C++ and easily linked into Python. For this, we use the publicly available library pybind11 (Jakob et al., 2017). We also use the library Sphinx to autogenerate documentation from Python source code files and Doxygen for C++ files.

To address the issue of making the developed software accessible for non-programmers, a JupyterHub server was configured to provide Jupyterlab interfaces. Jupyterlab is a webinterface serving interactive Python Jupyter Notebooks to user via a WebBrowser. Complex pipelines can then be provided as notebooks with users mainly having to use their mouse to interact with them.

3.1.4 Modules

Certain core modules were developed to provide users with expected features when working in the bone field.

3.1.4.1 File formats

The Any Image (AIM) format is used on the micro-CT manufacturer’s hardware. An interface was provided to load these images into Python. The imaging data is provided as numpy arrays

(the default for multidimensional arrays in Python) with physical quantities attached with the package pint (Grecco, 2020). The array itself is provided in a class that also provides further information such as the voxel size, the position relative to the total scanned region and the processing log (Figure 3.1.1). By default, the returned array is in calibrated units, though this feature is optional and can be disabled.

```
# import the aim module
from ifb_framework import aim

# call the read_aim function with the filepath
my_file = aim.load_aim('test_float.aim')

# print what we just read from file to the terminal
print('Voxelsize in mm:', my_file.voxelsize)
print('Processing log:', my_file.processing_log)
print('Position:', my_file.position)
print('Data:', my_file.data)
```

Fig. 3.1.1 Code listing showing how to interact with an AIM file in Python.

While the AIM format is the default for the scanner manufacturer's system, it also has some noticeable limitations. Metadata is hardcoded into the file format making changes to it difficult except for additions to the processing log which would require all meta data to be converted to strings. Multiple datasets that should be grouped logically can only be stored in separate AIM files. No parallel read and write for high performance applications is supported by the format. On the other hand, the hierarchical dataset format (HDF5) (The HDF Group, 2020) provides all these features and more. It is also the standard file format used for the micro-FE solver parOsol (Flaig, 2012) that is often used in the bone field.

Since the HDF5 format does not by default respect physical units attached to numpy arrays, we developed a light wrapper that provides this extra functionality (Figure 3.1.2). The files are stored as *.h5i, but they are fully compatible with the default HDF5 Python modules.

```

# import the h5i module
from ifb_framework import h5i

# open a file with the same syntax as a text file
# in Python
with h5i.File('my_test_file.h5i', 'r') as h5i_file:
    image, meta_data = h5i_file.read('segmented image')
    print(meta_data['threshold'])
    # Outputs e.g.: 450.0 milligram / centimeter ** 3

```

Fig. 3.1.2 Code listing showing how to interact with an h5i-File.

3.1.4.2 Image Processing

Many image processing tools such as Gaussian filters, erosion and dilation filters, and many more are already provided by Scikit-Image. Rigid image registration was based on the work by Thévenaz and colleagues (Thévenaz et al., 1998) was implemented in Python and its accuracy was tested against existing implementations. Furthermore, the machanoregulation analysis by Schulte and colleagues (Schulte et al., 2011) has been implemented.

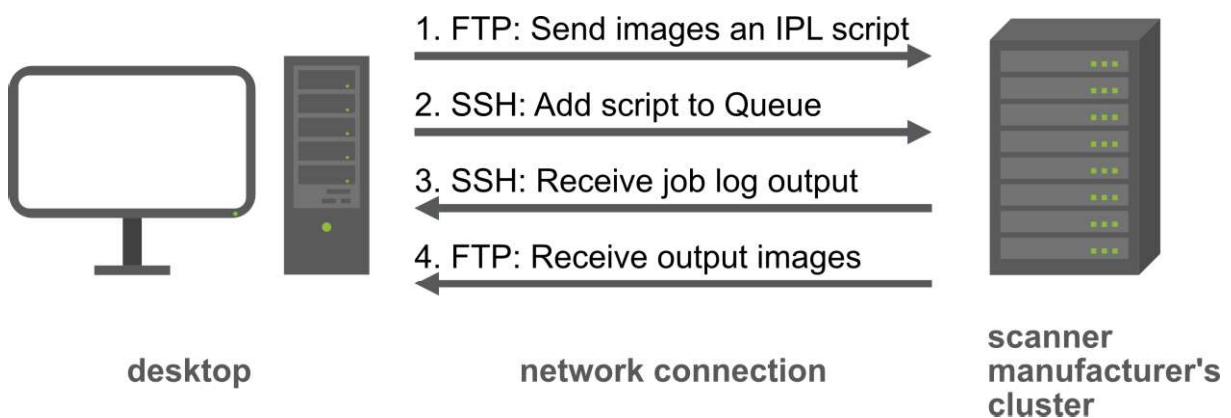


Fig. 3.1.3 While the Python user calls a simple morphometry function, the framework communicates over the network with the scanner manufacturer's cluster to run their black box code.

To have access to image processing pipelines programmed in IPL, an interface was developed that allows to wrap calls to IPL functions in Python. To realize this, an FTP connection is opened to the scanner manufacturer's cluster and the required images are transferred. Via an

SSH connection, IPL scripts and then sent to a queue of choice and the program waits for the job finishing. Finally, the results are obtained from the scanner manufacturer’s cluster (Figure 3.1.3). The interfaces are designed as if one would call a Python function (such as expecting a numpy array as an argument etc.). This allows to easily swap out the existing wrapper with implementations in Python, once the community agrees upon an open source implementation of these IPL functions.

3.1.4.3 External Job & Finite Element Solver

ParOsol is a high performance parallel octree solver that was wrapped to be callable from the framework. One of the key aspects when developing the wrapper for parOsol was to ensure that

```
from ifb_framework import Quantity
from ifb_framework.openVMS.legacy_ipl import tri

# binary_image = ...
# trabecular_mask = ...
voxel_size = Quantity(12., 'um')
tri(binary_image, trabecular_mask, voxel_size)

# Outputs something like
{'p_ratio': <Quantity(1.831678, 'dimensionless')>,
 'structural model index': <Quantity(1.206045, 'dimensionless')>,
 'surface': <Quantity(5799.261673, 'millimeter ** 2')>,
 'surface/volume': <Quantity(15.672709, '1 / millimeter')>,
 'total volume': <Quantity(1840.083252, 'millimeter ** 3')>,
 'trabecular number': <Quantity(1.575815, '1 / millimeter')>,
 'trabecular spacing': <Quantity(0.506982, 'millimeter')>,
 'trabecular thickness': <Quantity(0.12761, 'millimeter')>,
 'volume': <Quantity(370.022918, 'millimeter ** 3')>,
 'volume/total volume': <Quantity(0.20109, 'dimensionless')>}
```

Fig. 3.1.4 Code listing showing how a call to “tri” yields the known output of the IPL function “tri” including physical quantities.

running a Python script on a desktop PC would look identical to a script running on a large cluster or super-computer. This was achieved by abstracting away calls to external programs such as parOsol via the external job module. The module differentiates three different scenarios:

1. Being run on a Windows or Linux machine without a queuing system

In this case, the Python “submodule” package will be utilized to call the program in question in a child process.

2. Being run on a cluster without allocated resources

In this case, an appropriate job queue script will be generated. The script will be submitted into the queuing system and the Python script waits automatically until the job is done and the results can be read.

3. Being run on a cluster with allocated resource

In this case, the allocated resources of the system are determined and the program in question is run with these system resources as a child process.

```
import numpy as np
from ifb_framework.parosol import Parosol, compute_mechanics
from ifb_framework import Quantity

image_size = 25
script = (
    "Define Nodelist Bottom as Nodes on Plane with "\
    "x=0 y=0 z=0 u=0 v=0 w=1\n"\
    "Apply Displacement FixBottom on Nodelist Bottom using "\
    "uniform_displacement with z=1 d=0\n"\
    "Define Nodelist Top as Nodes on Plane with "\
    "x=0 y=0 z={height} u=0 v=0 w=1\n"\
    "Apply Displacement DisplaceTop on Nodelist Top using "\
    "uniform_displacement with z=-1 d=1%"
)
# Uniform Young's Modulus of 1 GPa
image = Quantity(np.ones((image_size,)*3), 'GPa')

res = compute_mechanics(
    image,
    script.format(height=image_size),
    voxel_size=Quantity(0.0105, 'mm'),
    output=['SED'])
```

Fig. 3.1.5 Code listing showing how a call parosol irrespective of the platform you are working on. The code fixed all nodes touching the bottom plane of the image in z direction (defined by x,y,z=0, w=1) in z direction while applying a 1% compression on the top plane.

In the specific case of parOsol, care was taken to make it easy for non-programmers to be able to run basic finite element tasks without assistance. For this, the simple boundary condition language was developed, implemented in C++ and linked into Python. A typical parOsol call can be seen in figure 3.1.5. Instead of writing code to generate a list of nodes for which to then apply certain boundary conditions, the user can use the simple boundary condition language (SBL) to define typical finite element computations.

For more complex loading scenarios, the parOsol calls can be combined with a load estimation algorithm (Christen et al., 2013) that was implemented in Python, to get more accurate local mechanical environments.

3.1.4.4 Mechanically Driven Bone Adaptation algorithm

The mechanostat based load adaptation algorithm (Schulte et al., 2013) was implemented using a combination of C++ and Python. For the finite element analysis, the parOsol wrapper was used. For the convolution of the adaptation solver, a C++ parallel convolution was implemented. The rest of the simulation was implemented in Python.

3.1.5 Limitations

However, this project is not without limitations. For once, the software has not been developed with a clear focus on open-sourcing it. Therefore, some extra effort has to be put into the software to make it ready for a public repository. However, this is also a chance to revise some design decisions with the experience gained during development. Another limitation is the lack of a complementing database such as the AiiDA Python project (Pizzi et al., 2016). The integration of such a database would have been beyond the scope of this project. Nevertheless, it is an important tool to promote reproducibility and data sharing.

3.1.6 Conclusion

The goal of this project was to develop a computational framework that unifies the way the bone community can collaborate. We have demonstrated, how our approach framework provides easy access to common existing bone analysis and simulation tools while offering all the power of Python to develop future tools.

References

- Anaconda, 2020. Anaconda Software Distribution [WWW Document]. URL <https://anaconda.com>
- Christen, P., Ito, K., Knippels, I., Müller, R., van Lenthe, G.H., van Rietbergen, B., 2013. Subject-specific bone loading estimation in the human distal radius. *J. Biomech.* 46, 759–766. <https://doi.org/10.1016/j.jbiomech.2012.11.016>
- Doube, M., Klosowski, M.M., Arganda-Carreras, I., Cordelières, F.P., Dougherty, R.P., Jackson, J.S., Schmid, B., Hutchinson, J.R., Shefelbine, S.J., 2010. BoneJ: Free and extensible bone image analysis in ImageJ. *Bone* 47, 1076–1079. <https://doi.org/10.1016/j.bone.2010.08.023>
- Flaig, C., 2012. A highly scalable memory efficient multigrid solver for μ -finite element analyses. ETH Zurich. <https://doi.org/10.3929/ethz-a-007613965>
- Grecco, H.E., 2020. Pint [WWW Document]. URL <https://pypi.python.org/pypi/Pint>
- Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jakob, W., Rhineland, J., Moldovan, D., 2017. pybind11 - Seamless operability between C++11 and Python.
- Martín, A., Ashish, A., Paul, B., Eugene, B., Zhifeng, C., Craig, C., Greg S., C., Andy, D., Jeffrey, D., Matthieu, D., Sanjay, G., Ian, G., Andrew, H., Geoffrey, I., Michael, I., Jia, Y., Rafal, J., Lukasz, K., Manjunath, K., Josh, L., Dandelion, M., Rajat, M., Sherry, M., Derek, M., Chris, O., Mike, S., Jonathon, S., Benoit, S., Ilya, S., Kunal, T., Paul, T., Vincent, V., Vijay, V., Fernanda, V., Oriol, V., Pete, W., Martin, Wattenberg, Martin, Wicke, Yuan, Y., Xiaoqiang, Z., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [WWW Document]. URL <https://www.tensorflow.org/>
- Mennel, U., Sala, M., 2006. Overview of parFE : A Scalable Finite Element Solver for Bone Modeling.
- Müller, R., Van Campenhout, H., Van Damme, B., Van Der Perre, G., Dequeker, J., Hildebrand, T., Rügsegger, P., 1998. Morphometric analysis of human bone biopsies: A

- quantitative structural comparison of histological sections and micro-computed tomography. *Bone* 23, 59–66. [https://doi.org/10.1016/S8756-3282\(98\)00068-4](https://doi.org/10.1016/S8756-3282(98)00068-4)
- Ohs, N., Collins, C.J., Atkins, P.R., 2020. Validation of HR-pQCT against Micro-CT for Morphometric and Biomechanical Analyses: A Review. *Bone Reports* 100711. <https://doi.org/10.1016/j.bonr.2020.100711>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Research}, journal={*Journal of M.L.*, 2011. Scikit-learn. *J. Mach. Learn. Res.* 12, 2825--2830.
- Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N., Kozinsky, B., 2016. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* 111, 218–230. <https://doi.org/10.1016/j.commatsci.2015.09.013>
- Python Software Foundation, 2020. Python Language Reference, Version 3.7 [WWW Document]. URL <https://www.python.org>
- Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2011. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. *Bone* 48, 433–442. <https://doi.org/10.1016/j.bone.2010.10.007>
- Schulte, F.A., Zwahlen, A., Lambers, F.M., Kuhn, G., Ruffoni, D., Betts, D., Webster, D.J., 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. <https://doi.org/10.1016/j.bone.2012.09.008>
- The HDF Group, 2020. Hierarchical Data Format, version 5 [WWW Document]. URL <http://www.hdfgroup.org/HDF5/>
- Thévenaz, P., Ruttimann, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41. <https://doi.org/10.1109/83.650848>
- Tourolle né Betts, D.C., Wehrle, E., Paul, G.R., Kuhn, G.A., Christen, P., Hofmann, S., Müller, R., 2020. The association between mineralised tissue formation and the mechanical

local in vivo environment: Time-lapsed quantification of a mouse defect healing model. *Sci. Rep.* 10, 1–10. <https://doi.org/10.1038/s41598-020-57461-5>

Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. <https://doi.org/10.1109/MCSE.2011.37>

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Contributors, S. 1. 0, 2019. SciPy 1.0--
Fundamental Algorithms for Scientific Computing in Python 1–22.

3.2 Towards *in silico* prognosis using big data

N. Ohs¹, Y.-W. W. Lee², C.-Y. J. Cheng², Peter Arbenz³, Ralph Müller¹, Patrik Christen¹

¹Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

²The Chinese University of Hong Kong, Department of Orthopaedics & Traumatology, Prince of Wales Hospital, Shatin, NT, Hong Kong SAR, China

³Computer Science Department, ETH Zurich, Zurich, Switzerland

Published in:

Current Directions in Biomedical Engineering, Volume 2, Issue 1 (2016)

Postprint version according to publisher copyright policy

Abstract

Clinical diagnosis and prognosis usually rely on few or even single measurements despite clinical big data being available. This limits the exploration of complex diseases such as adolescent idiopathic scoliosis (AIS) where the associated low bone mass remains unexplained. Observed low physical activity and increased RANKL/OPG, however, both indicate a mechanobiological cause. To deepen disease understanding, we propose an *in silico* prognosis approach using clinical big data, i.e. medical images, serum markers, questionnaires and live style data from mobile monitoring devices and explore the role of inadequate physical activity in a first AIS prototype. It employs a cellular automaton (CA) to represent the medical image, micro-finite element analysis to calculate loading, and a Boolean network to integrate the other biomarkers. Medical images of the distal tibia, physical activity scores, and vitamin D and PTH levels were integrated as measured clinically while the time development of bone density and RANKL/OPG was observed. Simulation of an AIS patient with normal physical activity and patient-specific vitamin D and PTH levels showed minor changes in bone density whereas the simulation of the same AIS patient but with reduced physical activity led to low density. Both showed unchanged RANKL/OPG and considerable cortical resorption. We conclude that our

integrative *in silico* approach allows to account for a variety of clinical big data to study complex diseases.

Keywords:

Clinical big data; cellular automaton; Boolean network; micro-finite element analysis; adolescent idiopathic scoliosis

3.2.1 Introduction

Understanding the physiology of the human body is a tremendous challenge, understanding and finally treating diseases is arguably an even more difficult task. Many diseases exhibit alterations in complex signalling pathways that lead to distinct phenotypes used in diagnosis and prognosis. However, both usually rely on few or even single clinical measurements despite clinical big data, i.e. medical images, serum markers, questionnaires and live style data from mobile monitoring devices being available. Exploiting these clinical big data would potentially enable one explore and thus further understand complex diseases such as adolescent idiopathic scoliosis (AIS) where many alterations in the signalling cascade and other factors such as a patient's lifestyle play an essential role.

AIS is a prevalent spinal deformity associated with systemic low bone mass (Z -score ≤ -1 with the reference to local matched population), which can persist into adulthood predisposing to osteoporosis in later life. If not treated properly, the curve will deteriorate leading to cardiopulmonary compromise, back pain, degenerative spine disease, and psychosocial disorder (Cheng et al., 2015). Low bone mass has been identified as a risk factor for curve progression. The underlying mechanism of low bone mass in AIS, however, remains unknown although it would potentially provide a basis for counteracting the associated complications (Yu et al., 2014). It has been suggested that low bone mass is related to inadequate physical activity (Lee et al., 2005) and an increase in RANKL/OPG (Cheng et al., 2015), which might lead to increased bone resorption. Most recently, low vitamin D levels were observed in AIS patients (Batista et al., 2014) indicating that multiple factors play a role in AIS.

Therefore, we propose a novel prognosis approach using clinical big data for the *in silico* simulation of complex diseases with the ultimate goal to improve clinical diagnosis and prognosis of these diseases as well as the evaluation of treatment options. Medical images containing bone tissue density are integrated into a cellular automaton (CA) model where the tissue densities define local states. These states are updated according to a local rule that takes into account the biological cells present in each CA cell. The cell number or activity is normalised representing a normal healthy value by default but with the possibility to either initialise patient-specific or experimental values. It allows to integrate serum markers of specific cells if they are available and at the same time does not require the definition of a value if it is not known. The cells itself are updated through mechanical tissue loading as

calculated with micro-finite element (micro-FE) analyses and a molecular factor determined in a Boolean network, which models the interaction of cells, molecules, and mechanical loading. The entire procedure is applied to obtain cell and tissue densities as well as molecular states at different time points that describe the time evolution of the system. Here, we describe and formulate the theoretical basis of this integrative *in silico* prognosis approach and present a first prototype demonstrating the integration of medical image, serum marker, and lifestyle data for AIS.

The purpose of this study is to formulate a theory for integrating varying clinical big data, and implement and employ it to simulate low bone mass typical of AIS.

3.2.2 Materials and Methods

We performed *in silico* simulations over a period of 6 months integrating high-resolution peripheral quantitative computed tomography (HR-pQCT) images of the distal tibia, physical activity scores, and vitamin D and PTH levels of an AIS patient. Two simulations were run including an AIS control case with normal physical activity score, and vitamin D and PTH levels as measured in the patient's serum, and an AIS case with reduced physical activity as observed in other studies but again with the same patient-specific measurements as in the control simulation.

3.2.2.1 Clinical patient assessment

To validate the proposed *in silico* prognosis, a 17-year-old AIS girl (patient number 666; major Cobb angle 37°) was randomly selected from the database of a prospective randomised controlled trial who received observation alone. Anthropometry, curve severity, pubertal assessment and bone parameters were assessed at baseline and at the 12-month time-point. Bone parameters were measured with dual-energy X-ray absorptiometry (DXA, XR-36; Norland Medical Systems, Fort Atkinson, WI, USA) and HR-pQCT (XtremeCT, Scanco Medical AG, Switzerland).

Blood was collected on the same date of clinical visit for the measurement of serum bone metabolic markers with standard protocols conducted by the University Pathology Unit. Physical activity was assessed by a standardised quantitative questionnaire adapted from

Slemenda et al. (1991). The amount of weight bearing activity is expressed as number of hours per day. The sum of hours spent on various types of weight bearing activity was used to summarise daily activities for each subject. Ethical approval was obtained from the University Clinical Research Ethics Committee (Ref. NO.; CRE-2008.054-T) (Lam et al., 2013).

3.2.2.2 Cellular automaton

The medical image of a bone, e.g. a HR-pQCT scan, is a 3D representation of the bone density in a cubic domain at a certain voxel resolution. It can be regarded as the regular grid, G , of cells of a CA. For each CA cell and a given time point, t , a density value, a strain energy density (SED) value, the number or activity of the different biological cells and the presence or absence of several molecules usually found in bone are defined:

$$g \in G, \quad g(t) = (\rho(t), \epsilon(t), n_0(t), n_1(t), \dots, n_N(t), m_0(t), m_1(t), \dots, m_M(t)),$$

with ρ being the tissue density value, ϵ being the SED value, N the number of cell types considered, n_i the number or activity of biological cells of type i inside each voxel, M the number of molecule types considered and m_j indicating whether the molecule of type j is present or absent. This tuple makes up the state of this CA cell at time t . Our first prototype uses three cell types: osteoblasts (n_{OBL}) forming new bone tissue and osteoclasts (n_{OCL}) removing old bone tissue resulting in a continuous remodelling of the tissue, and osteocytes (n_{OCY}) considered to respond to mechanical tissue loading and orchestrate the action of osteoblasts and osteoclasts accordingly.

In order to update the CA, rules are defined that are applied to each CA cell simultaneously and considering a certain neighbourhood. For each $g \in G$, $N(g)$ describes the set of all considered neighbours of g . In the presented prototype, a 3D von Neumann neighbourhood is defined. Finally, an update rule f that generates the new bone cells and tissue densities in the CA is given by

$$g(t + 1) = f(g(t), N(g(t)))$$

In our prototype AIS CA, the update rule is the same for every CA cell. Our prototype employs the following update rule: The new SED values for each voxel computed with the micro-FE method is an input. In a first update step, the cell numbers based on a function that transforms SED values to an increase or decrease of OBL/OCL are modified according to the SED-bone

remodelling relationship found earlier (Christen et al., 2016). To avoid infinite increase/decrease of cells, they are saturated at a minimum and maximum level for both, OBL and OCL. In a second step, the Boolean network for each voxel is evaluated, which will again result in modifications to the OBL/OCL counts. Finally, the tissue density is updated based on the OBL/OCL counts. If the density values are higher than bone, they are distributed to the neighbouring voxels. The initial bone cell count and distribution are determined in a preceding iteration where more OBLs are placed in regions of high SED and more OCLs in regions of low SED. The CA is only updated a few layers away from the top and bottom where the boundary conditions of the micro-FE analysis are applied to exclude boundary artefacts.

3.2.2.3 Boolean network

The interaction of the local molecules with the local biological cells under the influence of mechanical loading in each CA cell is modelled with a Boolean network. Each molecule is either present in a CA cell or not. The number of CA cells in which a specific molecule is present, is governed by the measurement of the appropriate serum marker. In the present AIS network, RANKL, OPG, vitamin D, PTH, OCL, OBL and mechanical tissue loading (Mech) are integrated. The corresponding Boolean functions are as follows:

$$RANKL = OBL \wedge (\neg Mech \vee VitD \vee PTH)$$

$$OPG = OBL \wedge Mech \wedge (\neg PTH \vee \neg VitD)$$

All molecules are either present or not. They are represented by a Boolean value. Mech is a thresholded Boolean value of ϵ . The cell numbers/activities of OBL/OCL are also thresholded. The molecule states are initialised randomly with a spatially uniform distribution and an average cover rate of 50%. Molecules that are never updated in the network are redistributed before every time iteration to avoid bias due to the specific configuration that was generated for the first iteration. OBL is then increased if PTH is true whereas OCL is increased if RANKL is true.

3.2.2.4 Computational implementation

The presented modelling approach is well suited for parallel implementation since each CA cell only depends on its nearest neighbourhood but is propagated independently with each time step. We therefore implemented the prototype for the execution on a graphics processing unit (GPU)

using OpenCL. This allows to exploit the large amount of computing units available on a GPU compared to only a few cores of a CPU on a normal computer workstation.

The analysis is performed with the dedicated micro-FE solver ParOSol (Flaig et al., 2011) on the CPU. Boundary conditions are defined according to a bone loading estimation algorithm (Christen et al., 2016), providing physiological *in vivo* loading for this particular patient. They include compression (zz-direction) and shear strains (zx- and zy-direction). For the CA, biological cell and tissue density, molecular states as well as the mechanical tissue loading are stored for the full CA domain and thus a so called full-space approach is followed. Tissue density and loading are stored as floating point numbers. Since GPU memory is very limited, the biological cell count/activity per voxel is restricted to 8-bit characters, which still provides 256 quantification levels. For each molecule, only 1 bit is needed in the Boolean network, thus, for each voxel all molecular information can be encoded in a single byte.

3.2.3 Results

In the control simulation, bone density stayed constant with only a minor change of 0.03% between start and end point, representing normal healthy bone remodelling with 1.40% formation and 1.43% resorption. Reducing physical activity led to a decrease of 20.65% in bone density with 1.40% and 22.05% bone formation and resorption, respectively (Figure 3.2.1). In both cases, RANKL/OPG did not change during the simulated time.

Calculating tissue loading with micro-FE analysis required 494 seconds on a supercomputer (96 CPU cores, CSCS Cray XC40) while the CA computation including the Boolean network required 45 seconds on a workstation GPU (AMD Radeon HD 6750M 1024 MB, MacBook Pro 2011).

3.2.4 Discussion

We here propose an *in silico* prognosis approach to combine varying clinical measurements and thus clinical big data to explore complex diseases. The theoretical basis is outlined as well as a first prototype simulating AIS related bone loss due to inadequate physical activity is presented. We successfully demonstrate this effect by feeding clinical big data including medical images,

serum markers, and physical activity levels into the *in silico* prognosis. Mobile monitoring data could potentially be included, too.

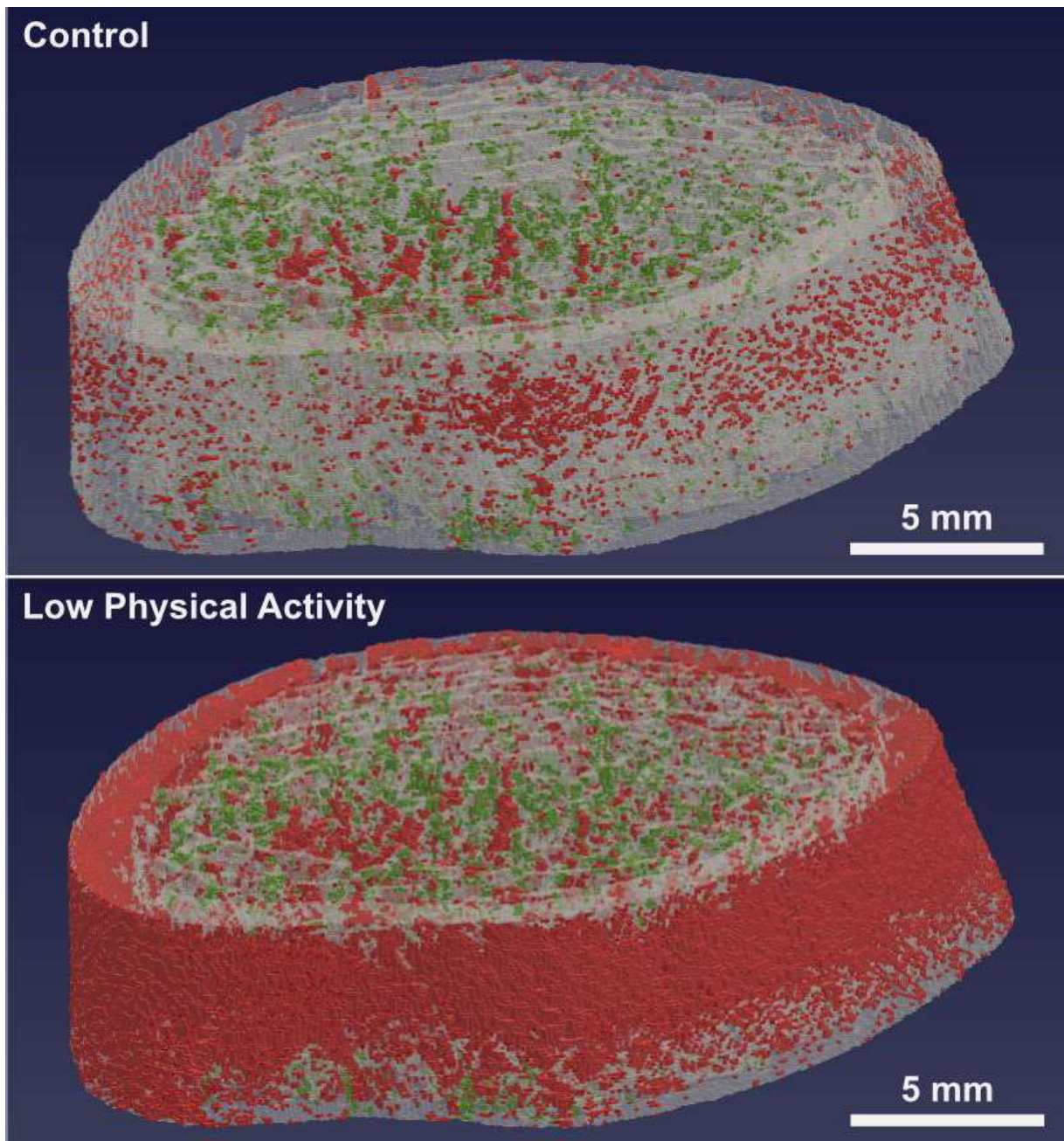


Fig. 3.2.1 Differences in bone density between the first and last iteration for a control and a low physical activity simulation of an AIS patient. Bone resorption is depicted in red, formation in green, and no change in white transparent.

Our results are in agreement with bone's capability to adapt to mechanical loading since bone density decreased with decreasing physical activity. The *in silico* method furthermore allows a

more local analysis of bone loss, showing that resorption occurred predominantly in cortical bone and the centre of trabecular bone indicating alterations in spatial bone remodelling activity that might be linked to the local mechanical loading conditions. Bone formation, in contrast, occurred more evenly distributed throughout the trabecular bone.

The current implementation is very fast with only a few seconds execution time. However, micro-FE calculations were performed on a supercomputer and are currently coupled through a shared results file with the CA requiring additional file reading/writing. Computation time could thus be reduced by including the micro-FE solver in the CA. Although the Boolean networks are modelled for each image voxel, they do not add much to the total computation time and thus, they might be extended to include several quantification levels similar to how the biological cell counts are implemented.

In conclusion, the present *in silico* prognosis method allows to account for a variety of clinical measurements to study complex diseases using the concept of big data.

Acknowledgements

This work has been supported by the Holcim Stiftung for the Advancement of Scientific Research and the Swiss National Supercomputing Centre (CSCS). 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.

Conflict of Interest

Authors state no conflict of interest.

References

- Batista, R., Martins, D.E., Hayashi, L.F., Lazaretti-Castro, M., Puertas, E.B., Wajchenberg, M., 2014. Association between vitamin D serum levels and adolescent idiopathic scoliosis. *Scoliosis* 9, 7161. <https://doi.org/10.1186/1748-7161-9-s1-o45>
- Cheng, J.C., Castelein, R.M., Chu, W.C., Danielsson, A.J., Dobbs, M.B., Grivas, T.B., Gurnett, C.A., Luk, K.D., Moreau, A., Newton, P.O., Stokes, I.A., Weinstein, S.L., Burwell, R.G., 2015. Adolescent idiopathic scoliosis. *Nat. Rev. Dis. Prim.* 1, 15030. <https://doi.org/10.1038/nrdp.2015.30>
- Christen, P., Ito, K., Ellouz, R., Boutroy, S., Sornay-Rendu, E., Chapurlat, R.D., Van Rietbergen, B., 2014. Bone remodelling in humans is load-driven but not lazy. *Nat. Commun.* 5. <https://doi.org/10.1038/ncomms5855>
- Christen, P., Schulte, F.A., Zwahlen, A., van Rietbergen, B., Boutroy, S., Melton, L.J., Amin, S., Khosla, S., Goldhahn, J., Müller, R., 2016. Voxel size dependency, reproducibility and sensitivity of an in vivo bone loading estimation algorithm. *J. R. Soc. Interface* 13, 20150991. <https://doi.org/10.1098/rsif.2015.0991>
- Flaig, C., 2012. A highly scalable memory efficient multigrid solver for μ -finite element analyses. ETH Zurich. <https://doi.org/10.3929/ethz-a-007613965>
- Lam, T.P., Ng, B.K.W., Cheung, L.W.H., Lee, K.M., Qin, L., Cheng, J.C.Y., 2013. Effect of whole body vibration (WBV) therapy on bone density and bone quality in osteopenic girls with adolescent idiopathic scoliosis: A randomized, controlled trial. *Osteoporos. Int.* 24, 1623–1636. <https://doi.org/10.1007/s00198-012-2144-1>
- Lee, W.T.K., Cheung, C.S.K., Tse, Y.K., Guo, X., Qin, L., Ho, S.C., Lau, J., Cheng, J.C.Y., 2005. Generalized low bone mass of girls with adolescent idiopathic scoliosis is related to inadequate calcium intake and weight bearing physical activity in peripubertal period. *Osteoporos. Int.* 16, 1024–1035. <https://doi.org/10.1007/s00198-004-1792-1>
- Slemenda, C.W., Miller, J.Z., Hui, S.L., Reister, T.K., Johnston, C.C., 1991. Role of physical activity in the development of skeletal mass in children. *J. Bone Miner. Res.* 6, 1227–1233. <https://doi.org/10.1002/jbmr.5650061113>

Yu, W.S., Chan, K.Y., Yu, F.W.P., Ng, B.K.W., Lee, K.M., Qin, L., Lam, T.P., Cheng, J.C.Y., 2014. Bone structural and mechanical indices in Adolescent Idiopathic Scoliosis evaluated by high-resolution peripheral quantitative computed tomography (HR-pQCT). *Bone* 61, 109–115. <https://doi.org/10.1016/j.bone.2013.12.033>

3.3 Upscaling of *in vivo* HR-pQCT images enables accurate simulations of human microstructural bone adaptation

N. Ohs¹, D. C. Tourolle né Betts¹, P. R. Atkins¹, S. Sebastian¹, B. van Rietbergen², M. Blauth³, P. Christen^{1,4}, R. Müller¹

¹Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

²Orthopaedic Biomechanics, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

³Department for Trauma Surgery, Innsbruck University Hospital, Innsbruck, Austria

⁴Institute for Information Systems, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

Under revision in:

Biomechanics and Modeling in Mechanobiology

Preprint version on BioRxiv:

DOI: 10.1101/2020.05.13.093328

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 upscaling 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 upscaling 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:

Upscaling; HR-pQCT; Simulation; Mechanoregulation; Microstructure; Bone Adaptation

3.3.1 Introduction

Simulations are considered the third pillar of modern science next to models and experiments. In the biomedical field the creation of a virtual physiological human (VPH) is seen as one of the most important goals (Fenner et al. 2008). The vision of the VPH is to provide researchers with a model that allows rapid hypothesis testing via *in silico* trials and provides doctors with a virtual patient as a decision-support system for their daily work (Viceconti and Hunter 2016). The role of bone, both structurally and physiologically, indicates that a validated model for microstructural bone adaptation and (re)modelling is a significant component to any VPH model. Previous studies have shown that various aspects of bone diseases and their treatments can be simulated in pre-clinical models (Müller 2005; Ruimerman et al. 2005; Schulte et al. 2013; Levchuk et al. 2014). Importantly, the advection based model by Adachi et al. (2001) can produce results comparable to population data when *ex vivo* images are used as an input (Badilatti et al. 2016). However, the translation of microstructural simulations to clinical image data has largely been constrained by the availability of high quality images and validation data.

With the introduction and increased use of HR-pQCT, large amounts of clinically relevant data have been gathered which provide the basis to validate and parameterise *in silico* models of bone (Boutroy et al. 2005, 2008; Sornay-Rendu et al. 2007, 2017; Kirmani et al. 2009; Nicks et al. 2012; Nishiyama et al. 2013; Nishiyama and Shane 2013; Yu et al. 2014; Zhu et al. 2014). However, combining current microstructural bone adaptation simulations with HR-pQCT is non-trivial. Existing simulations either utilize synthetic images (Ruimerman et al. 2005) or high-resolution micro-CT images which cannot be obtained clinically (Müller 2005; Schulte et al. 2013; Levchuk et al. 2014; Badilatti et al. 2016). Furthermore, HR-pQCT images tend to have more noise (Rajapakse et al. 2009) and other potential imaging artefacts, such as those due to movement (Pialat et al. 2012).

The reduction in resolution is a known obstacle for the translation of computational techniques from the lab into the clinical setting (Tjong et al. 2012; Manske et al. 2015; Christen et al. 2016). Alsayednoor et al. (2018), for example, showed that applying a single threshold to HR-pQCT images cannot yield both correct morphometric indices and mechanical properties on HR-pQCT images. However, *in silico* microstructural bone adaptations rely on having correct digital representations of morphometrics and mechanics as the simulation couples these two

properties. To study the effects of image resolution in detail, convergence studies have been performed which showed a clear dependence on image resolution for various mechanical and morphometric parameters (Müller et al. 1996; Christen et al. 2016). In order to counteract this algorithmic dependence on image resolution, upscaling low-resolution images to desktop micro-CT images on which the algorithms have been validated can help to produce accurate results (Rajapakse et al. 2009; Schulte et al. 2019). For example, upscaling of magnetic resonance imaging (MRI) data has been shown to produce micro-FE results in good agreement with those from micro-CT images of a higher resolution (Rajapakse et al. 2009). While it is clear that upscaling does not yield the same effects as scanning at a higher resolution (Cooper et al. 2007), as upscaling 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.

In previous studies using the algorithm by Adachi et al. (2001), initial iterations, which showed aberrant results, were regarded as part of the model initialization and excluded from analysis (Schulte et al. 2013; Badilatti et al. 2016). However, the exclusion of the initial iterations would lead to an inherent divergence between the clinical *in vivo* and *in silico* baseline models, which thus precludes direct comparison of simulation outcomes against clinically observed changes in bone microstructure, and hinders validation of the *in silico* model. 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 et al. (2001), and are related to mismatches between experimental input data and simulation parameters (Balmaseda and Anderson 2009; Mulholland et al. 2015). 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 goal of the present study was to determine if microstructural simulations of *in vivo* human HR-pQCT images based on the approach by Adachi et al. (2001) yield accurate results, such that they could be used as part of a VPH. Since high-resolution ground truth micro-CT images cannot be obtained *in vivo*, we applied an *ex vivo* approach to study the resolution dependence

and the effects of upscaling on morphometric accuracy over the course of a microstructural bone adaptation simulation. However, we first had to address two initialisation issues: initialisation shocks and disagreement of mechanics and morphometrics in model inputs at HR-pQCT resolution. Thus, we developed an input regularisation approach to reduce initialisation shocks observed in microstructural bone adaptation simulations and studied upscaling as a method to improve the accuracy of model inputs with respect to both mechanics and morphometry. Finally, we compared our *in silico* results to simulations run on *in vivo* images to investigate whether additional *in vivo* image artefacts affect simulation outcomes.

3.3.2 Materials

3.3.2.1 High resolution *ex vivo* micro-CT images

Five distal radii were obtained from female cadavers at the Amsterdam Medical Center as part of a previous study (De Jong et al. 2016). The donors' ages varied between 58 and 95 years and the bone volume fraction (BV/TV) of the samples varied from 7 to 20% where BV/TV was inversely related with age. The medical history of the cadaveric specimens was unknown. High resolution CT images were obtained at an isotropic voxel-size of 25 μm with a vivaCT 80 (70 kV, 114 μA , 300 ms integration time), a micro-CT device by Scanco Medical AG (Switzerland). Images were Gauss-filtered ($\sigma = 1.2$, support = 1) and the trabecular region was hand-contoured by a trained operator for each scan using the software of the scanner manufacturer.

3.3.2.2 *In vivo* HR-pQCT images

Five patients (four female, one male) were recruited at Innsbruck Medical University as part of a radius fracture study. Patients provided informed consent and participated in a study approved by the ethics committee of the Medical University of Innsbruck. The age of the patients ranged from 26 to 80 years and their BV/TV from 21 to 7%. In this study, images of the unfractured, contralateral radius were used. Scans were performed with an isotropic voxel-size of 61 μm using an XtremeCT II (68 kV, 1470 μA , 43 ms integration time), a clinical HR-pQCT device by Scanco Medical AG (Switzerland). Images were Gauss-filtered ($\sigma = 1.2$, support = 1) and the trabecular region was hand-contoured by a trained operator using the software of the scanner manufacturer.

3.3.2.3 Generation of low-resolution images

The high-resolution *ex vivo* micro-CT grey-scale images were downsampled to resolutions of 40, 61, and 82 μm . These three resolutions will be referred to as low-resolutions in this paper. Currently, 61 and 82 μm are the highest resolutions available for clinical CT scanners (XtremeCT I and II, Scanco Medical AG, Switzerland), these resolutions will be referred to as the clinically relevant resolutions. Resizing was performed using the scikit-image (Van Der Walt et al. 2014) rescale function in Python (Python Software Foundation 2020) with third-order interpolation and anti-aliasing enabled. The binary hand-contoured masks for the high-resolution images were converted to a floating-point data-type and resized like the micro-CT images. Finally, a threshold of 50% of the maximum image value was applied to obtain binary masks for the low-resolution images.

3.3.2.4 Generation of upscaled images

Upscaled images were created from the low-resolution images by applying the scikit-image `resize` Python function, again with third-order interpolation and anti-aliasing enabled, to a resolution of 25 μm . When creating the upscaled images, the conversion of image dimensions between the different resolutions was not unique (i.e. due to rounding of the integer image dimensions after scaling with a floating point number, differences in size of one voxel could occur). To ensure that the upscaled images have the exact same dimensions as the original images, the scikit image `resize` function was used. The `resize` function is identical to the `rescale` function with the exception that it resizes images to a target image dimension instead of resizing using a target scaling factor. For the upscaled images, the same hand-contoured masks for the trabecular region were used as for the high-resolution *ex vivo* micro-CT grey-scale images to avoid influences of differences in masks on our results. Herein, the images upscaled from 40, 61, and 82 μm to 25 μm are referred to as u40 μm , u61 μm , and u82 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.

3.3.3 Methods

3.3.3.1 Remodelling simulations

3.3.3.1.1 Micro-FE analysis

For the micro-FE analysis, we used the parallel octree solver parOsol (Flaig 2012) 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. (2013). 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 (Christen et al. 2016). 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 (Christen et al. 2013). For all experiments, we computed the load-estimation using the high-resolution files and applied the same loading conditions to the low-resolution and upscaled 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).

3.3.3.1.2 Remodelling algorithm

The strain-adaptive *in silico* microstructural bone adaptation simulation by Adachi et al. (2001) was re-implemented in Python using NumPy (Van Der Walt et al. 2011) and pybind11 (Jakob et al. 2017). 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.

3.3.3.1.3 Binary model generation from micro-CT data

The high-resolution images were segmented using a threshold of 450 mg HA/cm³, which has also been chosen in previous studies on this dataset (Müller et al. 1996; Christen et al. 2016), 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 / cm³ and background voxels were set to 0 mg HA / cm³.

3.3.3.1.4 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 3.3.1, 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 / cm³).

3.3.3.1.5 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.

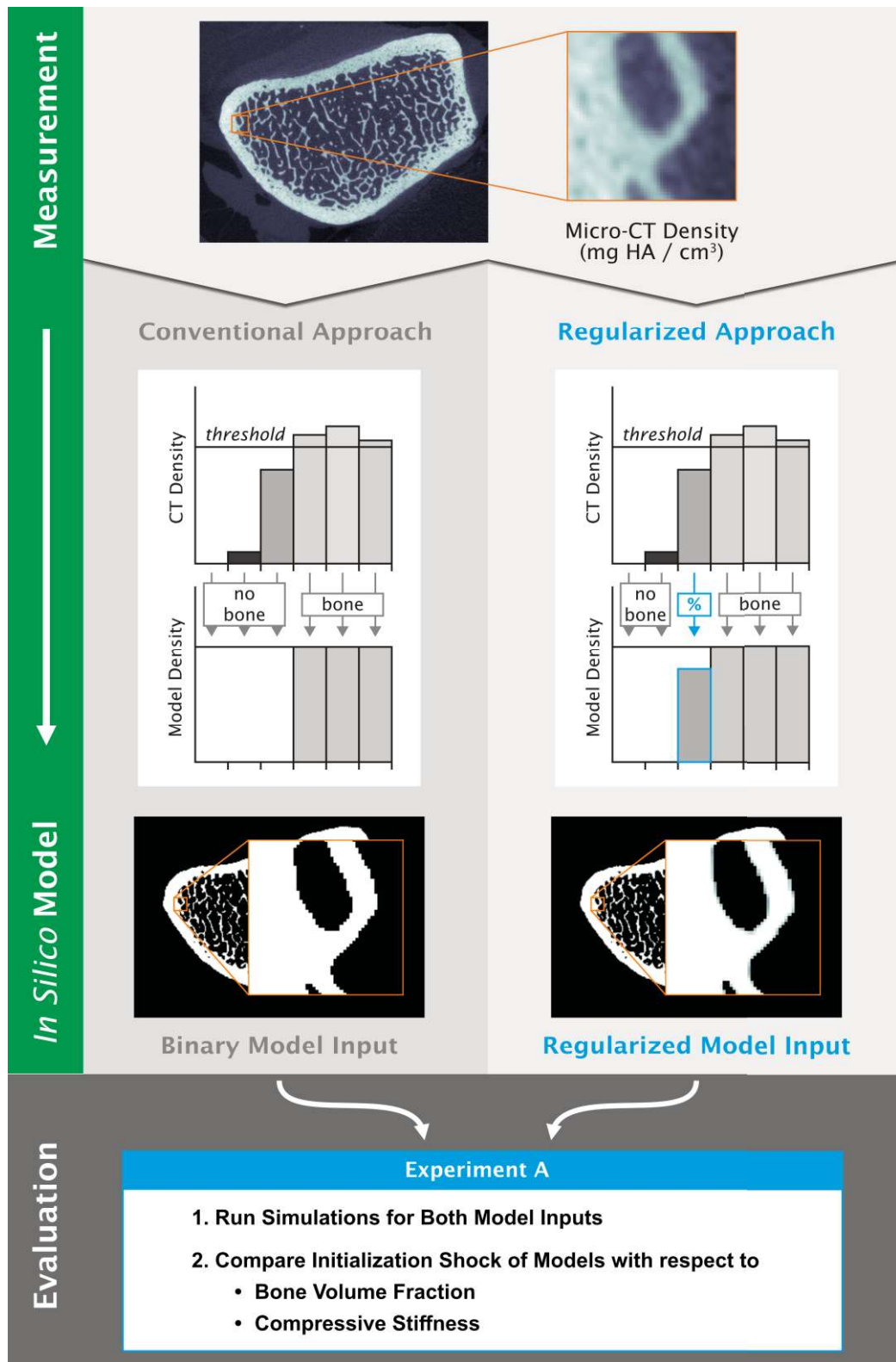


Fig. 3.3.1 Experiment A (Effect of regularized input models on Initialisation shock) evaluated the reduction of initialisation shock behaviour in bone volume fraction and compressive stiffness due to a novel input regularization approach compared to the conventional input approach. Left: Illustration of the conventional threshold approach used to convert a CT image into a valid input for the load adaptation simulation. Right: Regularization approach.

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

3.3.3.2.1 Experiment A: Effect of regularized input models on initialisation shock

Two simulation input model generation approaches were compared (Figure 3.3.1), the current state of the art through binary representation and a regularized model with partially filled voxels at the surface. Both models were generated from the high resolution *ex vivo* micro-CT image data set. The partially filled voxels approximate a bone surface with sub-voxel precision. Microstructural bone adaptation simulations were run and the discontinuity in BV/TV and compressive stiffness for the initial simulation steps were quantified for both methods.

3.3.3.2.2 Experiment B: Effect of upscaling on mechanical and morphometric accuracy

For experiments B, regularized input models were produced for all three micro-CT image sets: high-resolution, low-resolution, and upsampled low-resolution images. In the following, these are called reference, low-resolution, and upsampled regularized input models. For all images, the

reference BV/TV was always the one obtained from the respective binary high-resolution image.

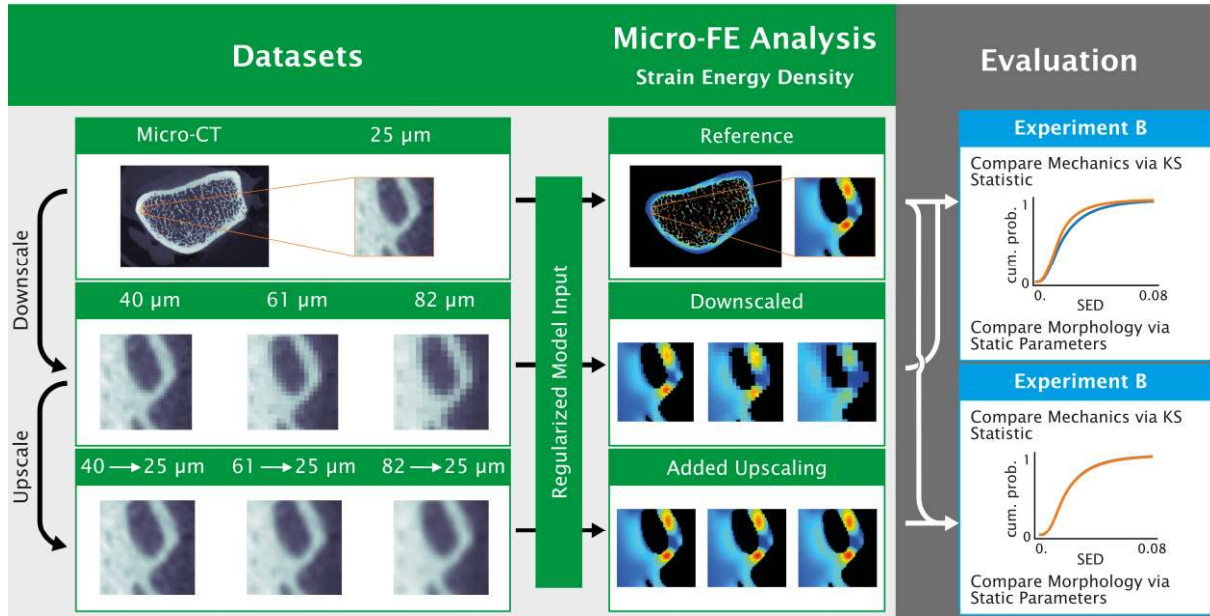


Fig. 3.3.2 Experiment B (Effect of upscaling on mechanical and morphometric accuracy) evaluated the mechanical agreement between the different resolutions for matched BV/TV. For each micro-CT image, high-, low-resolution, and upscaled images were created (left column). Low-resolution: 40 µm, 61 µm, and 82 µm and upscaled images: Same downsampled resolutions as the low-resolution image but upscaled back to 25 µm. Micro-finite-element (micro-FE) models were generated using the regularization model generation approach of experiment A (Figure 3.3.1) and matching bone volume fraction (BV/TV) for each image to the reference (25 µm). A micro-FE analysis was run and strain energy density (SED) (shown in the jet colour-map) and static parameters were computed for each model. SED distributions between low and upscaled resolution images were compared using the Kolmogorov-Smirnov statistic to see which one more closely matches the reference high-resolution regularized input model mechanically. Mean static parameters were compared to see which one more closely matches the reference model morphologically.

To quantify the agreement in mechanical properties between reference, low-resolution, and upscaled regularized input models of the same bone structure, micro-FE analyses were performed on the data set at each of these resolutions, respectively (Figure 3.3.2). SED distributions, mean SED, mean static parameters (BV/TV, trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular spacing (Tb.Sp), and structural model index (SMI)), and standard deviations for all mean values were computed for comparison between the three regularized

input model types. Additionally, the Kolmogorov-Smirnov statistic was computed between the SED of each sample of the low-resolution and upscaled regularized input models and the corresponding reference regularized input model to quantify mechanical agreement. Finally, the adequacy of the chosen threshold was assessed through comparison of the SED distribution from FE analysis of the low-resolution regularized input models for a range of thresholds (575 - 775 mg HA/cm³). The magnitudes of the peak SED were compared to the reference SED distribution.

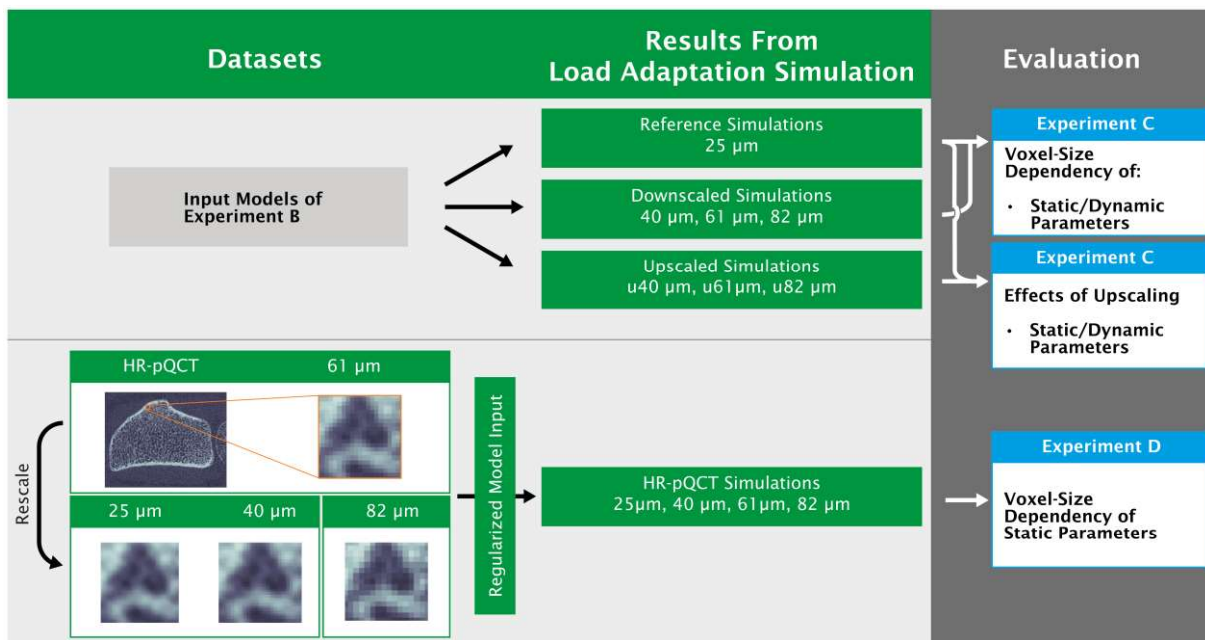


Fig. 3.3.3 Overview of experiments C (Effect of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation) and D (Effect of *in vivo* image artefacts on convergence of upscaled HR-pQCT simulation). Top: Regularized input models from experiment B were taken and microstructural bone adaptation simulations were run for all input models. Bottom: A separate HR-pQCT dataset was also converted to regularized input models and simulations were run for all input models. For experiment C, static parameters and dynamic parameters were computed. The goal of experiment C was to compare how accurate low-resolution versus upscaled resolution simulations were relative to the reference simulations. For experiment D, static parameters were computed as well. The goal of experiment D was to compare the voxel size dependency of down-scaled micro-CT and rescaled HR-pQCT images to determine if additional artefacts introduced by the HR-pQCT images had a strong influence on the outcome of the bone-adaptation simulation.

3.3.3.2.3 Experiment C: Effect of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation

Microstructural bone adaptation simulations were run on the input models generated for experiment B. The simulations run on the reference regularized input models were considered the best approximations of the *in vivo* remodelling process and were used as reference to quantify errors. These simulations are referred to as the reference simulations (Figure 3.3.3). Static parameters (BV/TV, Tb.N, Tb.Th, Tb.Sp, and SMI) and formed and resorbed volume over time were computed.

3.3.3.2.4 Experiment D: Effect of *in vivo* image artefacts on convergence of upscaled HR-pQCT simulation

Microstructural bone adaptation simulations were run on regularized input models of the *in vivo* HR-pQCT images and their rescaled version (Figure 3.3.3) using the same simulation parameters and calculating the same static parameters as in experiment C. The reference BV/TV for the regularized input model generation for all resolutions was based on the respective original HR-pQCT resolution binary structure.

Evaluation and statistics

All simulations and evaluations were performed within the trabecular mask. SED was evaluated for non-empty voxels. SED distributions were represented using the SciPy Gaussian kernel density approximation (Virtanen et al. 2019). NumPy (Van Der Walt et al. 2011) was used to compute the Kolmogorov-Smirnov (KS) statistic, mean and standard deviations, as well as BV/TV, which was computed through integration of the model density within the trabecular mask. All other static parameters (Tb.N, Tb.Th, Tb.Sp, SMI) were computed using the scanner manufacturer's image processing language (IPL) (Hildebrand et al. 1999). Before calling the IPL functions, models were up-scaled to 25 μm to remove the voxel-size dependency of IPL functions as a confounding factor. Finally, formed and resorbed volume over time was computed by integration of positive and negative density changes using NumPy.

Comparisons for experiment A were done using a paired Student t-test. For experiments B and C, to determine significance, two-way analysis of variance (two-way ANOVA) was performed for each measured parameter as an omnibus test with the two categorical groups: resolution and upscaling. If heteroscedasticity was detected using a Levene test, heteroscedasticity consistent

covariance matrices of type HC3 were used. Post-hoc group comparisons were done using paired Student t-tests and p-values were corrected using the Holm-Bonferroni correction for multiple comparisons. For experiment D, paired Student t-tests were performed and p-values were corrected using the Holm-Bonferroni correction for multiple comparisons. The level of significance was set to 0.05. For the Student t-tests and Levene tests, scipy 1.3.1 was used. The ANOVA was done using statsmodels (Seabold and Perktold 2010) 0.10.2.

3.3.4 Results

3.3.4.1 Experiment A: Effect of regularized input models on initialisation shock

The usage of regularized input models removed initialization shocks found in simulations run on conventional input models. For the conventional binary input models, the change in apparent compressive stiffness after the first iteration was a factor of 5.9 ± 0.8 larger than the maximum of all other iteration steps (Figure 3.3.4). For BV/TV, a factor of 2.3 ± 1.6 increase was observed in the first iteration (Figure 3.3.4). Visually, we observe that for some samples a clear shock in BV/TV was present. In contrast, for the first iteration of the regularized input model approach, the change in apparent compressive stiffness is indistinguishable from the rest of the simulation with a computed factor of 0.5 ± 0.5 increase, which is significantly lower than for the conventional method ($p < 0.001$). For BV/TV, we obtained a factor of 2.1 ± 0.6 increase (Figure 3.3.4) which is not significantly different from the threshold method. Since the regularized input model approach removed the initialization shock in apparent stiffness across all samples, we performed all other experiments using this approach.

The average regularization threshold for the high-resolution images was 563.6 ± 6.6 mg HA / cm^3 . For the low-resolution images, the average regularization thresholds were 601.8 ± 10.0 , 605.2 ± 14.7 , and 590.6 ± 18.9 mg HA / cm^3 for 40, 61, and 82 μm , respectively.

3.3.4.2 Experiment B: Effect of upscaling on mechanical and morphometric accuracy

For all tested thresholds, the SED distributions for the low-resolution regularized input models did not visually match the SED distribution of the reference regularized input models (Figure 3.3.5a). However, for BV/TV matched thresholds, the peaks of the SED distributions aligned

well. Qualitatively, the difference in SED distribution from the reference model increased with voxel-size (Figure 3.3.5b and 3.3.5c). The SED distributions for upscaled regularized input models were almost an order of magnitude closer to the reference than for the low-resolution regularized input models ($p < 0.05$ for all; Table 3.3.1). Specifically, deviations in mean SED for upscaled resolutions were less than 5% while deviations were up to 42% for low-resolutions and the KS statistic was below 0.03 for all upscaled resolutions compared to up to 0.22 for low-resolutions (Table 3.3.1).

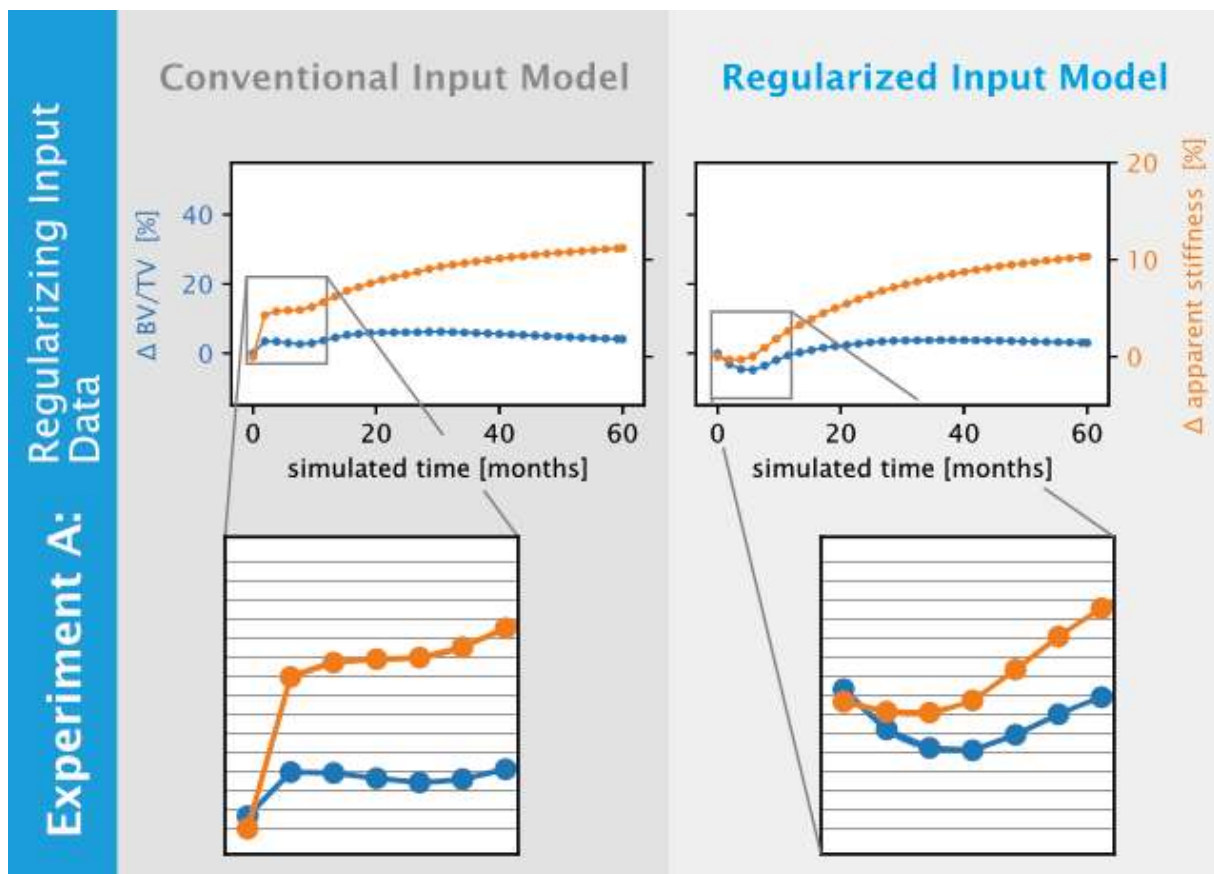


Fig. 3.3.4 Initialisation shocks in bone volume fraction (BV/TV) and relative apparent stiffness are reduced for regularized input models in comparison to conventional input models, shown for a representative sample of experiment A (Effect of regularized input models on initialisation shock). The initialization shock visible in BV/TV and stiffness (left) for the conventional model generation was not present when using the regularization approach (right).

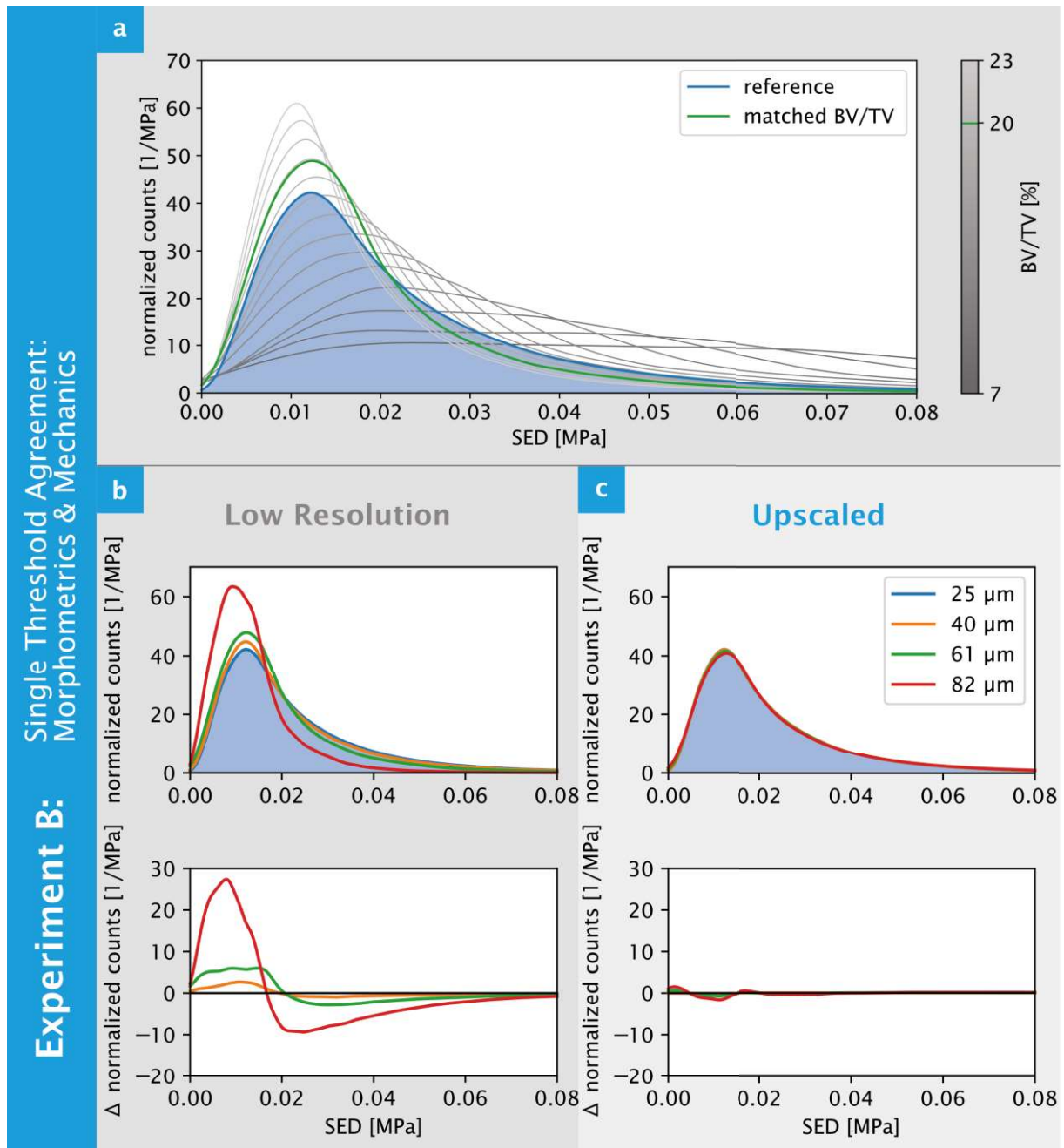


Fig. 3.3.5 Visual agreement in strain energy density (SED) distributions are improved for upscaled bone volume fraction (BV/TV) matched regularized input models for a representative sample of experiment B (Effect of upscaling on mechanical and morphometric accuracy). (a) The strain energy density (SED) distributions for the reference resolution (25 μm) (blue) was flattened 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. (b) Comparing across all resolutions, the low-resolution models did not capture the SED distribution of the reference model. (c) However, upscaled 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 upscaling.

Resolution, applied upscaling, and the interaction between resolution and upscaling had a significant effect on all measured static morphometric parameters, except for the interaction of the effects for SMI and the effect of upscaling on Tb.Th (Table 3.3.1). Without upscaling, significant differences in the mechanical and static parameters (except SMI) were observed for all lower resolutions (40, 61, and 82 μm) (Table 3.3.1). No differences in BV/TV were observed, as BV/TV was matched. With upscaling, deviations in the static parameters were significantly lower for each of the lower resolutions (u40, u61, and u82 μm , respectively) (Table 3.3.1).

The average regularization threshold for the upscaled images were 561.7 ± 6.8 , 530.9 ± 9.6 , and 496.6 ± 13.7 mg HA / cm^3 for the three upscaled resolutions (u40, u61, and u82 μm), respectively.

3.3.4.3 Experiment C: Effect of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation

For the reference simulations, BV/TV was initially reduced in the range of 1.8% to 14.8%, followed by an increase in BV/TV in the range of 3.8% to 38.8% (Figure S3.3.1). Tb.N decreased in the range of 5.2% to 22.8%. Trabecular thickness increased in the range of 29.1% to 42.4%, except for one sample for which it decreased by 8.8%. SMI was dependent on the specific sample, with differences in the range of -15.0% to 18.5%; two samples experienced an increase and three samples a decrease in SMI.

Resolution, upscaling, 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.

3.3.4.3.1 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 3.3.6). Deviations in static parameters significantly increased over the course of the simulation compared to the initial models (Table 3.3.1, Figure 3.3.7). Compared to the reference simulations, BV/TV and Tb.N were underestimated for all low-resolution simulations ($p < 0.01$), SMI and Tb.Sp were overestimated ($p < 0.05$), and Tb.Th did not follow a clear trend (Figure 3.3.7).

Table 3.3.1 Comparison of microstructural bone adaptation simulation outcomes for regularized model inputs of micro-CT images that were downscaled from low-resolution and then upscaled back to high-resolution. Parameters were compared against the high-resolution micro-CT image reference simulation and relative deviations in percent are shown.

		Downscaled from 25 μm			Upscaled to 25 μm			
		40 μm	61 μm	82 μm	u40 μm	u61 μm	u82 μm	
Regularized Input Model Deviation	Mechanics	Mean SED	4.1 \pm 1.3% *	17.2 \pm 2.0% ***, ###	42.2 \pm 7.1% **, ††	0.4 \pm 0.6% ‡	2.4 \pm 1.6% ‡‡	5.1 \pm 3.4% ‡‡
		KS	0.022 \pm 0.007 *	0.083 \pm 0.023 *, #	0.257 \pm 0.083 *, †	0.003 \pm 0.004 ‡	0.012 \pm 0.009 ‡	0.026 \pm 0.021 ‡
	Static Parameters	Tb.N	-7.9 \pm 2.7% **	-19.8 \pm 4.5% **, ##	-30.1 \pm 5.5% **, †††	-0.5 \pm 0.1% **, ‡‡	-6.6 \pm 1.3% **, ‡‡, ##	-13.4 \pm 2.7% **, ‡‡, ††
		Tb.Sp	9.8 \pm 2.6% **	26.8 \pm 6.0% **, ##	46.5 \pm 9.8% **, ††	0.5 \pm 0.1% **, ‡‡	7.1 \pm 1.4% **, ‡‡, ##	15.9 \pm 3.4% **, ‡‡, ††
		Tb.Th	-3.0 \pm 1.7% **	6.5 \pm 3.8% **, ##	16.8 \pm 5.8% **, ††	0.5 \pm 0.1% **	5.8 \pm 1.2% ##	12.4 \pm 2.7% *, ††
SMI	5.5 \pm 4.5%	13.9 \pm 11.4%	20.5 \pm 17.0%	0.03 \pm 0.2%	1.8 \pm 2.1%	6.6 \pm 5.0%		
Maximum Deviation Over the Course of the Simulation	Static Parameters	BV/TV	-6.7 \pm 2.4% **, ¥¥	-21.0 \pm 5.2% **, ¥¥, ##	-45.2 \pm 6.2% ***, ¥¥¥, ††	-0.3 \pm 0.4% ‡‡	-1.4 \pm 1.7% ‡‡	-3.2 \pm 3.7% ‡‡‡
		Tb.N	-11.1 \pm 3.8% **, ¥	-33.8 \pm 8.5% **, ¥¥, ##	-57.2 \pm 7.4% ***, ¥¥¥, ††	-0.6 \pm 0.2% **, ‡‡	-6.8 \pm 1.8% **, ‡‡, ##	-14.9 \pm 4.2% **, ‡‡‡, ††
		Tb.Sp	14.1 \pm 4.2% *, ¥	57.9 \pm 16.4% *, ¥, #	154.4 \pm 33.3% **, ¥¥, ††	0.6 \pm 0.3% *, ‡	7.5 \pm 2.1% *, ‡, #	18.2 \pm 5.8% *, ‡‡, †
		Tb.Th	-6.3 \pm 0.8% ***, ¥	10.3 \pm 4.0% *, ¥, ##	18.6 \pm 6.6% *, †	0.5 \pm 0.1% **, ‡‡‡	5.9 \pm 1.2% **, ##	12.9 \pm 3.0% **, ††
	SMI	12.6 \pm 3.8% *, ¥	31.4 \pm 8.9% *, ¥, #	48.4 \pm 13.5% *, ¥¥, #, †	0.2 \pm 0.2% ‡	2.8 \pm 0.9% *, ‡, #	8.2 \pm 2.3% *, ‡, †	
Dynamic Parameters	Bone Formation	70.9 \pm 1.9% ***	80.3 \pm 2.1% ***, ###	86.2 \pm 2.7% ***, †	1.5 \pm 0.7% ***, ‡‡‡	14.7 \pm 2.1% ***, ‡‡‡, ###	29.0 \pm 3.6% ***, ‡‡‡, †††	
	Bone Resorption	73.0 \pm 5.4% ***	186.7 \pm 23.4% ***, ##	365.6 \pm 61.3% **, ††	0.8 \pm 0.9% ‡‡‡	4.7 \pm 2.3% *, ‡‡‡, #	11.5 \pm 7.2% ‡‡	

SED: strain energy density; KS: Kolmogorov Smirnov statistic; Tb.N: Trabecular number, Tb.Sp: Trabecular spacing; Tb.Th: Trabecular thickness; SMI: Structure model index; s40, s61, and s82 μm : 40, 61, and 82 μm upscaled to 25 μm , respectively

Statistical significance is indicated as follows: Difference from zero (the reference simulation) (*), differences between the same resolution with and without upscaling (‡), difference between 40 and 61 μm (#), difference between 61 and 82 μm (†), and difference between regularized input model and maximum deviations (¥)

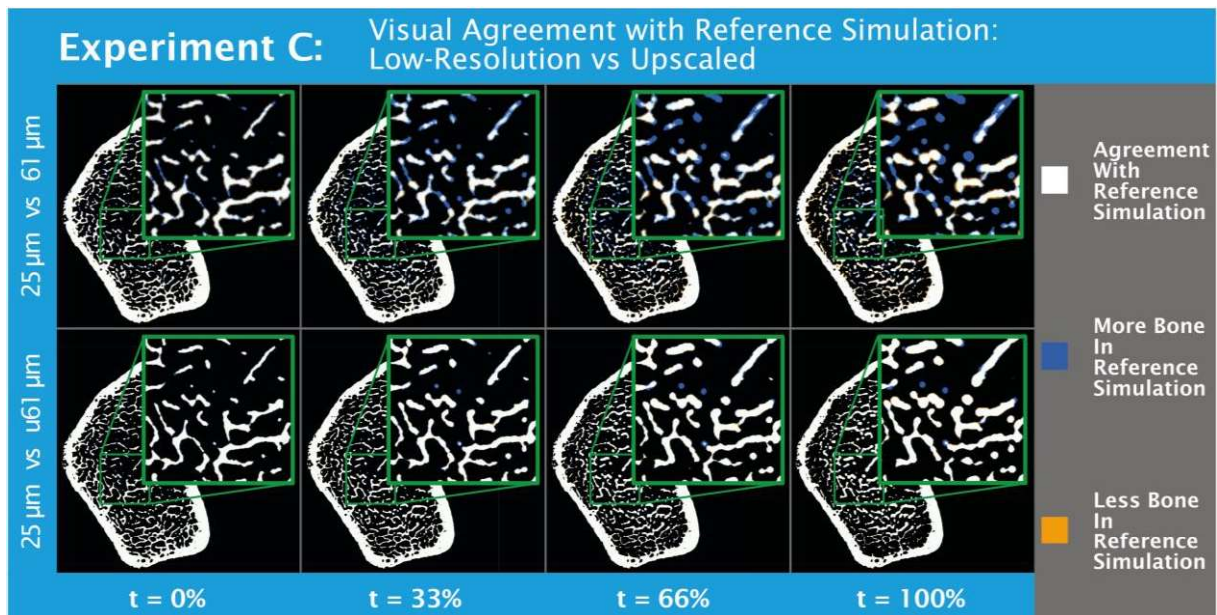


Fig. 3.3.6 Visual difference after 0, 33, 66, and 100% of the simulated time are reduced for upscaled regularized input models for a representative sample of experiment C (Effect of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation). Top: comparison of reference to low-resolution simulation. Bottom: comparison of reference to upscaled simulation. For the upscaled 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 upscaled images.

Bone formation and resorption rates were significantly larger for low-resolution simulations compared to the reference simulations (Table 3.3.1). 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 3.3.8). For the resorption rate, peak delay and widening was also observed for the low-resolution simulations (Figure 3.3.8). 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.

3.3.4.3.2 Upscaled simulations

Visually, differences in bone structure over the course of the simulation compared to the reference simulations were drastically reduced for the upscaled simulations (Figure 3.3.6). 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

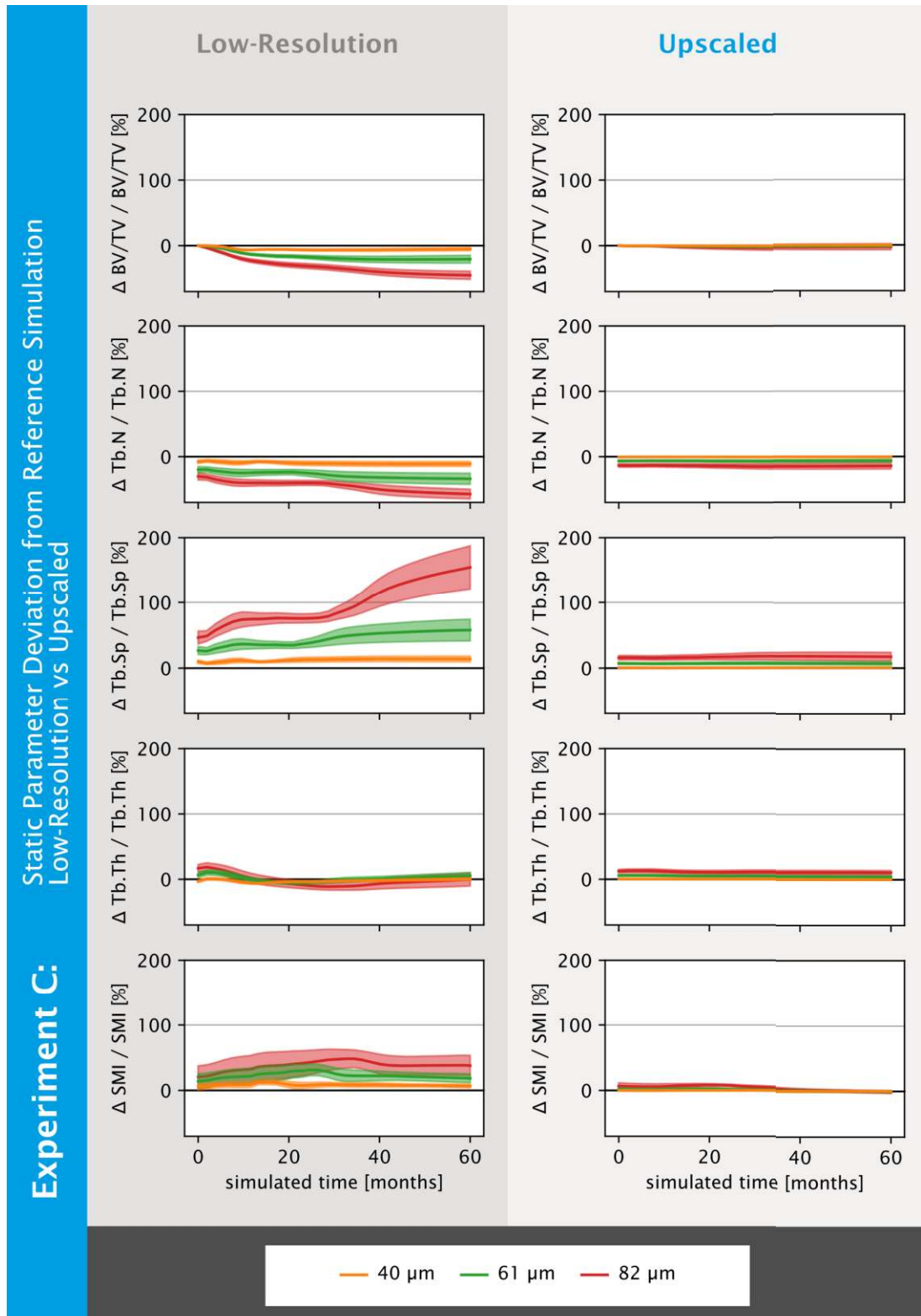


Fig. 3.3.7 Deviations from the reference simulation in static parameters are reduced for upscaled regularized input models across all resolutions as found in experiment C (Effect of

upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation). Deviations in static parameters from the reference simulation (25 μm) over the course of the simulation for low-resolution (left) and upscaled 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, upscaling improves the agreement with the reference (25 μm) simulations. Deviations were more predictable after upscaling of the image.

parameters was also significantly improved (Table 3.3.1, Figure 3.3.7). The maximum deviations in static parameters were not significantly greater than those of the input models for all upscaled resolutions.

For the dynamic parameters, accuracy of bone formation and resorption per time unit was significantly increased for the upscaled images (Table 3.3.1). 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 3.3.8).

3.3.4.4 Experiment D: Effect of *in vivo* image artefacts on convergence of upscaled HR-pQCT simulations

After an initial drop in BV/TV in the range of 7.2% to 24.2%, the simulations on the *in vivo* HR-pQCT data upscaled to 25 μm 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 in the range of 12.0% to 42.9%, whereas Tb.Th. increased over time in the range of 13.8% to 65.1%. Tb.Sp also increased in the range of 13.6% to 78.0%. SMI decreased for all samples in the range of 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 3.3.9). 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 3.3.9).

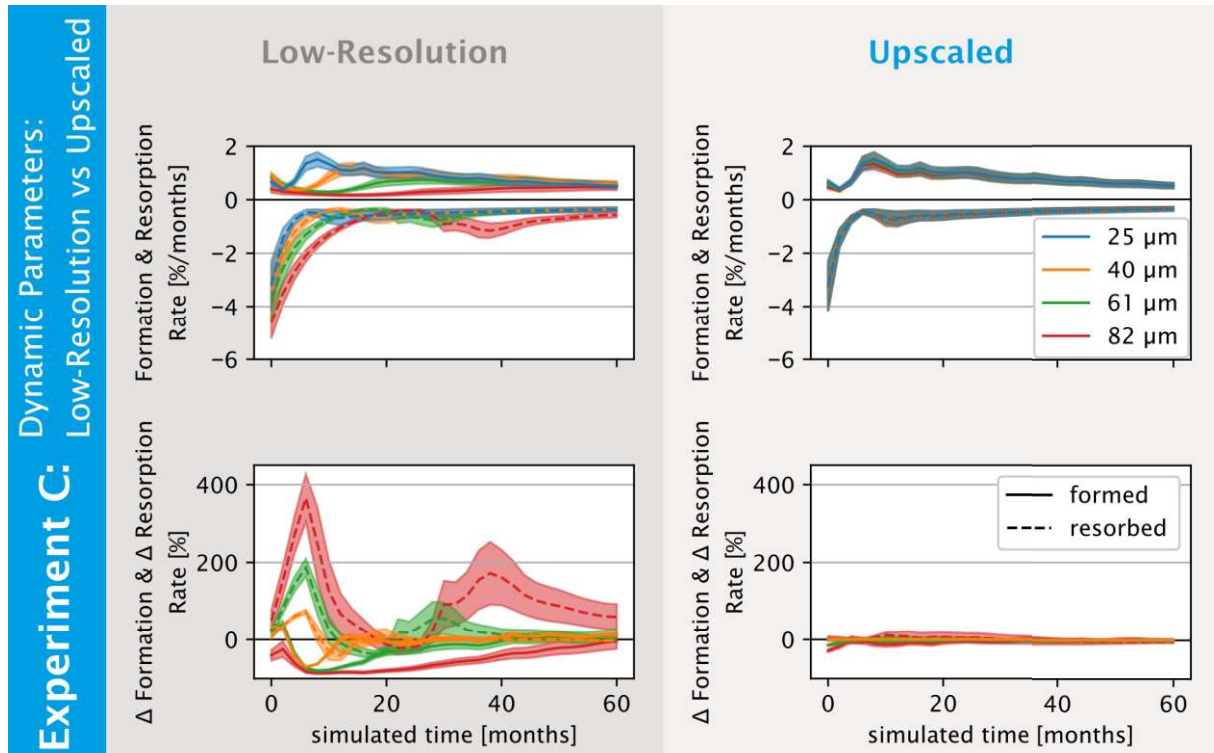


Fig. 3.3.8 Deviations from the reference simulation in dynamic parameters are reduced for upscaled regularized input models across all resolutions as found in experiment C (Effect of Upscaling on Morphometric Accuracy Throughout a Microstructural Bone Adaptation Simulation). Dynamic parameter results of simulation run on low-resolution (left) and upscaled 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 upscaled 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 upscaled regularized input models

3.3.5 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.

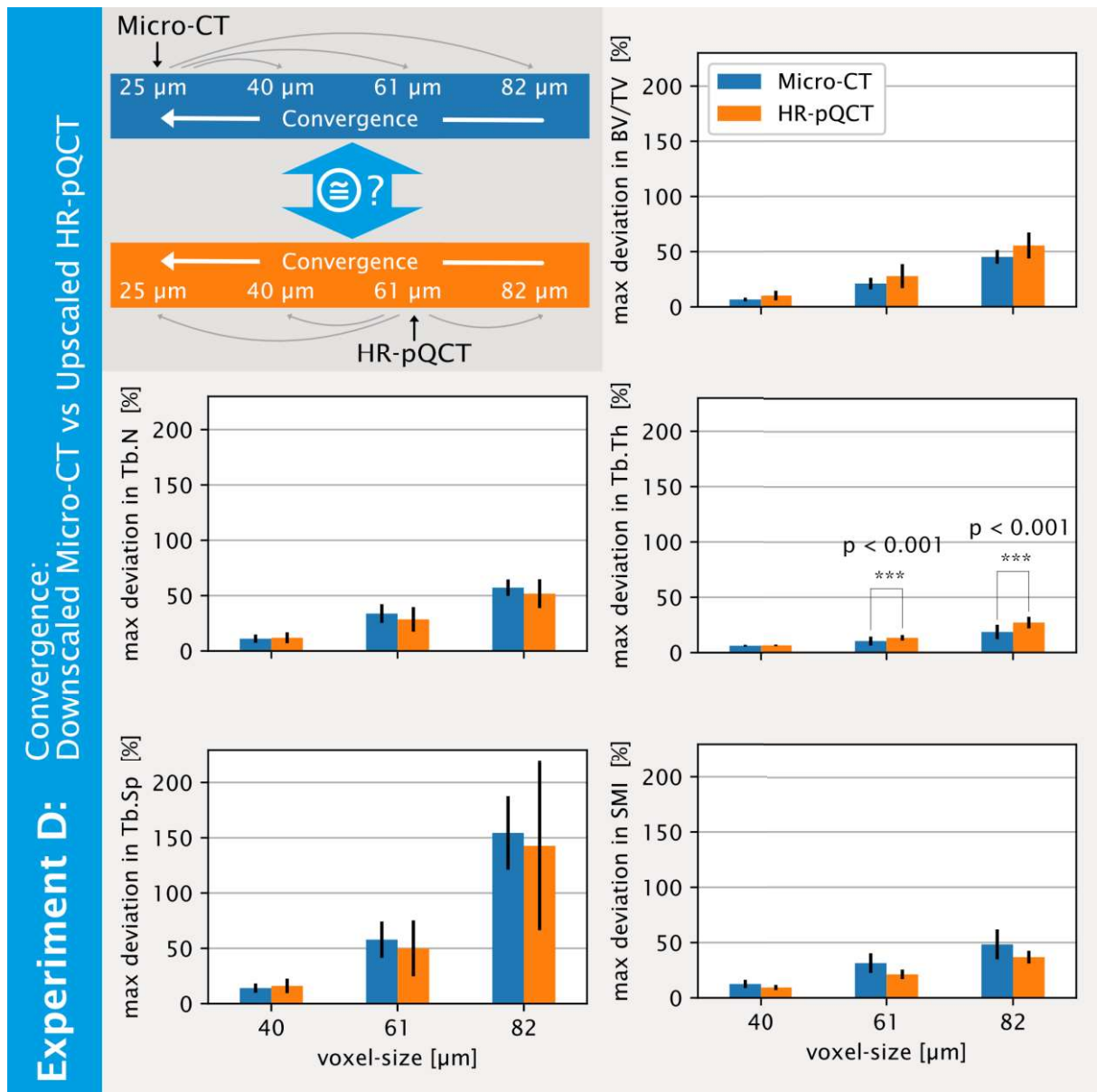


Fig. 3.3.9 Agreement in convergence between regularized input models from downscaled micro-CT and rescaled HR-pQCT as found in experiment D (Effect of *in vivo* image artefacts on convergence of upscaled HR-pQCT simulation). Convergence behaviour of low-resolution simulations from experiment C (Effect of Upscaling on Morphometric Accuracy Throughout a Microstructural Bone Adaptation Simulation) and from rescaled HR-pQCT image simulations (top left). 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.

3.3.5.1 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 (Viceconti and Hunter 2016), 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 3.3.4), 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, in the context of ocean climate models, Mulholland et al. (Mulholland et al. 2015) identified that 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. (2013) which tries to estimate the *in vivo* applied loads more closely than the uniaxial compression boundary conditions typically used with HR-pQCT radius data (Burghardt et al. 2011).

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 3.3.4). 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 3.3.4). Importantly, larger changes in BV/TV are expected for this type of simulation, as the structure adapts to the applied boundary conditions over the course of the simulation, yielding less changes as the structure reaches a shape optimized to the applied boundary conditions. We conclude that the first iteration step does not have to be excluded if this new approach is used, which allows direct comparisons of morphometrics and mechanics to *in vivo* measurements. Furthermore, the fact that the initialisation shock was removed might indicate that the regularized input model is a

more accurate mechanical representation of the *in vivo* bone structure than the conventional binary version.

3.3.5.2 Experiment B: Effect of upscaling on mechanical and morphometric accuracy

Another obstacle to overcome when running microstructural bone adaptation simulations on *in vivo* HR-pQCT images was finding an accurate digital representation of a bone captured *in vivo* with HR-pQCT with respect to mechanics and morphometry (Alsayednoor et al. 2018), which is an obvious requirement of bone adaptation simulations. We found that with the use of upscaling, the choice of a single threshold provided regularized input models that agreed well for BV/TV, mechanical properties, and other tested morphometric parameters (i.e. Tb.N, Tb.Sp, and SMI). Furthermore, the linear trend of the thresholds for the different upscaled resolutions indicated that even in the absence of a high resolution ground truth, an appropriate threshold for accurate morphometrics and mechanics can be chosen. In contrast, our results for images that were not upscaled agreed with previous research (Alsayednoor et al. 2018), which showed no agreement between mechanics and morphometrics for various thresholds (Figure 3.3.5). This lack of agreement also holds true for thresholds optimized to match BV/TV (Figure 3.3.5), a method used in a previous study by Christen et al. to investigate the voxel size dependence of a micro-FE based load estimation algorithm (Christen et al. 2016). The culmination of these results indicates that upscaling may also be useful for other applications using images of HR-pQCT resolution.

3.3.5.3 Experiment C: Effect of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation

To study the accuracy of microstructural bone adaptation simulations for images with HR-pQCT resolutions, we used high-resolution micro-CT images as ground truth, as this method has been previously validated to simulate realistic bone structures over time (Badilatti et al. 2016). The use of different low-resolution voxel-sizes (40, 61, and 82 μm) resulted in deviations in morphometric parameters of more than 30% in comparison to the reference simulations, however deviations as small as 15% may indicate disease, such as osteoporosis (Zhang et al. 2010). Importantly, observed differences in parameters were not consistent, thus could not be corrected or recovered via calibration curves, as has been possible in previous studies (Müller

et al. 1996). In comparison to static morphologic parameters, dynamic parameters were more dependent on image resolution. This can be understood by realizing that static parameters are averaged over the bone structure, such that local differences in morphometrics, e.g. trabecular number, may be concealed. On the other hand, dynamic parameters are additive and sum all changes over the given volume. Thus, even minor differences in structure result in larger deviations in dynamic parameters. Importantly, standard deviations of dynamic parameters from a previous study (Schulte et al. 2013) were lower by up to an order of magnitude compared to the deviations we observed for the highest clinically available resolution (80% vs 38% for formation, 187% vs 18% for resorption) (Table 3.3.1). Therefore, the accuracy we observed would be insufficient given the magnitude of natural variation previously observed for these dynamic parameters. Overall, our results show that microstructural bone adaptation simulations run on native clinical scanner resolutions suffer from poor accuracy in the assessment of static morphometric and dynamic parameters; thus, limiting future use as a model for human bone adaptation.

Running the same simulations on the upscaled images (u40, u61, and u82 μm) resulted in a drastic reduction in static parameter deviations to less than 10%. For BV/TV, these deviations were near 1% (Table 3.3.1), which is similar to the reproducibility limit of BV/TV for the clinical setting (0.84-1.14%) (MacNeil and Boyd 2008; Mueller et al. 2009). The improved accuracy in assessment of static parameters from the upscaled images is likely due to the need for higher resolution to correctly represent the reference model in terms of mechanics and morphology (Figure 3.3.6, Table 3.3.1). Any deviations of the regularized input model led to error accumulation throughout the simulation, such as missing thin individual trabecular structures that grow in thickness in the reference simulations (Figure 3.3.6). These errors ultimately led to larger morphometric deviations of the final structure (Figure 3.3.7). Only the dynamic parameters obtained from the upscaled simulations correctly captured the overall curve profile of the high-resolution simulations, matching the position, width, and height of peak formation and resorption rates. The large deviations of the downsampled resolutions indicate that there may have also been an intrinsic voxel-size dependence of the algorithm, independent of the initial model. The observed maximum deviations in dynamic parameters are also at least a factor of two smaller than the natural variation observed in inbred mice indicating sufficient accuracy of the method for human *in vivo* applications (Schulte et al. 2013).

3.5.5.4 Experiment D: Effect of *in vivo* image artefacts on convergence of upscaled HR-pQCT simulations

Finally, we investigated the effects of clinically observed image artefacts, e.g. higher noise levels, by using *in vivo* HR-pQCT images. Comparing the simulations run on these images to those run on upscaled versions of the same images, no significant difference was observed in the convergence of the different static parameters (Figure 3.3.9), except for Tb.Th. Since the *ex vivo* and *in vivo* dataset are not from the same study participants, it is possible that this difference in Tb.Th is due to unknown physiological differences between the *in vivo* subject group and the *ex vivo* samples. Furthermore, with respect to the overall deviation observed in Tb.Th, the observed significant difference is still small, especially given the fact that Tb.Th is known to be difficult to capture with HR-pQCT resolution (MacNeil and Boyd 2007; Manske et al. 2015) and are therefore likely not clinically relevant.

3.3.5.5 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 used in Experiment D. However, the artefacts most commonly associated with *in vivo* HR-pQCT images, such as motion artefact, are difficult to recreate with cadaveric specimen, while micro-CT images cannot be obtained from patients due to the radiation dosage and imaging volume. 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 upscaled 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.

3.3.6 Conclusion

In conclusion, we found model resolution to be the dominating image property which drove convergence errors in microstructural bone adaptation simulations. Importantly, upscaling

drastically reduced errors in mechanical analyses, static morphological parameters, and dynamic parameters, 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 upscaling 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. Hence, 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.

Conflict of Interest

The authors declare that they have no competing interests.

References

- 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. <https://doi.org/10.1115/1.1392315>
- Alsayednoor J, Metcalf L, Rochester J, et al (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. <https://doi.org/10.1007/s10237-018-1051-6>
- Badilatti SD, Christen P, Parkinson I, Müller R (2016) Load-adaptive bone remodeling simulations reveal osteoporotic microstructural and mechanical changes in whole human vertebrae. *J Biomech* 49:3770–3779. <https://doi.org/10.1016/j.jbiomech.2016.10.002>
- Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. *Geophys Res Lett* 36:1–5. <https://doi.org/10.1029/2008GL035561>
- Boutroy S, Bouxsein ML, Munoz F, Delmas PD (2005) In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. *J Clin Endocrinol Metab* 90:6508–6515. <https://doi.org/10.1210/jc.2005-1258>
- Boutroy S, Van Rietbergen B, Sornay-Rendu E, et al (2008) Finite element analysis based on in vivo HR-pQCT images of the distal radius is associated with wrist fracture in postmenopausal women. *J Bone Miner Res* 23:392–399. <https://doi.org/10.1359/jbmr.071108>
- Burghardt AJ, Link TM, Majumdar S (2011) High-resolution computed tomography for clinical imaging of bone microarchitecture. *Clin Orthop Relat Res* 469:2179–2193. <https://doi.org/10.1007/s11999-010-1766-x>
- Christen P, Ito K, Knippels I, et al (2013) Subject-specific bone loading estimation in the human distal radius. *J Biomech* 46:759–766. <https://doi.org/10.1016/j.jbiomech.2012.11.016>

- Christen P, Schulte FA, Zwahlen A, et al (2016) Voxel size dependency, reproducibility and sensitivity of an in vivo bone loading estimation algorithm. *J R Soc Interface* 13:20150991. <https://doi.org/10.1098/rsif.2015.0991>
- Cooper D, Turinsky A, Sensen C, Hallgrímsson B (2007) Effect of voxel size on 3D micro-CT analysis of cortical bone porosity. *Calcif Tissue Int* 80:211–219. <https://doi.org/10.1007/s00223-005-0274-6>
- De Jong JJA, Arts JJ, Meyer U, et al (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. <https://doi.org/10.2106/JBJS.O.00127>
- Fenner JW, Brook B, Clapworthy G, 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. <https://doi.org/10.1098/rsta.2008.0089>
- Flaig C (2012) A highly scalable memory efficient multigrid solver for μ -finite element analyses. ETH Zurich
- Hildebrand T, Laib A, Müller R, et al (1999) Direct Three-Dimensional Morphometric Analysis of Human Cancellous Bone: Microstructural Data from Spine, Femur, Iliac Crest, and Calcaneus. *J Bone Miner Res* 14:1167–1174. <https://doi.org/10.1359/jbmr.1999.14.7.1167>
- Jakob W, Rhineland J, Moldovan D (2017) pybind11 - Seamless operability between C++11 and Python
- Kirmani S, Christen D, Van Lenthe GH, et al (2009) Bone structure at the distal radius during adolescent growth. *J Bone Miner Res* 24:1033–1042. <https://doi.org/10.1359/jbmr.081255>
- Levchuk A, Zwahlen A, Weigt C, et al (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. <https://doi.org/10.1016/j.clinbiomech.2013.12.019>

- MacNeil JA, Boyd SK (2008) Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 30:792–799. <https://doi.org/10.1016/j.medengphy.2007.11.003>
- MacNeil JA, Boyd SK (2007) Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med Eng Phys* 29:1096–1105. <https://doi.org/10.1016/j.medengphy.2006.11.002>
- Manske SL, Zhu Y, Sandino C, Boyd SK (2015) Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone* 79:213–221. <https://doi.org/10.1016/j.bone.2015.06.006>
- Mueller TL, Stauber M, Kohler T, et al (2009) Non-invasive bone competence analysis by high-resolution pQCT: An in vitro reproducibility study on structural and mechanical properties at the human radius. *Bone* 44:364–371. <https://doi.org/10.1016/j.bone.2008.10.045>
- 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. <https://doi.org/10.1175/MWR-D-15-0076.1>
- 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. <https://doi.org/10.1007/s00198-004-1701-7>
- Müller R, Koller B, Hildebrand T, et al (1996) Resolution dependency of microstructural properties of cancellous bone based on three-dimensional mu-tomography. *Technol Health Care* 4:113–119
- Nicks KM, Amin S, Atkinson EJ, et al (2012) Relationship of age to bone microstructure independent of areal bone mineral density. *J Bone Miner Res* 27:637–644. <https://doi.org/10.1002/jbmr.1468>
- Nishiyama KK, Macdonald HM, Hanley DA, Boyd SK (2013) Women with previous fragility fractures can be classified based on bone microarchitecture and finite element analysis measured with HR-pQCT. *Osteoporos Int* 24:1733–1740. <https://doi.org/10.1007/s00198-012-2160-1>

- Nishiyama KK, Shane E (2013) Clinical imaging of bone microarchitecture with HR-pQCT. *Curr. Osteoporos. Rep.* 11:147–155
- Pialat JB, Burghardt AJ, Sode M, et al (2012) Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. *Bone* 50:111–118. <https://doi.org/10.1016/j.bone.2011.10.003>
- Python Software Foundation (2020) Python Language Reference, Version 3.7. <https://www.python.org>
- Rajapakse CS, Magland J, Zhang H, et al (2009) Implications of noise and resolution on mechanical properties of trabecular bone estimated by image-based finite-element analysis. *J Orthop Res* 27:1263–1271. <https://doi.org/10.1002/jor.20877>
- 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. <https://doi.org/10.1016/j.jbiomech.2004.03.037>
- Schulte FA, Christen P, Badilatti SD, et al (2019) Virtual supersampling as post-processing step preserves the trabecular bone morphometry in human peripheral quantitative computed tomography scans. *PLoS One* 14:1–11. <https://doi.org/10.1371/journal.pone.0212280>
- Schulte FA, Zwahlen A, Lambers FM, et al (2013) Strain-adaptive in silico modeling of bone adaptation - A computer simulation validated by in vivo micro-computed tomography data. *Bone* 52:485–492. <https://doi.org/10.1016/j.bone.2012.09.008>
- Seabold S, Perktold J (2010) statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference
- Sornay-Rendu E, Boutroy S, Duboeuf F, Chapurlat RD (2017) Bone Microarchitecture Assessed by HR-pQCT as Predictor of Fracture Risk in Postmenopausal Women: The OFELY Study. *J Bone Miner Res* 32:1243–1251. <https://doi.org/10.1002/jbmr.3105>
- Sornay-Rendu E, Boutroy S, Munoz F, Delmas PD (2007) Alterations of cortical and trabecular architecture are associated with fractures in postmenopausal women, partially independent of decreased BMD measured by DXA: The OFELY study. *J Bone Miner Res* 22:425–433. <https://doi.org/10.1359/jbmr.061206>

- Tjong W, Kazakia GJ, Burghardt AJ, Majumdar S (2012) The effect of voxel size on high-resolution peripheral computed tomography measurements of trabecular and cortical bone microstructure. *Med Phys* 39:1893–1903. <https://doi.org/10.1118/1.3689813>
- Van Der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng* 13:22–30. <https://doi.org/10.1109/MCSE.2011.37>
- Van Der Walt S, Schönberger JL, Nunez-Iglesias J, et al (2014) Scikit-image: Image processing in python. *PeerJ* 2014:1–18. <https://doi.org/10.7717/peerj.453>
- Viceconti M, Hunter P (2016) The Virtual Physiological Human: Ten Years After. *Annu Rev Biomed Eng* 18:103–123. <https://doi.org/10.1146/annurev-bioeng-110915-114742>
- Virtanen P, Gommers R, Oliphant TE, et al (2019) SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python. 1–22
- Yu WS, Chan KY, Yu FWP, et al (2014) Bone structural and mechanical indices in Adolescent Idiopathic Scoliosis evaluated by high-resolution peripheral quantitative computed tomography (HR-pQCT). *Bone* 61:109–115. <https://doi.org/10.1016/j.bone.2013.12.033>
- Zhang ZM, Li ZC, Jiang LS, et al (2010) Micro-CT and mechanical evaluation of subchondral trabecular bone structure between postmenopausal women with osteoarthritis and osteoporosis. *Osteoporos Int* 21:1383–1390. <https://doi.org/10.1007/s00198-009-1071-2>
- Zhu TY, Griffith JF, Qin L, et al (2014) Alterations of bone density, microstructure, and strength of the distal radius in male patients with rheumatoid arthritis: A case-control study with HR-pQCT. *J Bone Miner Res* 29:2118–2129. <https://doi.org/10.1002/jbmr.2221>

Supplementary material

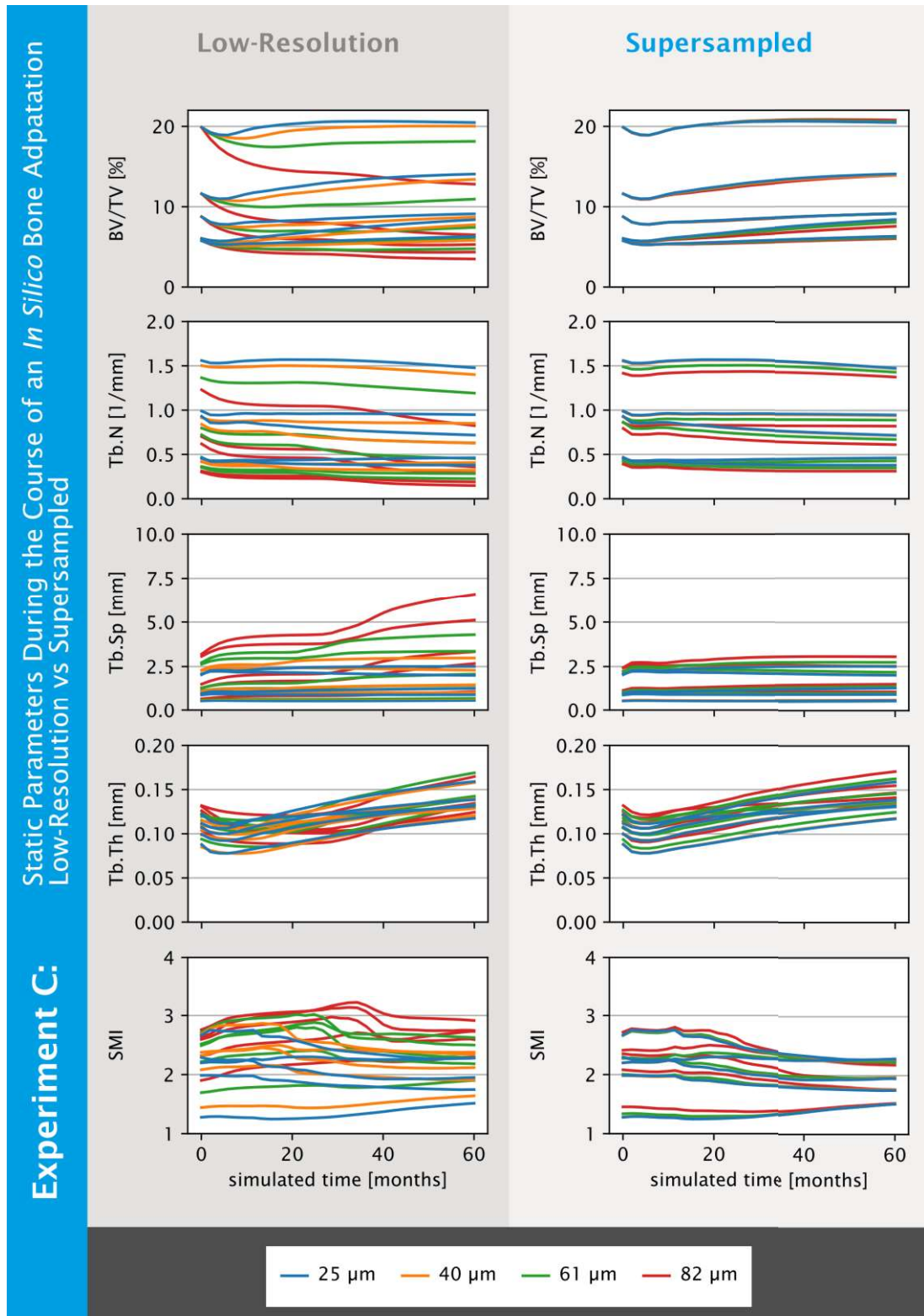


Fig. S3.3.1 Visual deviations from the reference simulation in static parameters are reduced for upscaled regularized input models across all resolutions as found in experiment C (Effect

of upscaling on morphometric accuracy throughout a microstructural bone adaptation simulation). Static parameters over the course of the simulation for low-resolution (left) and upscaled 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 upscaling improves the agreement with the reference (25 μm) simulations

3.4 Automated segmentation of fractured distal radii by 3D geodesic active contouring of *in vivo* HR-pQCT images

N. Ohs¹, D. C. Tourolle né Betts¹, P. R. Atkins¹, Caitlyn J. Collins¹, Bryant Schroeder¹, M. Blauth², P. Christen^{1,3}, R. Müller¹

¹Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

²Department for Trauma Surgery, Innsbruck University Hospital, Innsbruck, Austria

³Institute for Information Systems, FHNW University of Applied Sciences and Arts
Northwestern Switzerland, Olten, Switzerland

Abstract

Radius fractures are among the most common fracture types; however, there is limited consensus on the standard of care. A better understanding of the fracture healing process could help to shape future treatment protocols and thus improve functional outcomes of patients. High-resolution peripheral quantitative computed tomography (HR-pQCT) allows monitoring and evaluation of the radius on the micro-structural level, which is crucial to our understanding of fracture healing. However, current radius fracture studies using HR-pQCT are limited by the lack of automated contouring routines, hence only including small number of patients due to the prohibitively time-consuming task of manually contouring HR-pQCT images.

In the present study, a new method to automatically contour images of distal radius fractures based on 3D morphological geodesic active contours (3D-GAC) is presented. Contours of 60 HR-pQCT images of fractured and conservatively treated radii spanning the healing process up to one year post-fracture are compared to the current gold standard, hand-drawn 2D contours, to assess the accuracy of the algorithm. Furthermore, robustness was established by applying the algorithm to HR-pQCT images of intact radii of 73 patients and comparing the resulting morphometric indices to the gold standard patient evaluation. Reproducibility was evaluated using repeat scans of intact radii of 19 patients.

The new 3D-GAC approach offers accurate contours within inter-operator variability for images of fractured distal radii. The generated contours for images of intact radii yielded morphometric indices within *in vivo* reproducibility limits compared to the current gold standard. Additionally, the 3D-GAC approach shows an improved robustness against failure when dealing with cortical interruptions, fracture fragments, etc. Using the 3D-GAC approach assures consistent results, while reducing the need for time-consuming semi-automatic contouring.

Keywords:

HR-pQCT; fracture; active contours; automatic radius

3.4.1 Introduction

In our ageing population, bone fractures are increasingly common and have become a major socioeconomic burden (Johnell and Kanis, 2006). Specifically, fractures of the distal radius are among the most common fracture types and indicative of reduced bone quality (Court-Brown and Caesar, 2006). However, there is currently limited consensus on the optimal treatment protocol (Ng and McQueen, 2011). The conservative treatment of many radius fractures provides the opportunity to study the fracture healing process in humans. A better understanding of this process could help to shape future treatment protocols of distal radius fractures, ultimately resulting in better functional outcomes for patients.

With the introduction of high-resolution peripheral quantitative computed tomography (HR-pQCT), longitudinal changes in the microstructure of the radius can be monitored non-invasively (Burghardt et al., 2010b; Burt et al., 2017; Ellouz et al., 2014; Nishiyama et al., 2015; Shanbhogue et al., 2017). Use of this technology has revealed critical changes in bone microstructure as a result of aging as well as pharmaceutical and surgical interventions. For example, an increase in cortical porosity, which is highly associated with bone frailty, was revealed in a longitudinal study of patients who had undergone kidney transplantation (Nishiyama et al., 2015). Preliminary evidence suggests that HR-pQCT might also be a viable technique to monitor bone fractures *in vivo* (De Jong et al., 2014). The microstructure in such studies is of fundamental importance, as techniques such as micro-finite-element analysis can be utilized to estimate the capability of a bone scanned with HR-pQCT to withstand mechanical load, a key parameter to assess successful fracture healing.

Image-based clinical research, such as that using HR-pQCT, requires accurate, reproducible and scalable tools for image processing. For this reason, automatic contouring approaches have been developed to reduce the need for hand-drawn contours (Buie et al., 2007; Burghardt et al., 2010a, 2007; Zebaze et al., 2013) in HR-pQCT studies. The method developed by Buie et al. (Buie et al., 2007), specifically, is integrated into the software of the manufacturer of HR-pQCT devices (XtremeCT I and II, Scanco Medical AG, Switzerland), making it the de-facto standard. All of these automatic contouring approaches segment the target bone (i.e. radius) by first finding the outer contour of the bone with the assumption that the largest cortical interruptions imaged with HR-pQCT are small (on the sub-mm length scale) and that the cortex forms a well-defined, high-contrast edge in the HR-pQCT image. However, these assumptions break down

when processing images of fractured radii where the cortex is fragmented, and the presence of cartilaginous callus may blur the boundary between bone and soft tissue. Furthermore, radius fractures often occur in the ultra-distal region, contributing to the complexity of the image segmentation, as the cortex is both thinner and less mineralized. To our knowledge, approaches to generate automatic outer contours for HR-pQCT images of distal radius fractures have yet to be proposed.

Active contouring (Kass et al., 1988) is a promising technique that has already been shown to successfully segment HR-pQCT images of healthy distal radii (Hafri et al., 2016b, 2016a). The advantage of the active contouring algorithm is the inherent ability to pass over larger gaps in the bone. This is possible because the requirements for a contour to be both close to the object edges and smooth are balanced. Recently, three-dimensional morphological geodesic active contours (3D-GAC) have been implemented, allowing for efficient calculation of active contours on 3D images (Caselles et al., 1997; Marquez-Neila et al., 2014). Hence, 3D-GAC appears to be a promising approach to create outer contours of HR-pQCT images of distal radius fractures.

The goal of present study was to investigate the use of fully automatic 3D-GAC to generate outer contours on HR-pQCT images of both intact and fractured radii. We validated the accuracy of these contours against hand-drawn contours using HR-pQCT images acquired throughout the first year of the healing process from 10 patients with fractured distal radii. We further assessed the robustness of the algorithm by comparing computed morphometric indices from intact distal radii HR-pQCT images of 73 subjects using contours generated from (i) the 3D-GAC and (ii) the scanner manufacturer default approach. Lastly, the reproducibility of the 3D-GAC based algorithm was assessed based on contours of 19 subjects imaged six times using HR-pQCT.

3.4.2 Materials and Methods

3.4.2.1 Data

HR-pQCT images (XtremeCT II, 60.7 μm voxels, 68 kV, 1470 μA , 43 ms integration time) were obtained from the database of a previous fracture study conducted at Innsbruck Medical University. Patients provided informed consent and participated in a study approved by the

ethics committee of the Medical University of Innsbruck (UN 0374344/4.31). For each patient, six scans of the fractured (up to 504 slices per scan distributed over three stacks) and contralateral (up to 168 slices per scan) radius were taken over the course of one year (1, 3, 5, 13, 26, and 52 weeks post-fracture).

3.4.2.1.1 Fractured radii

Images of radius fractures for 10 out of the 75 patients that completed the study were selected from the database based on the highest available visual grading scores (VGS) (Pialat et al., 2012) and well aligned stacks resulting in 60 HR-pQCT fractured radius images. For each image, the most distal slice before the appearance of the lunate fossa was identified, and all evaluations on these images were performed on slices proximal to that slice.

3.4.2.1.2 Intact radii

Seventy-three patients were selected which had at least one HR-pQCT image of the contralateral, intact radius with a VGS of three or better resulting in 438 images. Of these 73 patients, 19 had a VGS of three or better across all time points and were used in the reproducibility study.

3.4.2.2 Image pre-processing

HR-pQCT images of each patient were registered using rigid image registration (Thévenaz et al., 1998) to allow direct comparison of generated contours across all time-points per patient.

3.4.2.3 Contours: Current gold standards

For ten randomly selected slices per fractured radius HR-pQCT image, hand-drawn 2D contours were generated by three researchers experienced in the processing of medical images (DCT, PRA, CJC), respectively, using the software of the scanner manufacturer. For this procedure, the 2D slice is visualised in the scanner manufacturer's software, and using a computer mouse, the contour is drawn onto the image. The software provides zooming functionality and local edge detection to assist the operator in defining the outer contour. These hand-drawn 2D contours were used as the gold standard to validate the results of the 3D-GAC algorithm for fractured radii and to quantify the inherent inter-observer variability in hand-drawn contours of fractured radii.

For intact radii, the default manufacturer pipeline based on a 3D algorithm by Buie et al. (Buie et al., 2007), written in the Image Processing Language (IPL) of the scanner manufacturer (Scanco Medical AG, Switzerland), was used as the gold standard. All contours generated using the manufacturer's software were exported as binary 3D images to h5 files (The HDF Group, 2020), a hierarchical data format, for further processing using Python (Python Software Foundation, 2020).

3.4.2.4 Novel contouring algorithm

The proposed algorithm uses 3D morphological geodesic active contours (3D-GAC) (Caselles et al., 1997) implemented in the Python library Scikit-Image (Van Der Walt et al., 2014) using morphological operators (Marquez-Neila et al., 2014). The guiding principle for geodesic active contours is to minimize an energy term consisting of both internal and image energy. The internal energy penalizes non-smooth contours, while the image energy penalizes contours away from the voxels of interest. The image energy landscape, in this approach, is generated using an inverse Gaussian gradient, where the Gaussian blur removes local minima and the gradient is an operation to detect object edges.

The proposed contouring algorithm can be separated into four distinct steps. First, the image is segmented into sections that only contain one bone (here, radius or ulna). In a second step, an initial guess of the radius contour is generated. Third, the image is pre-processed and converted to an energy landscape. Finally, the 3D-GAC algorithm is applied in an iterative loop to isolate the surface of the radius.

3.4.2.4.1 Segment radius and ulna

3D-GAC are drawn to all intensity-based edges in an image and do not differentiate between objects of interest and other objects. Therefore, it is necessary to remove extraneous objects. Here, the ulna must be removed, as it is in close proximity to the radius. The watershed algorithm implemented in Scikit-Image (Van Der Walt et al., 2014) is used to separate these two bones, based on seed voxels of the different objects.

Seed voxels of radius and ulna

Since the radius and ulna are typically only a few voxels apart in ultra-distal scans, segmenting using a global threshold has the risk of identifying both bones as a single object. Therefore, the radius and ulna are identified in a proximal slice (here, 35-39 slices from proximal to avoid end

effects from registration) using a threshold of 250 mg HA / cm³ and component labelling implemented in the Python library NumPy (Van Der Walt et al., 2011). The largest component identified is the radius and the second largest one the ulna. Finally, a seeds image is generated by copying the labels of the radius and ulna in the proximal slice into the entire proximal half of a zeroed image. The distal part was left empty to avoid incorrect assignments of seeds in the distal region, where radius and ulna are not well separated.

Watershed algorithm

The energy landscape for the watershed algorithm is generated from the HR-pQCT image by setting all negative values to zero and inverting the sign of all remaining elements. The algorithm is applied to the entire image such that every voxel – background or bone – is assigned to exactly either the radius or ulna.

3.4.2.4.2 Initial guess

To generate the initial guess contour for the 3D-GAC algorithm, the image is first normalized to the range of zero to one. Afterwards, the image is equalized in the two orthogonal longitudinal planes independently using a contrast limited adaptive histogram equalization implemented in Scikit-Image (Van Der Walt et al., 2014) and the resulting two images are averaged. A threshold of 0.5 is applied and the ulna portion of the image identified during pre-processing is removed from the thresholded image. Finally, the greatest connected component is extracted via component labelling, the structure is closed using a closing distance of 50 voxels, the remaining interior holes are filled using an additional component-labelling step and the structure is dilated using five iterations to obtain the initial guess.

3.4.2.4.3 Pre-processing of image

To prepare the image from the scanner to be used as an input to the 3D-GAC algorithm, the image is first cropped to the bounds of the initial guess contour to reduce the image size, which reduces computational cost. The image is then normalized to the range of zero to one and equalized as is done during the creation of the initial guess. The catchment basins from the pre-processing watershed algorithm are used to set those voxels not belonging to the radius section to zero. Finally, the voxels outside the initial guess are set to the mean value of the initial guess surface voxels, which are those voxels that make up the perimeter of the initial guess.

3.4.2.4.4 3D geodesic active contours (3D-GAC)

The 3D-GAC algorithm takes an initial guess, an energy landscape and outputs an optimized contour. The energy landscape needs to be generated from the input image. Since the algorithm is easily trapped in local minima, an iterative application of 3D-GAC to energy landscapes containing increasing levels of details is part of the proposed approach.

Density to Energy Landscape Conversion

The pre-processed image is used to create energy landscapes. For this, the image is first padded in the longitudinal direction by 20 voxels at each end using the NumPy (Van Der Walt et al., 2011) edge mirroring setting. The padding is followed by the application of two iterations of a Gaussian filter. The sigma of the Gaussian filter is a parameter that is decreased step-wise during the iterative application of 3D-GAC. Finally, the padding voxels are removed.

Iterative application of 3D-GAC

To avoid finding an unwanted local minima (that is, a contour that does not match the surface of the radius), Gaussian blurring is used to smooth the energy landscape. However, finer details of the radius can get lost when a Gaussian blurring with a large sigma is used. Therefore, three iterations of 3D-GAC are run using decreasing sigmas (14.0, 3.0, 1.5 voxels) for the Gaussian blurring to allow for an iterative approximation of the true radius contour. Additionally, for the first iteration, image resolutions are halved after blurring to speed up computations. Each 3D-GAC application is then run using the default parameters of SciPy (Virtanen et al., 2019) with the number of iterations set to five. As a final step to the entire algorithm, a component labelling followed by a selection of everything not connected to the background is performed to remove holes in the contour that can appear at the distal image boundary.

3.4.2.5 Morphometric indices

The Scanco system provides the standard patient evaluation script for XtremeCT II devices, which is the gold standard approach of generating cortical and trabecular masks (Burghardt et al., 2010a) and computing bone morphometric indices. Total volumetric bone mineral density (Tt.vBMD), trabecular volumetric bone mineral density (Tb.vBMD), cortical volumetric bone mineral density (Ct.vBMD), trabecular area (Tb.Ar), cortical area (Ct.Ar), trabecular bone volume fraction (Tb.BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), cortical perimeter (Ct.Pm), and cortical thickness (Ct.Th) were computed for this study. The default setup uses outer contours generated by the approach of

Buie et al. as input (Buie et al., 2007). However, the standard patient evaluation pipeline also accepts alternative outer contours as input. Outer contours generated from the proposed novel 3D-GAC method were used as input to the standard patient evaluations and the resultant morphometrics were compared to those of the manufacturer's default pipeline.

3.4.2.6 Study design

To assess the quality of the generated contours, we compared them to the respective gold standard for two different datasets (fractured and intact distal radii).

3.4.2.6.1 Fractured radii: Comparison of the 3D-GAC approach to hand-drawn contours

The 3D-GAC method was run for all fracture images. Contours were visually categorised as acceptable, mistakenly including parts of the ulna, having obvious missing parts, or having both of these issues. Only images categorised as acceptable were used for following quantitative analysis steps. Agreement between hand-drawn contour operators was assessed by computing the Dice score between their respective 2D contours for each image. The Dice score is a typical measure to compare automatically generated contours against reference contours yielding values in the range from zero to one with zero indicating no overlap while one indicates a perfect match. The Dice score is close to percent agreement when comparing contours that are similar to each other. Additionally, the distance transform of the area between the edges of 2D contours between operators was computed. Accuracy of the 3D-GAC approach was determined by taking the median Dice score between the automatic and the three hand-drawn 2D contours for each of the slices per image for which hand-drawn 2D contours exists. Furthermore, all three hand-drawn 2D contours were combined to yield a smallest and largest contour. The voxels between these two contours describes an area of uncertainty for the hand-drawn contours. Voxels, which were in the smallest hand-drawn 2D contour but not in the automatic one and voxels, which were in the 3D-GAC contour but not in the largest hand-drawn 2D contour were extracted and a distance transform was applied as a measure for deviations of the 3D-GAC contours from the hand-drawn 2D ones. Bland-Altman plots between the 3D-GAC and the closest hand-drawn 2D contour in terms of area from all three operators for each slice were generated to identify systematic differences between hand-drawn contours and the 3D-GAC outer contours. The default manufacturer pipeline was also run on all fracture images and categorisation as for the 3D-GAC contours was performed.

3.4.2.6.2 Intact radii: Comparison of the 3D-GAC and manufacturer's approach

The 3D-GAC approach and the default manufacturer pipeline were run on all intact radii. Contours of both approaches were visually assessed and categorized as was done for the contours of the fractured radii. For each patient, one image for which both approaches were categorized as being acceptable was randomly chosen for further analysis. Accuracy of the 3D-GAC method was assessed by computing the Dice score between each contour of the 3D-GAC method and the respective contour of the manufacturer's method. Additionally, the distance transform of the volume between the surfaces of both approaches was computed as another measure of how much the two methods deviated from each other. Bland-Altman plots were generated to assess systematic differences between the two methods. Reproducibility was determined by computing the Dice score of each follow-up image with the baseline image for each patient. Morphometrics were computed to assess how strongly the computation of standard morphometric indices is influenced by the choice of outer contour.

3.4.2.7 Bland-Altman plots & statistics

Bland-Altman plots were created to assess systematic deviations for morphometric indices. To allow visual comparison of the magnitudes of deviations between different parameters, all Bland-Altman plots had their y-axis scaled by 13% of the maximum x-value throughout this paper, which was the maximum value not leading to clipping of data. Robust linear models were fitted for Bland-Altman plots with Huber's T for M estimation using the statsmodels package in Python (Seabold and Perktold, 2010). This package also provides two-sided Student's t-tests for the linear models with zero slope and intercept of each Bland-Altman plot being the null-hypotheses, respectively.

Statistical significance for the comparison of morphometrics was performed using a paired Student's t-test. Significance level was set to $P < 0.05$.

3.4.3 Results

3.4.3.1 Fractured radii: Comparison of the 3D-GAC approach to hand-drawn contours

From the 60 images, 20 were successfully contoured by both 3D-GAC and the default manufacturer pipeline. The latter approach failed 40 times, while the 3D-GAC approach failed for four images (Figure 3.4.1a). The default manufacturer pipeline failed due to missing parts of the radius as a result of fracture gaps in the cortex (N=33), including parts of the ulna (N=2) or both (N=5). The 3D-GAC approach failed due to missing parts of the radius (N=2), including parts of the ulna (N=1), or including parts of the background (N=1). Furthermore, the default manufacturer pipeline only showed a high success rate for images taken one year post fracture (Figure 3.4.1b). It should be noted, that the default manufacturer pipeline has not been designed to handle images of fractured radii. Visually, contours matched the given bone structures well within the inter-operator variability (Figure 3.4.2). Agreement was found between the 3D-GAC approach and hand-drawn contours with a mean Dice score of 0.992 ± 0.004 . This is within the range of inter-operator agreement: (operator 1 vs operator 2) 0.993 ± 0.005 , (operator 1 vs operator 3) 0.993 ± 0.006 , and (operator 2 vs operator 3) 0.994 ± 0.005 . Evaluating the six appointments individually, Dice scores were high (>0.985), though some increase in the inter-quartile range of the Dice scores was observed between three and thirteen weeks post fracture (Figure 3.4.1c). This coincided with a decrease in contouring success (Figure 3.4.1b). For voxels of the 3D-GAC deviating from the area of uncertainty, 82% of these contour voxels were only one voxel away from at least one hand-drawn contour. This was within the inter-observer variability, as we observed (operator 1 vs operator 2) 57%, (operator 1 vs operator 3) 60%, and (operator 2 vs operator 3) 58% of deviations being less than one voxel away between operators (Figure 3.4.1d). Comparing the contour areas between the 3D-GAC method and the three operators, agreement was high with errors being: $-1.14 \pm 0.76\%$, $-0.40 \pm 0.62\%$, and $-0.22 \pm 0.73\%$ for each operator, respectively. An underestimation of the total area was observed for the 3D-GAC contours (Figure 3.4.3). This underestimation is less than e.g. comparing operator 1 to operator 3: $-0.011[\pm 0.004] + 0.090[\pm 0.134]$, i.e.: slope (-0.003 vs 0.011) and intercept (-0.076 vs 0.090 (both not significant)) (Figure 3.4.3).

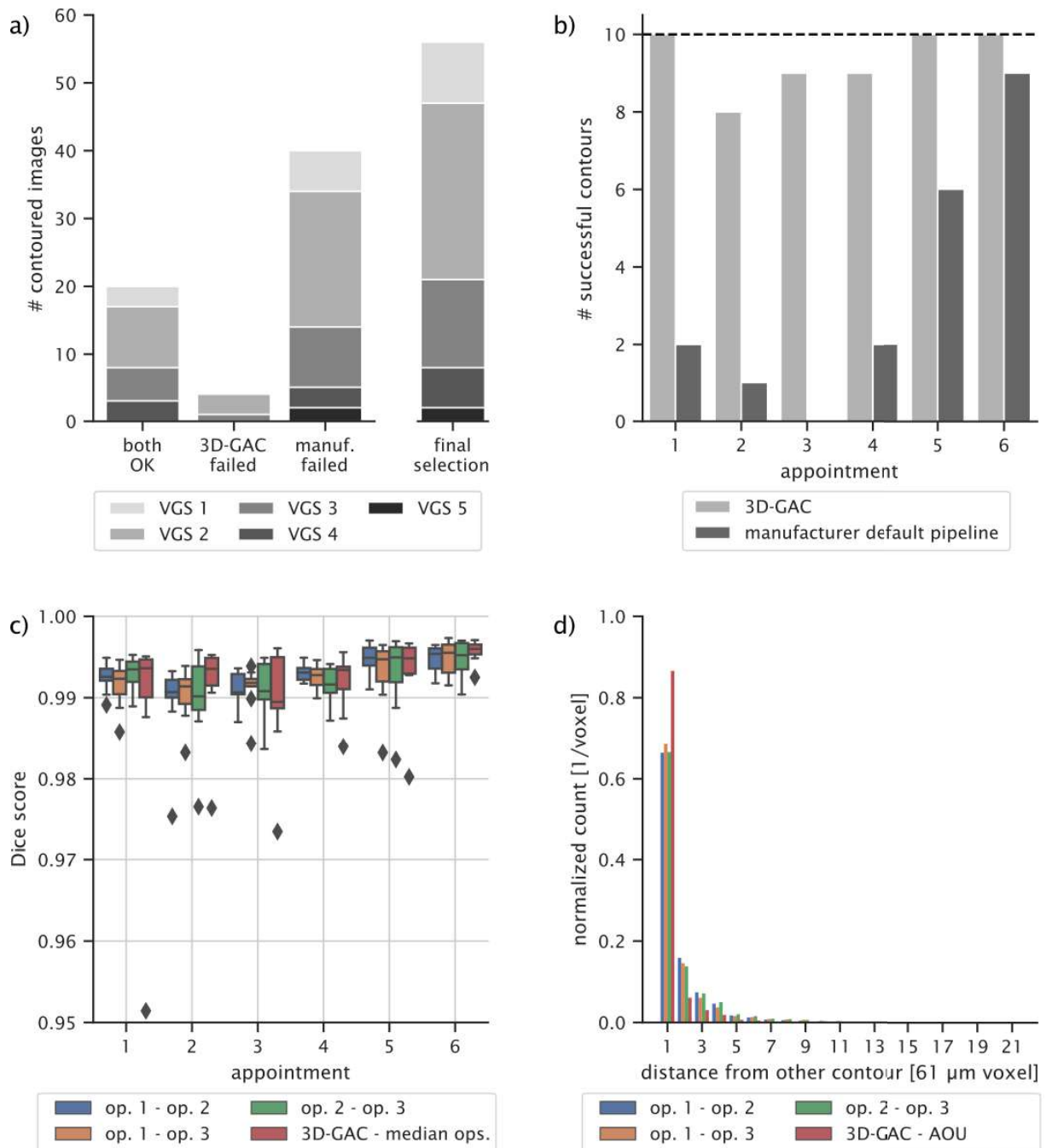


Fig. 3.4.1 Accuracy of the 3D-GAC approach vs hand-drawn contours for scans of fractured radii. a) Histogram of successful and failed contours for the 3D-GAC and the default manufacturer pipeline. The final bar shows the images selected for further analysis. Note that this pipeline was not designed to handle scans of fractured radii. b) Histogram of successful contours for the two approaches per patient time-appointment. Dashed line indicates 100% success. c) Dice scores computed for all images between the hand-drawn slices of all operators (op) respectively. The corresponding slices of the 3D-GAC contours were also compared against the hand-drawn slices, respectively, and the average median Dice score is shown. d) Histogram of the distance of all voxels different between the 3D-GAC and the area of uncertainty (AOU) for the hand-drawn contours of each operator, respectively. The AOU is defined as the area between the contour obtained by only accepting voxels present in all hand-

drawn contours and the contour obtained by accepting voxels present in at least one hand-drawn contour.

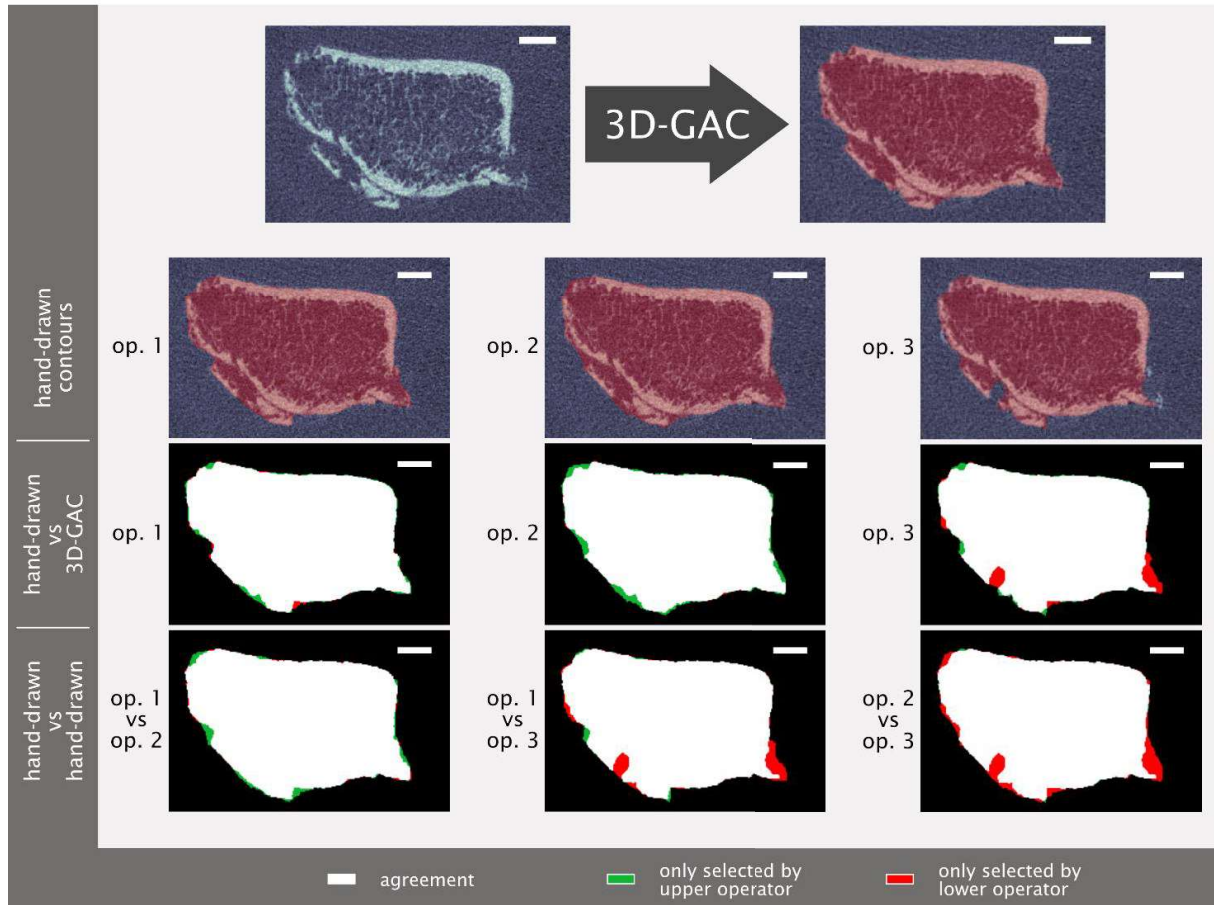


Fig. 3.4.2 Visual comparison of hand-drawn contours (by three different operators (op)) and 3D geodesic active contour (3D-GAC) generated contours for a representative challenging slice from one week post-fracture. Visual differences between the 3D-GAC and hand-drawn contours are similar to those between hand-drawn contours of different operators. Scale bar is 2 mm.

3.4.3.2 Intact Radii: Comparison of the 3D-GAC and Manufacturer's Approach

From the 438 images of intact radii, 341 were successfully contoured by both approaches. The default manufacturer pipeline failed for 85 images, the 3D-GAC approach failed 3 times, and both approaches failed 9 times (Figure 3.4.4a). The default manufacturer pipeline failed due to the inclusion of parts of the ulna (N=42) into the radius contour, omitting large parts of the radius (N=37) (Figure 3.4.5), having parts of the background in the contour (N=5), or including

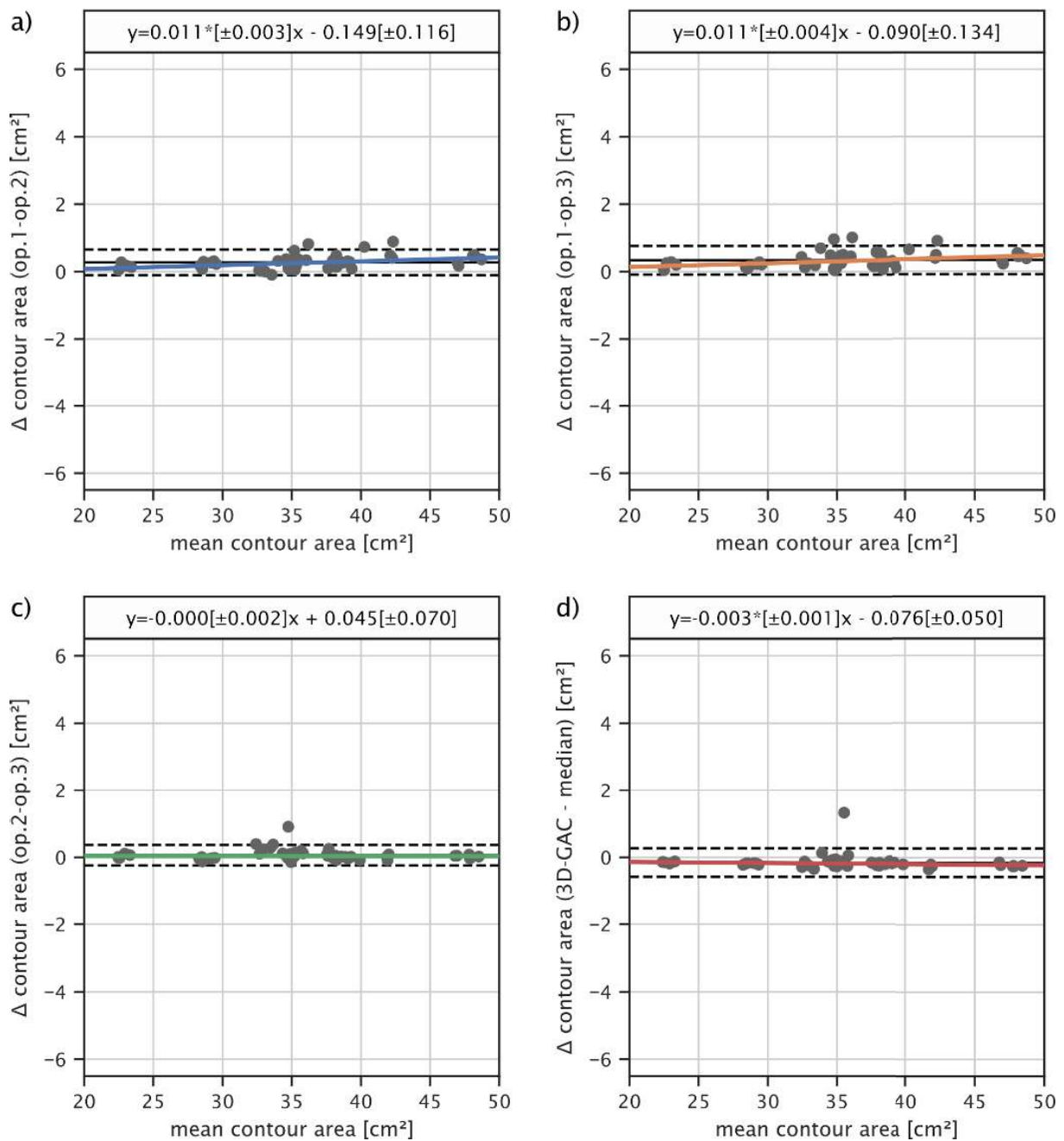


Fig. 3.4.3 Bland-Altman plots of the mean contour area per image. a) Comparing operator 1 (op. 1) to operator 2 (op. 2). b) Comparing operator 1 to operator 3 (op. 3). c) Comparing operator 2 to operator 3. d) Comparing the automatic 3D-GAC contours with the median (computed per slice) of all three operators. Area computed with the 3D-GAC approach agrees within the observable inter-operator variability with the area computed from hand-draw contours.

parts of the ulna while omitting large parts of the radius (N=10). The 3D-GAC approach failed due inclusion of background (N=8), large parts of the radius missing (N=1), missing parts of the radius and inclusion of parts of the ulna (N=2), and all mentioned failure modes combined

(N=1). The VGS for the failed contours of the manufacturer's approach ranged from one to five (N=7 for VGS 1, N=14 for VGS 2, N=16 for VGS 3, N=30 for VGS 4, N=27 for VGS 5). For the thirteen contours that only failed for the 3D-GAC approach (N=1 for VGS 2, N=5 for VGS 4, N=6 for VGS 5).

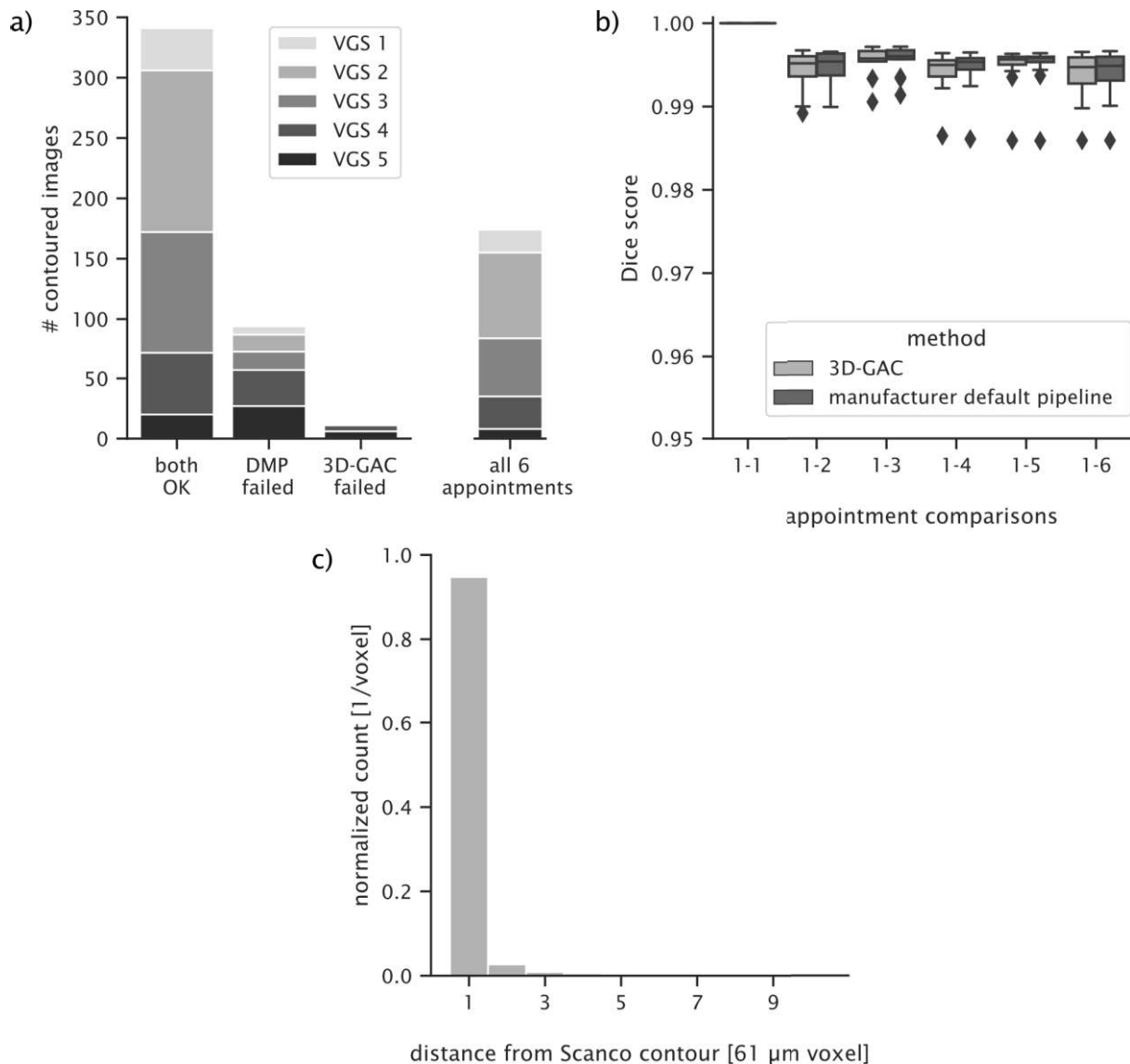


Fig. 3.4.4 Accuracy and reproducibility of the 3D-GAC vs the default manufacturer pipeline (DMP). a) Histogram of successful and failed contours for the 3D-GAC and the DMP with indicated image quality. The final bar shows the images selected for the reproducibility analysis. b) Dice scores of contours of all time points relative to the initial scan for both approaches show the reproducibility of each method. c) Histogram of the distance of all voxels different between the 3D-GAC and the DMP contour from the respective DMP contour.

Of the 341 images, 72 images (one image per patient) with sufficient image quality (VGS of three or better) and acceptable contours from both methods were used for further analysis. Overall agreement between both approaches was high with a mean Dice score of 0.996 ± 0.001 , with typical Dice score values for contours with missing parts of the radius being around 0.85. Reproducibility is on the same level as the default manufacturer pipeline, with agreement with the baseline image for both methods and for all appointments (median Dice score > 0.994 for all time-points) (Figure 3.4.4b). For those voxels that were either part of the 3D-GAC contour and not part of the contour from the default manufacturer pipeline or vice versa, 91% are at most one voxel away from the manufacturer's contour (Figure 3.4.4c). Bland-Atman plots reveal agreement in morphometric indices with a slight thickness and density based bias in cortical parameters and no bias in trabecular parameters (Figure 3.4.6). Mean deviations were below 1% for all parameters except for Ct.Th (1.341%) (Table 3.4.1).

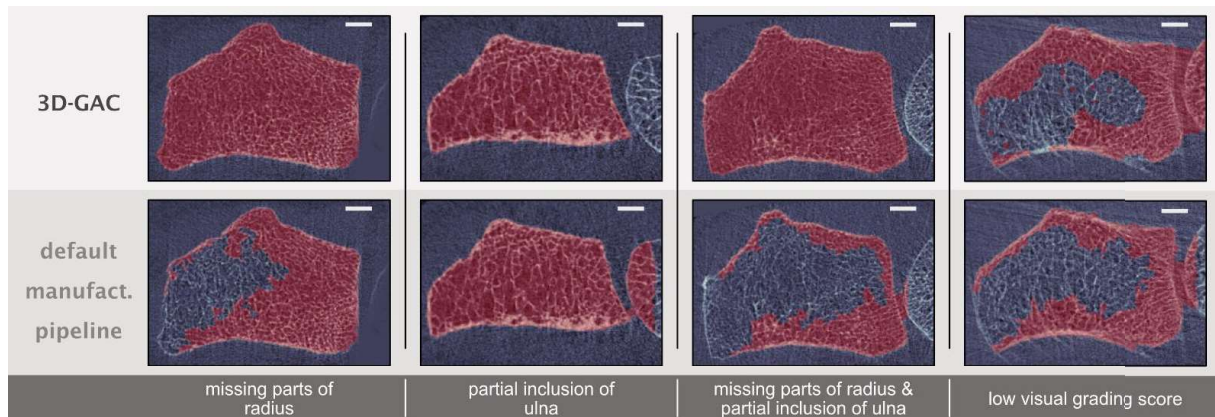


Fig. 3.4.5 Examples of failure modes when contouring distal radii. Contours from the 3D-GAC and the default manufacturer pipeline are compared for four challenging contouring cases: A thin cortex leading to missing parts of the radius, the ulna (partly being visible on the right) being close to the radius (bone on the left) and hence being included in the contour, the combination of the previous two, and a low visual grading score. Only the 3D-GAC approach manages to generate visually correct contours in most cases. Scale bar is 2 mm.

Table 3.4.1 Deviations of the bone morphometrics derived using 3D-GAC based contours compared to the manufacturer's default pipeline.

Morphometric Index	Deviation from Gold Standard
Tt.vBMD	0.190 ± 3.668%
Tb.vBMD	0.662 ± 3.226%
Ct.vBMD	-0.131 ± 2.437%
Tb.Ar	0.510 ± 4.920%
Ct.Ar	-0.710* ± 2.530%
Tb.BV/TV	0.540 ± 2.830%
Tb.N	-0.093 ± 1.497%
Tb.Th	0.147 ± 1.017%
Tb.Sp	0.181 ± 1.625%
Ct.Pm	-0.333 ± 2.402%
Ct.Th	-1.341* ± 4.911%
Morphometric Indices: total volumetric bone mineral density (Tt.vBMD), trabecular volumetric bone mineral density (Tb.vBMD), cortical volumetric bone mineral density (Ct.vBMD), trabecular area (Tb.Ar), cortical area (Ct.Ar), trabecular bone volume fraction (Tb.BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), cortical perimeter (Ct.Pm), cortical thickness (Ct.Th). *indicates significant deviation (P<0.05)	

3.4.4 Discussion

3.4.4.1 Fractured radii: Comparison of the 3D-GAC approach to hand-drawn contours

56 out of 60 images were contoured without major issues by the 3D-GAC approach. We observed clear inter-operator variability that is a known issue of hand-contouring bone (de Waard et al., 2018). On average, the generated contours by the 3D-GAC algorithm agree well with operators. The slightly larger inter-quartile range especially for appointments 2 and 3 can be explained by the formation of a callus in this time period and the resulting low mineralization

of background voxels visible in the HR-pQCT images. This makes it harder for both operators and the algorithm to decide on the correct contour. Bland-Altman plots show no systematic deviations between operators and the 3D-GAC approach that is larger than between operators themselves (Figure 3.4.3). Hence, the 3D-GAC contouring approach provides accurate contours for HR-pQCT images of fractured distal radii compared to the gold standard.

Interestingly, the default manufacturer pipeline only reliably contoured HR-pQCT images taken one year post fracture (Figure 3.4.1b). Given the fact that this pipeline was not designed to work with images of impaired cortices, this might indicate structural weaknesses in the cortex even 6 months post fracture. This observation is in agreement with recent findings suggesting that fracture repair is a long term process which might take as long as two years (de Jong et al., 2016).

3.4.4.2 Intact radii: Comparison of the 3D-GAC and manufacturer's approach

Comparing the 3D-GAC approach to the manufacturer default pipeline, the failure rate of the latter approach was much higher (94 vs 12 failed contours). Since this study used HR-pQCT images of fracture patients, the used cohort may inherently have poorer bone quality than standard populations, which could be one reason for the high number of failed contours for the default manufacturer approach. One of the major reasons for failure was the inclusion of the ulna into the outer contour of the radius, which is a known issue of this approach that is implementation dependent (Buie et al., 2007). The second reason for failure is having obvious missing parts, which appears to happen if the cortex has very thin or low-density structures (Figure 3.4.5). The fallback option of the default manufacturer pipeline is to create hand-drawn

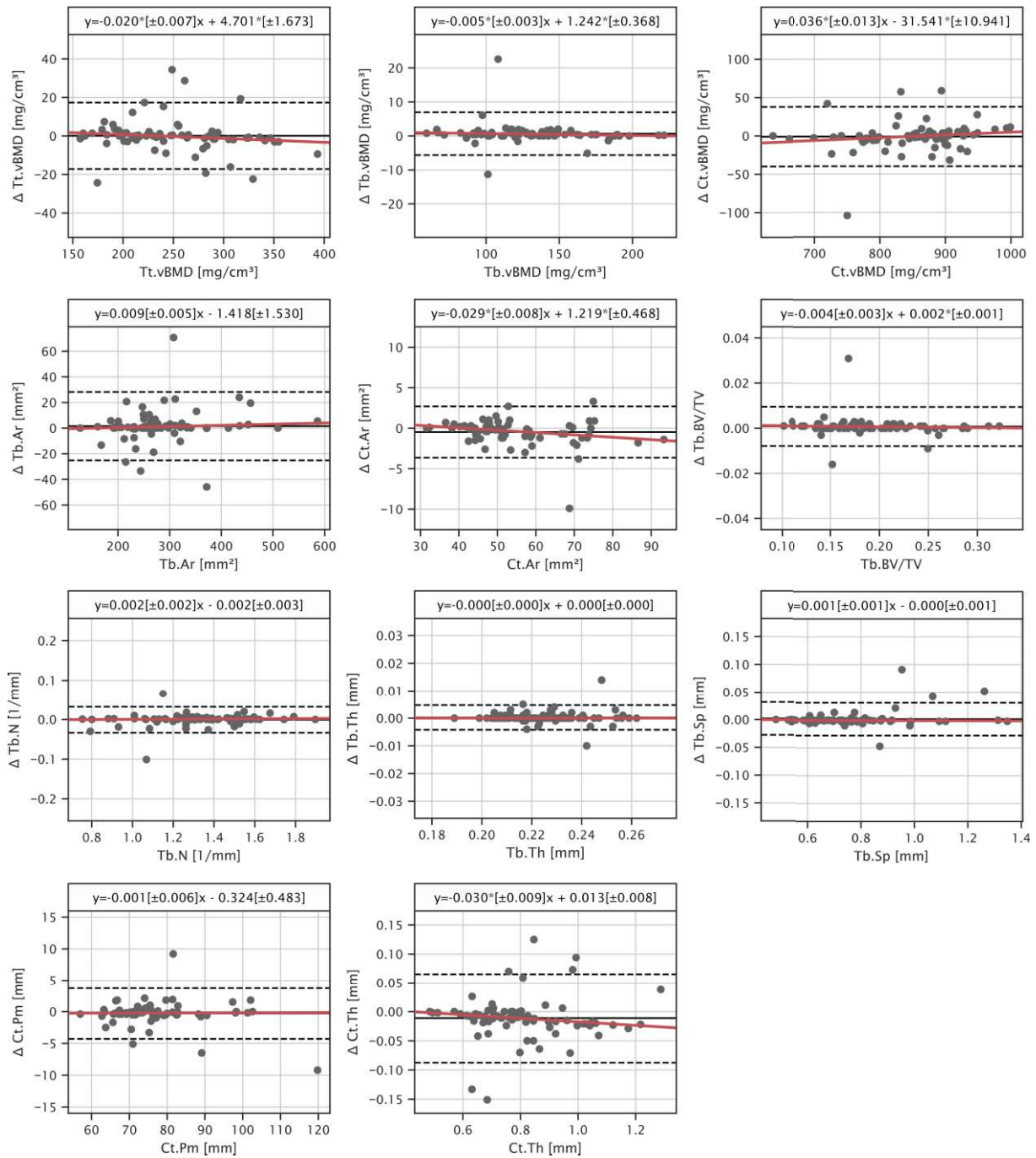


Fig. 3.4.6 Bland-Altman plots of morphometric indices characterize potential differences in 3D-GAC and contours from the default manufacturer pipeline. Robust linear model fits are shown in red. Regression parameters are presented above each plot with slope and intercept parameters statistically significantly different from zero being indicated by (*, P < 0.05). Indices are total volumetric bone mineral density (Tt.vBMD), trabecular volumetric bone mineral density (Tb.vBMD), cortical volumetric bone mineral density (Ct.vBMD), trabecular area (Tb.Ar), cortical area (Ct.Ar), trabecular bone volume fraction (Tb.BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), cortical perimeter (Ct.Pm), cortical thickness (Ct.Th).

contours, which can be done using a built-in tool of the manufacturer's software. Another proposed solution for this is to increase of the number of dilation steps of the algorithm (Buie et al., 2007), however this can lead to unwanted smoothing of the contour (Buie et al., 2007). Defining per-image numbers of dilation steps on the other hand would reduce the degree of automation drastically. The 3D-GAC approach only failed for one image with sufficient image quality to be typically included in HR-pQCT studies (de Waard et al., 2018; Manske et al., 2017; Nishiyama et al., 2015), showing and overall improvement in robustness compared to the current gold standard.

Accuracy of the generated contours was high with the mean Dice score being 0.996 ± 0.001 . Using these generated masks, trabecular morphometric indices agreed well with those generated using the gold standard method with deviations well below the precision error of these indices (Mueller et al., 2009). Interestingly, for the cortical parameters, accuracy was slightly worse with Ct.Th showing deviations of 1-2%; however, these differences are still less than 1 voxel in thickness and thus at the accuracy limit of the device (Figure 3.4.6). Furthermore, differences in cortical thickness were at the reproducibility limit of the standard patient evaluation (1.3-3.9% dependent on subject age) (Burghardt et al., 2010a) and were still below differences observed between patient groups from the literature, e.g. healthy and osteopenic women (12.8%) (Nishiyama et al., 2010), indicating that the 3D-GAC approach is a viable option for clinical studies of non-fractured radii as well.

3.4.4.3 Limitations

This study is not without limitations. The repeat scans of the contralateral site were spaced over a year, thus may include variations in bone due to natural remodelling. Additional scans for reproducibility were not possible due to radiation concerns. However, it would be expected that these natural variations would only decrease the reproducibility found. The same intervals between scans was used for the images of the fractured radii. Due to the large structural changes happening during fracture repair, the images of the fracture site could not be regarded as repeat scans. Therefore, no direct assessment of reproducibility was possible for images of fractured radii. However, the high accuracy observed for 60 images does indicate also high levels of reproducibility. Another limitation is the low number of hand-drawn contour slices (270). While time was a limiting factor (operators reporting to take roughly 1-2 minutes per slice) for creating more hand-drawn contours, we included images of the intact contralateral site to cover a large

variety of bone morphologies, which we could benchmark against existing automatic approaches for intact radii. Another limitation is that the 3D-GAC contours were generated only proximal to the lunate fossa. However, the morphometric analysis of the standard patient evaluation has only been developed for that region, making this region the most relevant for clinical studies. Furthermore, no hand-corrected contours were generated for intact radii to estimate the bias introduced by the automated protocol. However, recent studies have already done this for the gold standard (Whittier et al., 2020) and, given the good agreement of the 3D-GAC approach with the gold standard, the requirement for minor manual corrections should be on a similar level.

3.4.5 Conclusion

Our proposed 3D-GAC algorithm provides a unified pipeline for generating outer contours from HR-pQCT images of distal radii, including both fractured and intact bones. Importantly, our method automatically contours radii both accurately and reproducibly, while showing robustness when dealing with a wide variety of cortex structures. This will facilitate future clinical studies using large patient cohorts to support the development of improved treatment protocols of radius fractures.

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.

Conflict of Interest

The authors declare that they have no competing interests.

References

- Buie, H.R., Campbell, G.M., Klinck, R.J., MacNeil, J.A., Boyd, S.K., 2007. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* 41, 505–515. <https://doi.org/10.1016/j.bone.2007.07.007>
- Burghardt, A.J., Buie, H.R., Laib, A., Majumdar, S., Boyd, S.K., 2010a. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. *Bone* 47, 519–528. <https://doi.org/10.1016/j.bone.2010.05.034>
- Burghardt, A.J., Kazakia, G.J., Majumdar, S., 2007. A local adaptive threshold strategy for high resolution peripheral quantitative computed tomography of trabecular bone. *Ann. Biomed. Eng.* 35, 1678–1686. <https://doi.org/10.1007/s10439-007-9344-4>
- Burghardt, A.J., Kazakia, G.J., Sode, M., De Papp, A.E., Link, T.M., Majumdar, S., 2010b. A longitudinal HR-pQCT study of alendronate treatment in postmenopausal women with low bone density: Relations among density, cortical and trabecular microarchitecture, biomechanics, and bone turnover. *J. Bone Miner. Res.* 25, 2558–2571. <https://doi.org/10.1002/jbmr.157>
- Burt, L.A., Hanley, D.A., Boyd, S.K., 2017. Cross-sectional Versus Longitudinal Change in a Prospective HR-pQCT Study. *J. Bone Miner. Res.* 32, 1505–1513. <https://doi.org/10.1002/jbmr.3129>
- Caselles, V., Kimmel, R., Sapiro, G., 1997. Geodesic Active Contours. *Int. J. Comput. Vis.* 22, 61–79. <https://doi.org/10.1023/A:1007979827043>
- Court-Brown, C.M., Caesar, B., 2006. Epidemiology of adult fractures: A review. *Injury* 37, 691–697. <https://doi.org/10.1016/j.injury.2006.04.130>
- de Jong, J.J., Heyer, F.L., Arts, J.J.C., Poeze, M., Keszei, A.P., Willems, P.C., van Rietbergen, B., Geusens, P.P., van den Bergh, J.P., 2016. Fracture Repair in the Distal Radius in Postmenopausal Women: A Follow-Up 2 Years Postfracture Using HRpQCT. *J. Bone Miner. Res.* 31, 1114–1122. <https://doi.org/10.1002/jbmr.2766>

- De Jong, J.J.A., Willems, P.C., Arts, J.J., Bours, S.G.P., Brink, P.R.G., van Geel, T.A.C.M., Poeze, M., Geusens, P.P., van Rietbergen, B., van den Bergh, J.P.W., 2014. Assessment of the healing process in distal radius fractures by high resolution peripheral quantitative computed tomography. *Bone* 64, 65–74. <https://doi.org/10.1016/j.bone.2014.03.043>
- de Waard, E.A.C., Sarodnik, C., Pennings, A., de Jong, J.J.A., Savelberg, H.H.C.M., van Geel, T.A., van der Kallen, C.J., Stehouwer, C.D.A., Schram, M.T., Schaper, N., Dagnelie, P.C., Geusens, P.P.M.M., Koster, A., van Rietbergen, B., van den Bergh, J.P.W., 2018. Reliability of HR-pQCT Derived Cortical Bone Structural Parameters When Using Uncorrected Instead of Corrected Automatically Generated Endocortical Contours in a Cross-Sectional Study: The Maastricht Study. *Calcif. Tissue Int.* 103, 252–265. <https://doi.org/10.1007/s00223-018-0416-2>
- Ellouz, R., Chapurlat, R., van Rietbergen, B., Christen, P., Pialat, J.B., Boutroy, S., 2014. Challenges in longitudinal measurements with HR-pQCT: Evaluation of a 3D registration method to improve bone microarchitecture and strength measurement reproducibility. *Bone* 63, 147–157. <https://doi.org/10.1016/j.bone.2014.03.001>
- Hafri, M., Jennane, R., Lespessailles, E., Toumi, H., 2016a. Dual active contours model for HR-pQCT cortical bone segmentation, in: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, pp. 2270–2275. <https://doi.org/10.1109/ICPR.2016.7899974>
- Hafri, M., Toumi, H., Boutroy, S., Chapurlat, R.D., Lespessailles, E., Jennane, R., 2016b. Fuzzy energy based active contours model for HR-PQCT cortical bone segmentation, in: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 4334–4338. <https://doi.org/10.1109/ICIP.2016.7533178>
- Johnell, O., Kanis, J.A., 2006. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos. Int.* 17, 1726–1733. <https://doi.org/10.1007/s00198-006-0172-4>
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *Int. J. Comput. Vis.* 1, 321–331. <https://doi.org/10.1007/BF00133570>
- Manske, S.L., Davison, E.M., Burt, L.A., Raymond, D.A., Boyd, S.K., 2017. The Estimation of Second-Generation HR-pQCT From First-Generation HR-pQCT Using In Vivo

- Cross-Calibration. *J. Bone Miner. Res.* 32, 1514–1524. <https://doi.org/10.1002/jbmr.3128>
- Marquez-Neila, P., Baumela, L., Alvarez, L., 2014. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 2–17. <https://doi.org/10.1109/TPAMI.2013.106>
- Mueller, T.L., Stauber, M., Kohler, T., Eckstein, F., Müller, R., van Lenthe, G.H., 2009. Non-invasive bone competence analysis by high-resolution pQCT: An in vitro reproducibility study on structural and mechanical properties at the human radius. *Bone* 44, 364–371. <https://doi.org/10.1016/j.bone.2008.10.045>
- Ng, C.Y., McQueen, M.M., 2011. What are the radiological predictors of functional outcome following fractures of the distal radius? *J. Bone Jt. Surg. - Ser. B* 93 B, 145–150. <https://doi.org/10.1302/0301-620X.93B2.25631>
- Nishiyama, K.K., Macdonald, H.M., Buie, H.R., Hanley, D.A., Boyd, S.K., 2010. Postmenopausal women with osteopenia have higher cortical porosity and thinner cortices at the distal radius and tibia than women with normal aBMD: An in vivo HR-pQCT study. *J. Bone Miner. Res.* 25, 882–890. <https://doi.org/10.1359/jbmr.091020>
- Nishiyama, K.K., Pauchard, Y., Nikkel, L.E., Iyer, S., Zhang, C., McMahon, D.J., Cohen, D., Boyd, S.K., Shane, E., Nickolas, T.L., 2015. Longitudinal HR-pQCT and image registration detects endocortical bone loss in kidney transplantation patients. *J. Bone Miner. Res.* 30, 456–463. <https://doi.org/10.1002/jbmr.2358>
- Pialat, J.B., Burghardt, A.J., Sode, M., Link, T.M., Majumdar, S., 2012. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. *Bone* 50, 111–118. <https://doi.org/10.1016/j.bone.2011.10.003>
- Python Software Foundation, 2020. Python Language Reference, Version 3.7 [WWW Document]. URL <https://www.python.org>
- Seabold, S., Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference.

- Shanbhogue, V. V., Støvning, R.K., Frederiksen, K.H., Hanson, S., Brixen, K., Gram, J., Jørgensen, N.R., Hansen, S., 2017. Bone structural changes after gastric bypass surgery evaluated by HR-pQCT: A two-year longitudinal study. *Eur. J. Endocrinol.* 176, 685–693. <https://doi.org/10.1530/EJE-17-0014>
- The HDF Group, 2020. Hierarchical Data Format, version 5 [WWW Document]. URL <http://www.hdfgroup.org/HDF5/>
- Thévenaz, P., Ruttimann, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41. <https://doi.org/10.1109/83.650848>
- Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. <https://doi.org/10.1109/MCSE.2011.37>
- Van Der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. Scikit-image: Image processing in python. *PeerJ* 2014, 1–18. <https://doi.org/10.7717/peerj.453>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Contributors, S. 1. 0, 2019. SciPy 1.0-- Fundamental Algorithms for Scientific Computing in Python 1–22.
- Whittier, D.E., Mudryk, A.N., Vandergaag, I.D., Burt, L.A., Boyd, S.K., 2020. Optimizing HR-pQCT workflow: a comparison of bias and precision error for quantitative bone analysis. *Osteoporos. Int.* 31, 567–576. <https://doi.org/10.1007/s00198-019-05214-0>
- Zebaze, R., Ghasem-Zadeh, A., Mbala, A., Seeman, E., 2013. A new method of segmentation of compact-appearing, Transitional and trabecular compartments and quantification of cortical porosity from high resolution peripheral quantitative computed tomographic images. *Bone* 54, 8–20. <https://doi.org/10.1016/j.bone.2013.01.007>

Chapter 4

Longitudinal assessment of bone mechanoregulation of fracture healing

4.1 Mechanoregulation in distal radius fractures assessed *in vivo* by time-lapsed HR-pQCT images

N. Ohs¹, F.C. Marques¹, P. R. Atkins¹, C. J. Collins¹, M. Blauth², P.Christen^{1,3}, D. C. Tourolle né Betts¹, R. Müller¹

¹Institute for Biomechanics, ETH Zurich, Zurich, Switzerland

²Department for Trauma Surgery, Innsbruck University Hospital, Innsbruck, Austria

³Institute for Information Systems, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

Abstract

Radius fractures are among the most common fracture types; however, there is limited consensus on the optimal treatment protocol. A better understanding of the fracture healing process could help to shape future protocols and thus improve functional outcomes of patients. High-resolution peripheral quantitative computed tomography (HR-pQCT) allows close monitoring and evaluation of the radius on the micro-structural level, which is crucial to our understanding of the fracture healing process. Since local mechanical tissue loading is essential for bone adaptation and fracture healing, the question of whether or not HR-pQCT can be utilized to study mechanoregulation in distal radii during the fracture healing process arises.

In the present study, we investigate local mechanoregulation in HR-pQCT of fracture patients. For this, *in silico* bone adaptation simulations are first used to test the ability of an existing micro-CT-based method for quantifying mechanoregulation to handle HR-pQCT images as input. Three loading scenarios, (i) homeostatic, (ii) catabolic, and (iii) anabolic bone adaptation, were simulated. For all scenarios, images were downscaled from 25 μm to various resolutions and various levels of noise were added, including the resolution and noise level of second generation HR-pQCT (61 μm). In virtual experiments, images with added noise and reduced resolution from these simulations were taken and the correct classification of remodelling events were quantified. Second, mechanoregulation of seven patients with distal radius fractures whose fractured and contralateral radii were scanned six times over the course of a year was quantified using the tested method.

Virtual experiments on *in silico* data show that mechanoregulation can be extracted from images with resolution and signal to noise ratio matching clinical HR-pQCT. Mechanoregulation was detected in the patient data with the formation and resorption processes during fracture repair being more mechanosensitive.

Keywords:

HR-pQCT; fracture; radius; mechanoregulation; bone adaptation; fracture healing

4.1.1 Introduction

Bone fractures are increasingly common and have become a major socioeconomic burden in our ageing population (Johnell and Kanis, 2006). Even though fractures of the distal radius are among the most common fracture types and indicative of a reduction in bone quality (Court-Brown and Caesar, 2006), there is currently limited consensus on the optimal treatment protocol (Ng and McQueen, 2011). Conservative treatment of many radius fractures, namely cast immobilization, provides a unique opportunity to study the fracture healing process in humans. A deeper insight into what drives this process could help to shape future treatment protocols of distal radius fractures, ultimately resulting in better functional outcomes for patients.

Local mechanical tissue loading is essential for both bone adaptation and remodelling as well as fracture healing (Augat et al., 2005, 1996; Boerckel et al., 2012; Carter, 1987; Carter et al., 1998, 1987; Claes et al., 1998; Giannoudis et al., 2007; Huiskes et al., 2000; Isaksson et al., 2006; Klein et al., 2003). In a mouse defect-healing model, microstructural tissue formation has already been linked to the local mechanical environment using longitudinal, micro-CT imaging (Tourolle né Betts et al., 2020). With the availability of high-resolution peripheral quantitative computed tomography (HR-pQCT) in clinics, it has now become possible to non-invasively and longitudinally observe bone microstructural changes in patients (Burghardt et al., 2010; Burt et al., 2017; Ellouz et al., 2014; Nishiyama et al., 2015; Shanbhogue et al., 2017). Furthermore, efforts have been made to combine HR-pQCT images with finite element analysis to obtain information on the local mechanical environment (Ohs et al., 2020a). Preliminary findings indicate that these techniques are also viable to monitor fracture healing in patients, longitudinally (De Jong et al., 2014).

To assess the local mechanoregulation, the mechanical loading at the tissue level has to be correlated with bone tissue formation and resorption. Only one study so far has attempted this using HR-pQCT images (Christen et al., 2014). However, this pilot study used an alternative method than the established one for animal experiments introduced by Schulte and colleagues (Schulte et al., 2011), making comparisons to the existing literature difficult. Using the approach developed for animal experiments on HR-pQCT images, on the other hand, is not trivial as this approach was developed for micro-CT images. The difference in image modality between HR-pQCT and micro-CT images can influence the outcome of image analysis tools

such as segmentation and filtering, mandating a separate validation of any tool developed for micro-CT for use with HR-pQCT images (Ohs et al., 2020a).

The challenge with validating longitudinal assessment approaches, such as the quantification of mechanoregulation, is finding a proper ground truth. Typically, cadaveric bone tissue is used in the validation of methods for processing and analysis of HR-pQCT data as they can be scanned with high-resolution micro-CT. However, cadaveric samples do not enable the assessment of longitudinal data. Therefore, this common validation technique cannot be used for the validation of any method quantifying *in vivo* mechanoregulation. Instead, the use of simulated data has been suggested as a possible ground truth for the validation of longitudinal assessment approaches (Ohs et al., 2020a). With recent advancements in bone adaptation simulations, it is now possible to simulate bone formation and resorption in response to mechanical loading on HR-pQCT images (Ohs et al., 2020b). These simulations not only provide data at a high resolution (25 μm) compared to HR-pQCT (61 μm) but also have the advantage of every occurring adaptation event being 100% driven by mechanics, which would not be known for any scanned *in vivo* bone structure. Furthermore, simulations allow one to obtain diverse adaptation scenarios easily; for example, homeostatic or anabolic bone (re)modelling, which result in a more robust validation. Finally, these simulations can be used to conduct virtual experiments with added noise and/or reduced image resolution to study differences in image modalities and their impact on the resultant mechanoregulation.

The goal of the present study was to investigate mechanoregulation during fracture healing using clinical resolution HR-pQCT. We first quantified to which extent mechanoregulation can be extracted from HR-pQCT images using the approach of Schulte and colleagues (Schulte et al., 2011) by performing virtual experiments on simulated data for three bone adaptation scenarios: (i) homeostatic, (ii) catabolic, (iii) and anabolic bone adaptation. Here, we accounted for variations in signal to noise ratio and image resolution. Then, we investigated differences in mechanoregulation between fractured and intact distal radii of seven patients using HR-pQCT images that were acquired throughout the first year of the healing process.

4.1.2 Materials and Methods

4.1.2.1 Data

HR-pQCT images (XtremeCT II, Scanco Medical AG, Switzerland; 60.7 μm voxels, 68 kV, 1470 μA , 43 ms integration time) were obtained from the database of a previous fracture study conducted at Innsbruck Medical University. Patients provided informed consent and participated in a study approved by the ethics committee of the Medical University of Innsbruck. For each of the 75 patients that completed the study, six scans of the fractured (up to 504 slices per scan distributed over three stacks) and contralateral (up to 168 slices per scan, one stack only) radius were taken over the course of one year (1, 3, 5, 13, 26, and 52 weeks post-fracture). A subset of patients were selected from the database based on visual grading scores (VGS) (Pialat et al., 2012) of three or better for all images and minimal stack-artefacts as well as low fracture fragment movement. Data from one patient was selected for the bone adaptation simulations. Data from seven others were selected for the mechanoregulation analysis.

4.1.2.2 Image pre-processing

Images of the contralateral radii were registered using a pyramid-based rigid registration approach (Thévenaz et al., 1998). For images of the fractured radii, the middle stacks were registered first using the same method, followed by the outer stacks using the temporal misalignment of the centre stack to find overlapping regions of outer stacks to centre stacks of other time points. Any gaps between stacks were filled by linearly interpolating between the two adjacent slices. Outer contours of the radius were generated for each image using a previously developed algorithm (Chapter 3.4).

4.1.2.3 Mechanoregulation analysis

Surface-based mechanoregulation analysis was performed on segmented images as is described in the work of Schulte and colleagues (Schulte et al., 2011). To quantify the amount of correctly classified mechanoregulation, the correct classification rate (CCR) was used as described by Tourolle and colleagues (Tourolle né Betts et al., 2020).

4.1.2.4 Trabecular region mask

Since mechanoregulation is analysed only for the trabecular region in the current study, a mask of this region was automatically generated for each image using the following steps. First, the density values in the image were rescaled to the interval [0, 1]. Next, the rescaled density image and the outer contour were mirrored along the z-axis (longitudinal axis of the radius) at the top and bottom of the image to provide a continuous boundary at these edges. A binary erosion operation was then applied to the outer contour for one iteration, assuming a ‘von Neumann’ neighbourhood. Rescaled density values outside the outer contour mask were set to 1 and a Gaussian filter was applied with a sigma of 22, truncated at 4 standard deviations. The filtered density was binarized by only keeping values above 0.5. The greatest connected component of this structure, i.e. the combined background and high-density cortex, was identified and binary inverted, resulting in a mask only covering the trabecular region. Finally, the mirrored slices at the top and bottom of this mask were removed and the common region of all time-points per patient were used as the mask for this patient.

4.1.2.5 Bone adaptation simulation

The mechanostat based bone adaptation simulation by Schulte and colleagues (Schulte et al., 2013) was used with the implementation in Python being described in chapter 3.3 to generate a series of images with changes in bone structure that are 100% mechanically driven. The simulation was only run on the masked trabecular region of the selected sample.

4.1.2.5.1 Input model generation

Bone volume fraction (BV/TV) of the sample was determined by applying a threshold of 320 mg HA/cm³ to the image. The image was rescaled from the original voxel size of 61 to 25 μ m using cubic interpolation to achieve accurate results, and simulation inputs were generated from this image as was done in chapter 3.3. To generate a valid input model for the simulation, a threshold was applied that resulted in identical BV/TV as was computed for the original resolution image. Then, surface voxels of this intermediate binary structure were filled with the original grayscale value of each voxel relative to the applied threshold, respectively, and the structure was multiplied by a density value of 750 mg/cm³. Finally, the greatest connected component of the structure was identified using component labelling and disconnected bone fragments were discarded since no realistic loading can be calculated for them.

4.1.2.5.2 Boundary Conditions

For the micro-finite element (micro-FE) step of the simulation, a longitudinal (z-axis), high-friction compressive boundary condition was chosen. The applied compressive force was estimated using a force estimation algorithm (Christen et al., 2013), considering 0.02 MPa as the reference value for the target homogeneous strain energy density value in the structure (Badilatti et al., 2016; Mullender and Huiskes, 1995).

4.1.2.5.3 Adaptation scenarios

The mechanostat used was identical to the one used in a previous study (Ohs et al., 2020b). The simulation time was set to five years, from which the first 1.5 years were used as an adaptation phase to the new boundary conditions and dismissed (Badilatti et al., 2016; Schulte et al., 2013), yielding 25 iteration steps of the remaining 3.5 years for three adaptation scenarios based on changes in BV/TV: (i) homeostatic, (ii) catabolic, (iii) and anabolic bone adaptation. The anabolic and catabolic cases were defined by a change in BV/TV of approximately 5% at the end of the simulation period (Badilatti et al., 2016), which was achieved by scaling the boundary forces by 1.5 (catabolic), 2.6 (homeostatic), 4.6 (anabolic). The simulations were performed at the Swiss National Supercomputing Centre (Piz Daint, CSCS, Lugano, Switzerland).

4.1.2.5.4 Virtual experiments

Virtual experiments were performed on the simulation data to determine to what extent the mechanoregulation governing the simulation could be recovered given different image voxel sizes and noise levels.

For this, the simulated data (25 images for each of the three adaptation scenarios, respectively) with a voxel size of 25 μm was downsampled to three different resolutions covering, in total, the range from micro-CT resolution to clinically available HR-pQCT resolution: 40 μm , 61 μm and 82 μm . Typical noise levels measured in HR-pQCT images followed a Gaussian distribution with a standard deviation of approximately 100 mg HA/cm³. Therefore, from each set of the 25 μm and rescaled images, seven images were adapted by adding a Gaussian noise distribution with a standard deviation ranging from 0 to 120 mg HA/cm³ in steps of 20 mg HA/cm³.

For each of the 28 different configurations (4 resolutions times 7 noise levels), FE simulations were run on all images using a longitudinal compression boundary condition. The applied force for each configuration was determined by running the force estimation (Christen et al., 2013) algorithm on the last image of the respective adaptation scenario for that configuration.

Finally, for each configuration, mechanoregulation was extracted for all possible time steps and the median, minimum, and maximum CCR was computed for all configurations.

4.1.2.6 HR-pQCT grayscale-based micro-FE

Registered images were cropped to only contain parts of the bone visible in all images across all time points. Next, images were filtered using a Gaussian filter ($\sigma = 0.8$, support = 1). Grayscale images (units mg HA/cm³, provided by the manufacturer software) were converted to Young's modulus using a linear relationship proposed by Mulder and colleagues (Mulder et al., 2007). Additionally, values below 0.2 MPa were set to 0.2 MPa to avoid convergence issues in the analysis. Linear micro finite element analysis (micro-FEA) was run for each image using the matrix-free and octree-based parallel solver parOsol (Flaig, 2012). The boundary condition set for the micro-FEA was 1% compression in the longitudinal direction with a fully fixed bottom slice, dubbed "high-friction compression" in the literature. The effective strain (EFF) results were saved for each analysis for further processing, since partial volume effects in grayscale-based FEA lead to unrealistic surface SED values.

4.1.2.7 Morphometric indices

Morphometric indices for the contralateral radii were computed using the scanner manufacturer's default pipeline but with the outer contours generated as stated above, which has been shown to yield accurate results (chapter 3.4). Computed indices were: trabecular volumetric bone mineral density (Tb.vBMD), trabecular bone volume fraction (Tb.BV/TV), trabecular area (Tb.Ar), trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp).

4.1.2.8 Mechanoregulation analysis on patient data

As stated before, mechanoregulation was computed using the approach by Schulte and colleagues (Schulte et al., 2011). To obtain the necessary binary inputs, threshold of 320 mg HA/cm³ was applied to each Gaussian filtered image (same parameters as for the grey-scale micro-FE), which is the default threshold of the manufacturer for trabecular bone. The CCR was computed for each consecutive pair of images (weeks: 1-3, 3-5, 5-13, 13-26, 26-52), respectively. EFF was used for the local mechanics and all EFF values were cut off at and normalized by the value of 0.258, which was visually determined as the best value to exclude

unrealistic (extremely high) EFF values from the FE analysis across all samples while discarding as little EFF values as possible.

4.1.2.9 Statistics

Statistical significance for the comparisons of CCR between fractured and intact radii were performed using a paired Student's t-test. Significance level was set to $p < 0.05$. Holm-Bonferroni corrections for multiple comparisons were applied.

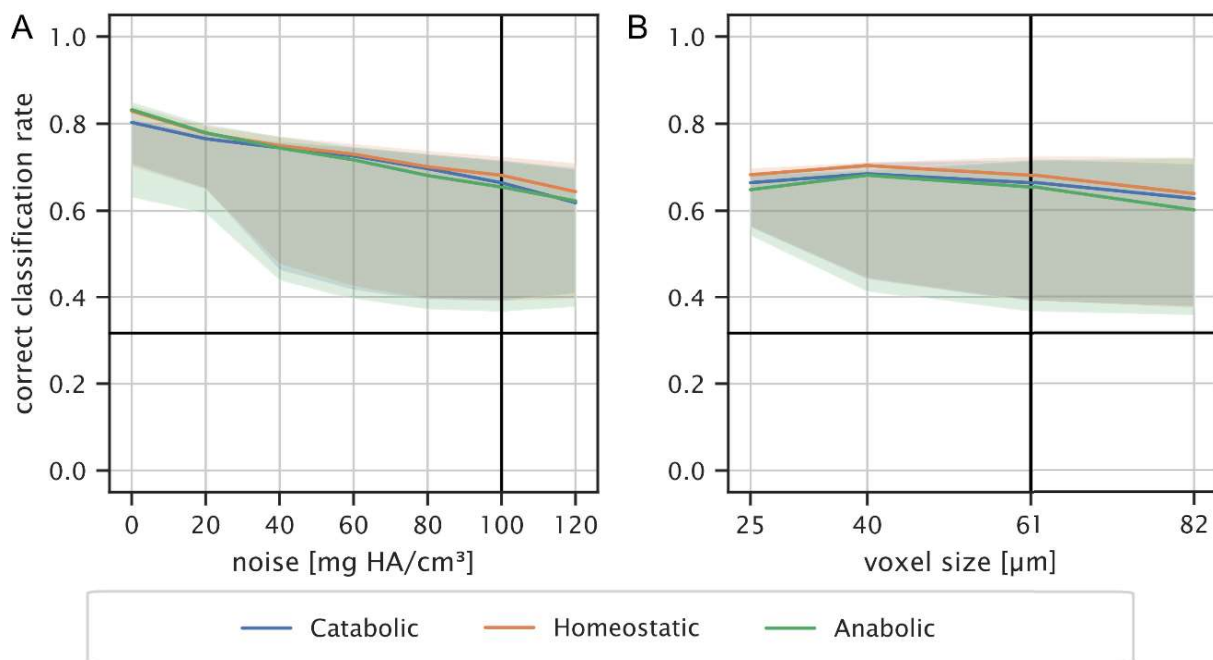


Fig. 4.1.1 HR-pQCT noise levels and voxel size allow for assessment of mechanoregulation in patient data. Correct classification rate (CCR) of mechanoregulation is shown for data with (A) native HR-pQCT resolution with different levels of noise and (B) clinical HR-pQCT noise levels with different image resolutions. Coloured lines show median values of CCR for three simulated adaptation scenarios: catabolic, homeostatic, and anabolic. Coloured shades show the range between max and min CCR found. Vertical lines show native HR-pQCT noise levels and voxel size, respectively. Horizontal lines show the 33% line, i.e. the random classification/detection limit.

4.1.3 Results

Virtual experiments on *in silico* bone adaptation simulations showed median CCRs of larger than 60% for all three adaptation scenarios and all tested voxel sizes and noise levels (Figure 4.1.1). Minimum CCRs were observed for images with close to no structural differences, but

were still above the 61 μm detection limit (Figure 4.1.1). Median CCR for native HR-pQCT noise levels and resolution were: 68% (homeostatic), 66% (catabolic), and 65% (anabolic). These values are lower than what can be found for the highest tested resolution (25 μm) and no added noise: 78% (homeostatic), 73% (catabolic), and 77% (anabolic).

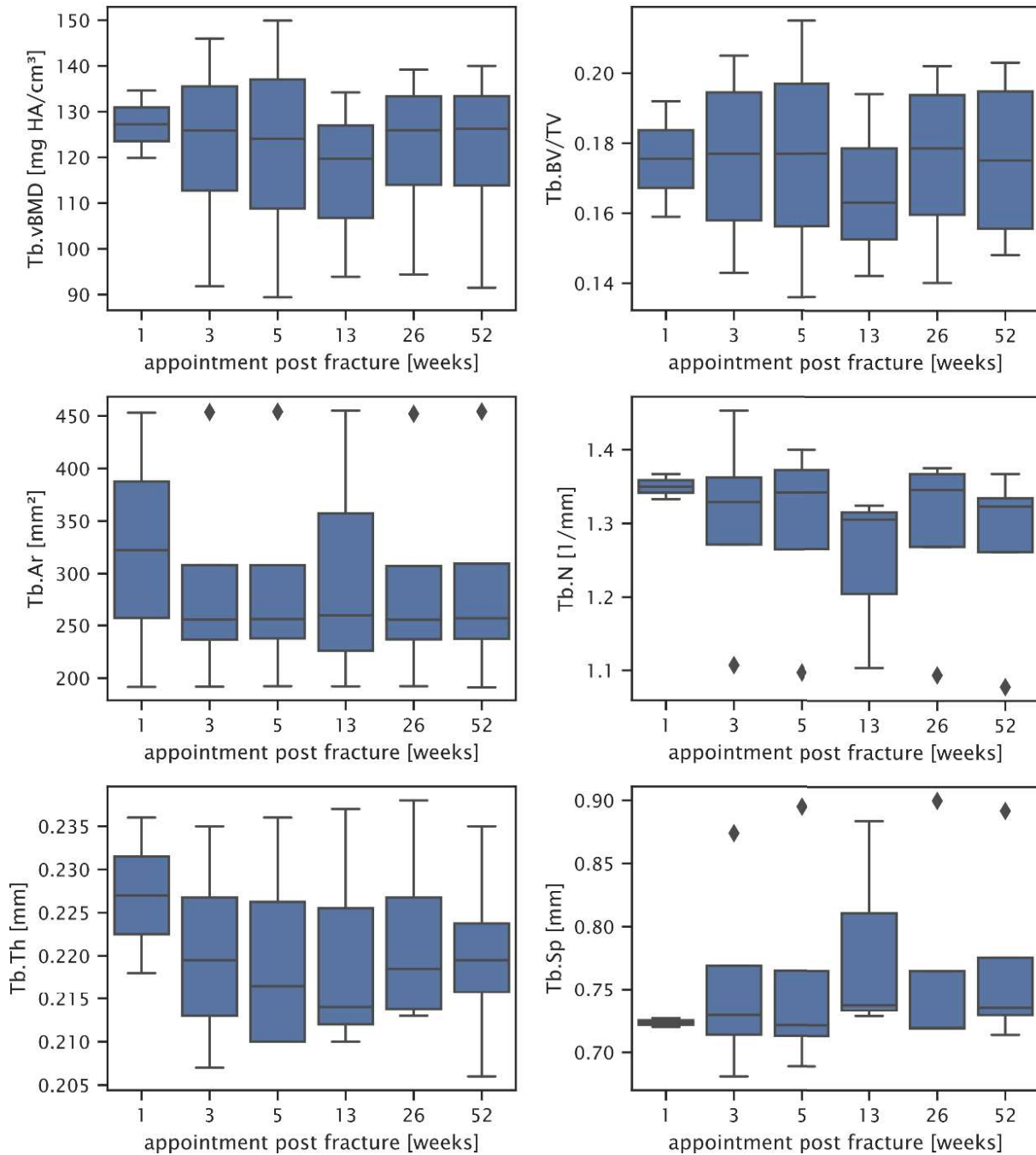


Fig. 4.1.2 Trabecular morphometric indices of the contralateral site are stable over time. Assessed parameters include: trabecular volumetric bone mineral density (Tb.vBMD),

trabecular bone volume fraction (Tb.BV/TV), trabecular area (Tb.Ar), trabecular number (Tb.N), trabecular thickness (Tb.Th), and trabecular separation (Tb.Sp).

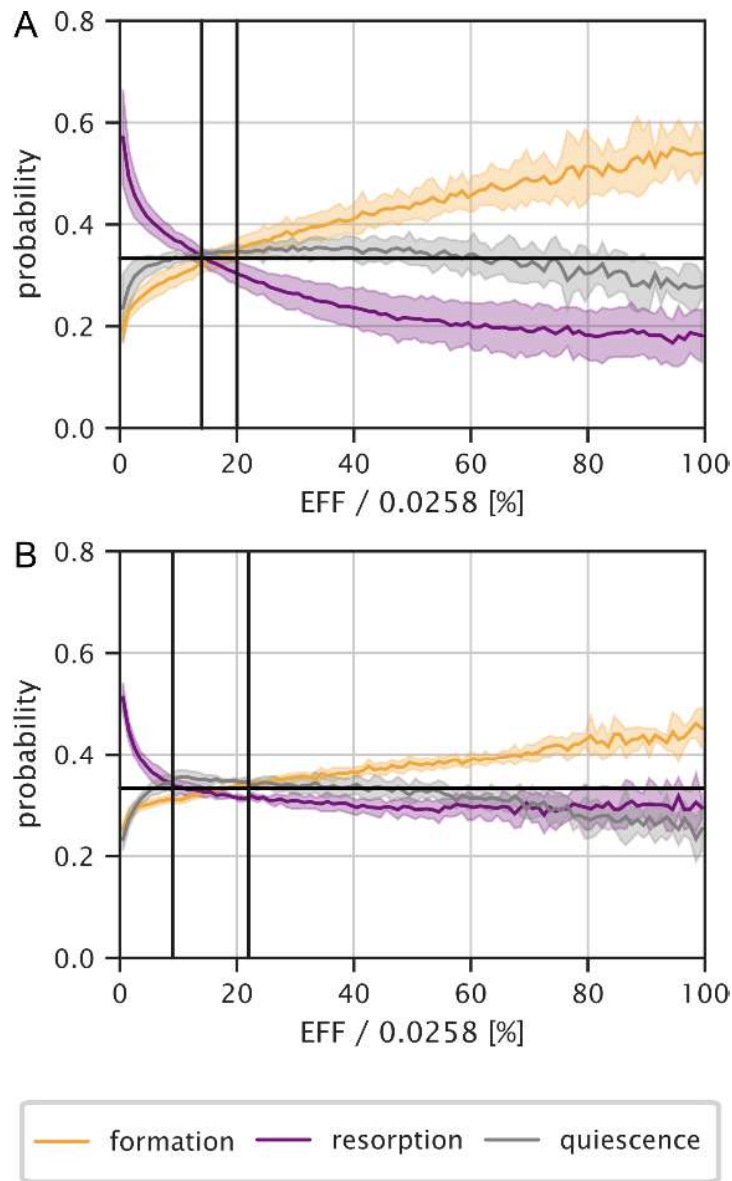


Fig. 4.1.3 Events of formation, quiescence, and resorption are better associated with mechanics in (A) fractured than in (B) intact radii. Per effective strain (EFF) bin, normalized probability for the three events is shown for different levels of EFF. Random classification (33%) is indicated by horizontal bold lines. Vertical lines show division of EFF into intervals of correct classification rate maximizing (left to right) resorption, quiescence, and formation.

Trabecular indices for the contralateral site were stable over time (Figure 4.1.2). Averages over all seven patients and time were: Tb.vBMD (121.37 ± 18.95 mg HA/cm³), Tb.BV/TV (0.17 ± 0.02), Tb.N (1.29 ± 0.12 mm⁻¹), Tb.Th (0.22 ± 0.01 mm), and Tb.Sp (0.76 ± 0.07 mm).

Visually, correct prediction of local remodelling was higher for the fractured radii compared to the intact contralateral radii when combining all time intervals (Figure 4.1.3). This can also be seen in the correct classification percentages of formation, resorption, and quiescence for the lowest and highest EFF values (Table 4.1.1). CCRs were 0.44 for the fractured and 0.39 for the intact radii when combining all time intervals. For the fractured side, classification of quiescence was close to random for most EFF values, while formation and resorption were determined by mechanics (Figure 4.1.3A). For the contralateral side, higher resorption occurred for low EFF; however, for high EFF, quiescence and resorption were close to random (Figure 4.1.3B). Conversely, formation events for the contralateral site were determined by mechanics over the entire range of EFF values (Figure 4.1.3B).

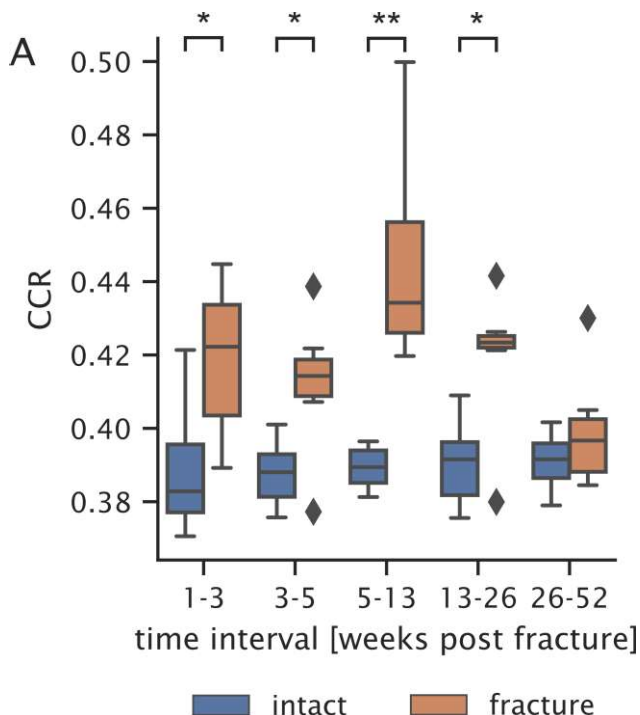


Fig. 4.1.4 Mechanics predict local bone formation, quiescence, and resorption better in fractured compared to intact radii for up to 26 weeks post fracture. Combined correct classification rate (CCR) of formation, resorption, and quiescence is plotted for fractured and corresponding contralateral radii scanned over the course of one year. Significance was determined using a paired t-test with: * $p < 0.05$, ** $p < 0.01$. Holm-Bonferroni was used to correct for multiple comparisons.

For the first half-year post fracture, CCR of mechanoregulation was significantly higher for the fracture site compare to the contralateral site ($p < 0.05$ for weeks 1 to 3, 3 to 5, and 13 to 26; $p < 0.01$ for weeks 5 to 13). No significant difference was found between sides for the latter half of the year. Mechanoregulation of the contralateral site visually appeared stable over the entire year (Figure 4.1.4).

Table 1 Classification of formation, resorption, and quiescence events in patient data for the lowest and highest calculated effective strain values.

	FRACTURE		INTACT	
	Rel. EFF < 1% 99%	Rel. EFF > 99%	Rel. EFF < 1% 99%	Rel. EFF > 99%
FORMATION	19 ± 3% *	54 ± 4% *	25 ± 1%	45 ± 4%
RESORPTION	57 ± 9%	18 ± 5% **	51 ± 3%	29 ± 2%
QUIESCENCE	24 ± 7%	28 ± 3%	23 ± 2%	25 ± 4%

Rel. EFF: Effective strain relative to the reference value 0.0258
 Statistical significance from paired t-test with Holm-Bonferroni correction: * ($p < 0.05$), ** ($p < 0.01$)

4.1.4 Discussion

The purpose of the virtual experiments was to determine if mechanoregulation can be observed with HR-pQCT, using standard clinical settings. Median CCRs of 60% (Figure 4.1.1) were observed with noise levels and voxel sizes comparable to clinical HR-pQCT, showing that mechanoregulation was in fact observable with clinical HR-pQCT image properties. Interestingly, even for zero noise and micro-CT voxel size the CCR does not reach 100%; this can be explained by the difference in using EFF for the analysis and the Gaussian dilated SED for the bone remodelling simulations (Schulte et al., 2013).

In the patient dataset, mechanoregulation was found for both fractured and intact radii (Figure 3). CCR results during fracture healing were around 42-44% (Figure 4.1.4) which is comparable to what has been found in fracture healing animal studies (35-43%) (Tourolle né Betts et al., 2020). The almost random events of high EFF resorption in intact radii (Figure 4.1.3B) are in line with the theory of the remodelling cycle also acting as a maintenance protocol of bone

(Kenkre and Bassett, 2018). Interestingly, this effect was less present in the fracture site. This could indicate that processes specific to fracture repair exhibit more mechanosensitivity to mechanical signals compared to bone remodelling events in intact bone. All seven patients showed similar mechanoregulation curves, which is probably related to the fact that their BV/TV values were all around 17% (Figure 4.1.2), hence a broader range in BV/TV should be explored to e.g. also include osteoporotic patients (BV/TV <10%) (Krause et al., 2014).

4.1.5 Limitations

This study is not without limitations. A limitation was the use of just one patient for the virtual experiments. However, using more samples would have been beyond our computing budget. Furthermore, since every simulation provided 25 iterations of data and three different adaptation scenarios were simulated, the total number of images used for validation was actually 75.

Another limitation was the number of samples (N=7) used, which is low compared to other HR-pQCT studies. However, for the only other study looking at mechanoregulation in humans without fracture, sample numbers are not much higher (N=9) (Christen et al., 2014). Furthermore, fracture data has the added difficulty of many samples showing fracture fragment movements, which still cannot be accurately registered over the entire bone volume (De Jong et al., 2017), further reducing the samples that can be used for such an analysis.

Another limitation was the use of a simple axial compression instead of a more complex combination of loading scenarios based on physiological load estimation algorithms (Christen et al., 2013). However, the axial compression is still the most used FEA boundary condition for HR-pQCT image data (Ohs et al., 2020a). Furthermore, the existing mechanoregulative algorithm works with SED and not with EFF. However, greyscale-based FEA yields unrealistic surface SED values due to partial volume effects. On the other hand, not using greyscale FEA results in poor local mechanical accuracy (chapter 3.3). Hence, EFF was used in the mechanoregulation analysis to circumvent issues with any surface SED values while still benefiting from the results of the greyscale-based micro-FEA.

Finally, a limitation of this study is the use of automatically generated masks of the trabecular compartment with a very conservative approach, such that trabeculae close to the cortex are not

included into the analysis. Although considered the gold-standard, hand-drawn contours are known to show operator-bias (chapter 3.4). Furthermore, this effect is magnified by the complicated and often unclear geometries found in HR-pQCT images of fractures and fracture calluses, which would result in a further reduced reproducibility of the study.

4.1.6 Conclusion

In conclusion, HR-pQCT images can be used to study mechanoregulation in patients. Furthermore, for the first time, mechanoregulation of local bone remodelling was observed in fractured distal radii *in vivo*, making this a viable option to study fracture healing locally on the microscale. Differences in mechanoregulation were found between the fractured and contralateral site, showing the need for further studies of the mechanosensitivity of fracture-repair-specific biological processes. We hope that future fracture studies will include mechanoregulation analyses to better understand fracture healing ultimately leading to improved treatment protocols and functional patient outcomes.

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.

Conflict of Interest

The authors declare that they have no competing interests.

References

- Augat, P., Merk, J., Ignatius, A., Margevicius, K., Bauer, G., Rosenbaum, D., Claes, L., 1996. Early, Full Weightbearing With Flexible Fixation Delays Fracture Healing. *Clin. Orthop. Relat. Res.* 328, 194–202. <https://doi.org/10.1097/00003086-199607000-00031>
- Augat, P., Simon, U., Liedert, A., Claes, L., 2005. Mechanics and mechano-biology of fracture healing in normal and osteoporotic bone. *Osteoporos. Int.* 16, 36–43. <https://doi.org/10.1007/s00198-004-1728-9>
- Badilatti, S.D., Christen, P., Parkinson, I., Müller, R., 2016. Load-adaptive bone remodeling simulations reveal osteoporotic microstructural and mechanical changes in whole human vertebrae. *J. Biomech.* 49, 3770–3779. <https://doi.org/10.1016/j.jbiomech.2016.10.002>
- Boerckel, J.D., Kolambkar, Y.M., Stevens, H.Y., Lin, A.S.P., Dupont, K.M., Guldberg, R.E., 2012. Effects of in vivo mechanical loading on large bone defect regeneration. *J. Orthop. Res.* 30, 1067–1075. <https://doi.org/10.1002/jor.22042>
- Burghardt, A.J., Kazakia, G.J., Sode, M., De Papp, A.E., Link, T.M., Majumdar, S., 2010. A longitudinal HR-pQCT study of alendronate treatment in postmenopausal women with low bone density: Relations among density, cortical and trabecular microarchitecture, biomechanics, and bone turnover. *J. Bone Miner. Res.* 25, 2558–2571. <https://doi.org/10.1002/jbmr.157>
- Burt, L.A., Hanley, D.A., Boyd, S.K., 2017. Cross-sectional Versus Longitudinal Change in a Prospective HR-pQCT Study. *J. Bone Miner. Res.* 32, 1505–1513. <https://doi.org/10.1002/jbmr.3129>
- Carter, D.R., 1987. Mechanical loading history and skeletal biology. *J. Biomech.* 20, 1095–1109. [https://doi.org/10.1016/0021-9290\(87\)90027-3](https://doi.org/10.1016/0021-9290(87)90027-3)
- Carter, D.R., Beaupré, G.S., Giori, N.J., Helms, J.A., 1998. Mechanobiology of Skeletal Regeneration. *Clin. Orthop. Relat. Res.* 355S, S41–S55. <https://doi.org/10.1097/00003086-199810001-00006>
- Carter, D.R., Fyhrie, D.P., Whalen, R.T., 1987. Trabecular bone density and loading history: Regulation of connective tissue biology by mechanical energy. *J. Biomech.* 20,

[https://doi.org/10.1016/0021-9290\(87\)90058-3](https://doi.org/10.1016/0021-9290(87)90058-3)

- Christen, P., Ito, K., Ellouz, R., Boutroy, S., Sornay-Rendu, E., Chapurlat, R.D., Van Rietbergen, B., 2014. Bone remodelling in humans is load-driven but not lazy. *Nat. Commun.* 5. <https://doi.org/10.1038/ncomms5855>
- Christen, P., Ito, K., Knippels, I., Müller, R., van Lenthe, G.H., van Rietbergen, B., 2013. Subject-specific bone loading estimation in the human distal radius. *J. Biomech.* 46, 759–766. <https://doi.org/10.1016/j.jbiomech.2012.11.016>
- Claes, L.E., Heigele, C.A., Neidlinger-Wilke, C., Kaspar, D., Seidl, W., Margevicius, K.J., Augat, P., 1998. Effects of Mechanical Factors on the Fracture Healing Process. *Clin. Orthop. Relat. Res.* 355S, S132–S147. <https://doi.org/10.1097/00003086-199810001-00015>
- Court-Brown, C.M., Caesar, B., 2006. Epidemiology of adult fractures: A review. *Injury* 37, 691–697. <https://doi.org/10.1016/j.injury.2006.04.130>
- De Jong, J.J.A., Christen, P., Plett, R.M., Chapurlat, R., Geusens, P.P., Van Den Bergh, J.P.W., Müller, R., Van Rietbergen, B., 2017. Feasibility of rigid 3D image registration of high-resolution peripheral quantitative computed tomography images of healing distal radius fractures. *PLoS One* 12. <https://doi.org/10.1371/journal.pone.0179413>
- De Jong, J.J.A., Willems, P.C., Arts, J.J., Bours, S.G.P., Brink, P.R.G., van Geel, T.A.C.M., Poeze, M., Geusens, P.P., van Rietbergen, B., van den Bergh, J.P.W., 2014. Assessment of the healing process in distal radius fractures by high resolution peripheral quantitative computed tomography. *Bone* 64, 65–74. <https://doi.org/10.1016/j.bone.2014.03.043>
- Ellouz, R., Chapurlat, R., van Rietbergen, B., Christen, P., Pialat, J.B., Boutroy, S., 2014. Challenges in longitudinal measurements with HR-pQCT: Evaluation of a 3D registration method to improve bone microarchitecture and strength measurement reproducibility. *Bone* 63, 147–157. <https://doi.org/10.1016/j.bone.2014.03.001>
- Flaig, C., 2012. A highly scalable memory efficient multigrid solver for μ -finite element analyses. ETH Zurich. <https://doi.org/10.3929/ethz-a-007613965>
- Giannoudis, P. V., Einhorn, T.A., Marsh, D., 2007. Fracture healing: The diamond concept. *Injury* 38, S3–S6. [https://doi.org/10.1016/S0020-1383\(08\)70003-2](https://doi.org/10.1016/S0020-1383(08)70003-2)

- Huiskes, R., Rulmerman, R., Van Lenthe, G.H., Janssen, J.D., 2000. Effects of mechanical forces on maintenance and adaptation of form in trabecular bone. *Nature* 405, 704–706. <https://doi.org/10.1038/35015116>
- Isaksson, H., Wilson, W., van Donkelaar, C.C., Huiskes, R., Ito, K., 2006. Comparison of biophysical stimuli for mechano-regulation of tissue differentiation during fracture healing. *J. Biomech.* 39, 1507–1516. <https://doi.org/10.1016/j.jbiomech.2005.01.037>
- Johnell, O., Kanis, J.A., 2006. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos. Int.* 17, 1726–1733. <https://doi.org/10.1007/s00198-006-0172-4>
- Kenkre, J.S., Bassett, J.H.D., 2018. The bone remodelling cycle. *Ann. Clin. Biochem.* 55, 308–327. <https://doi.org/10.1177/0004563218759371>
- Klein, P., Schell, H., Streitparth, F., Heller, M., Kassi, J.P., Kandziora, F., Bragulla, H., Haas, N.P., Duda, G.N., 2003. The initial phase of fracture healing is specifically sensitive to mechanical conditions. *J. Orthop. Res.* 21, 662–669. [https://doi.org/10.1016/S0736-0266\(02\)00259-0](https://doi.org/10.1016/S0736-0266(02)00259-0)
- Krause, M., Museyko, O., Breer, S., Wulff, B., Duckstein, C., Vettorazzi, E., Glueer, C., Püschel, K., Engelke, K., Amling, M., 2014. Accuracy of trabecular structure by HR-pQCT compared to gold standard μ CT in the radius and tibia of patients with osteoporosis and long-term bisphosphonate therapy. *Osteoporos. Int.* 25, 1595–1606. <https://doi.org/10.1007/s00198-014-2650-4>
- Mulder, L., Koolstra, J.H., den Toonder, J.M.J., van Eijden, T.M.G.J., 2007. Intratrabecular distribution of tissue stiffness and mineralization in developing trabecular bone. *Bone* 41, 256–265. <https://doi.org/10.1016/j.bone.2007.04.188>
- Mullender, M.G., Huiskes, R., 1995. Proposal for the regulatory mechanism of Wolff's law. *J. Orthop. Res.* 13, 503–512. <https://doi.org/10.1002/jor.1100130405>
- Ng, C.Y., McQueen, M.M., 2011. What are the radiological predictors of functional outcome following fractures of the distal radius? *J. Bone Jt. Surg. - Ser. B* 93 B, 145–150. <https://doi.org/10.1302/0301-620X.93B2.25631>
- Nishiyama, K.K., Pauchard, Y., Nikkel, L.E., Iyer, S., Zhang, C., McMahon, D.J., Cohen, D.,

- Boyd, S.K., Shane, E., Nickolas, T.L., 2015. Longitudinal HR-pQCT and image registration detects endocortical bone loss in kidney transplantation patients. *J. Bone Miner. Res.* 30, 456–463. <https://doi.org/10.1002/jbmr.2358>
- Ohs, N., Collins, C.J., Atkins, P.R., 2020a. Validation of HR-pQCT against Micro-CT for Morphometric and Biomechanical Analyses: A Review. *Bone Reports* 100711. <https://doi.org/10.1016/j.bonr.2020.100711>
- Ohs, N., Tourolle né Betts, D.C., Atkins, P.R., Sebastian, S., van Rietbergen, B., Blauth, M., Christen, P., Müller, R., 2020b. Upscaling of In Vivo HR-pQCT Images Enables Accurate Simulations of Human Microstructural Bone Adaptation. *bioRxiv*. <https://doi.org/10.1101/2020.05.13.093328>
- Pialat, J.B., Burghardt, A.J., Sode, M., Link, T.M., Majumdar, S., 2012. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. *Bone* 50, 111–118. <https://doi.org/10.1016/j.bone.2011.10.003>
- Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2011. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. *Bone* 48, 433–442. <https://doi.org/10.1016/j.bone.2010.10.007>
- Schulte, F.A., Zwahlen, A., Lambers, F.M., Kuhn, G., Ruffoni, D., Betts, D., Webster, D.J., 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. <https://doi.org/10.1016/j.bone.2012.09.008>
- Shanbhogue, V. V., Støvning, R.K., Frederiksen, K.H., Hanson, S., Brixen, K., Gram, J., Jørgensen, N.R., Hansen, S., 2017. Bone structural changes after gastric bypass surgery evaluated by HR-pQCT: A two-year longitudinal study. *Eur. J. Endocrinol.* 176, 685–693. <https://doi.org/10.1530/EJE-17-0014>
- Thévenaz, P., Ruttimann, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41. <https://doi.org/10.1109/83.650848>

Tourolle né Betts, D.C., Wehrle, E., Paul, G.R., Kuhn, G.A., Christen, P., Hofmann, S., Müller, R., 2020. The association between mineralised tissue formation and the mechanical local in vivo environment: Time-lapsed quantification of a mouse defect healing model. *Sci. Rep.* 10, 1–10. <https://doi.org/10.1038/s41598-020-57461-5>

Chapter 5

Synthesis

5.1 Background

Modern technology and medicine has allowed humans to grow older than ever before (United Nations, 2020). However, with the blessing of longevity come the challenges of age related health issues. Bone fractures for example are increasingly common and have become a major socioeconomic burden (Johnell and Kanis, 2006), with the radius fracture being the most common fracture type and also indicative of a generally reduced bone quality (Court-Brown and Caesar, 2006; Mallmin and Ljunghall, 1994). It is known that healthy bone is able to structurally support the human body (Rho et al., 1998), to maintain itself and adapt to changes (Hill, 1998), and to even fully repair itself without scar tissue (Claes et al., 2012). These processes are choreographed across multiple scales in a hierarchical fashion, from organ scale mechanical loads down to protein expressions in cells signalling the requirement for additional deposition of mineral (Gardner et al., 2006; Robling and Turner, 2009). For all the biological complexity of bone, it has been shown that mechanical loading is the key driver behind all these processes (Augat et al., 2005, 1996; Boerckel et al., 2012; Carter, 1987; Carter et al., 1998; Claes et al., 1998; Giannoudis et al., 2007; Huiskes et al., 2000; Klein et al., 2003). This in turn means that being able to monitor the response of bone to mechanical loading could give insight into all this underlying physiological complexity. Given the current lack of consensus regarding fracture treatment protocols, a deeper understanding of the role mechanics play in fracture healing could ultimately also lead to better protocols and hence to better functional outcomes for patients.

Micro-CT has been established as a common tool to study bone remodelling and fracture healing due to its ability to resolve bone at the microstructural level (Bouxsein et al., 2010; Morgan et al., 2009; Schulte et al., 2011; Tourolle né Betts et al., 2020; Wehrle et al., 2019). To analyse these micro-CT images, computational tools have been developed to quantify changes in bone morphology (Goulet et al., 1994; Hildebrand et al., 1999; Hildebrand and Rügsegger, 1997; Odgaard and Gundersen, 1993; Rügsegger et al., 1996; Wachter et al., 2001; Whitehouse, 1974), to compute local mechanical stresses and strains (Chen et al., 2017; Hambli, 2013; Rügsegger et al., 1996; van Rietbergen et al., 1995), to estimate physiological loading (Christen et al., 2012) and to quantify mechanoregulation, i.e. how much of the observed difference in bone structure can be attributed to local mechanical environments (Schulte et al., 2011; Tourolle né Betts et al., 2020). However, micro-CT only allows for longitudinal animal studies or cadaver studies. Since not all processes in animals and humans

are identical, e.g. age related bone loss known to appear in humans couldn't be observed in mouse models (Lambers et al., 2015), the development of high-resolution peripheral quantitative computed tomography (HR-pQCT) allows for the monitoring of distal radii and tibia in patients longitudinally.

With the introduction of HR-pQCT, different computational approaches compared to those for micro-CT were used for many of the morphological parameters, to cope with the lower resolution of HR-pQCT (82 μm) devices compared to micro-CT (1-25 μm) (Ohs et al., 2020). For the assessment of mechanoregulation, a similar reason was given to develop an independent approach (Christen et al., 2014). The obvious drawback is limited comparability with existing animal studies. Validation studies showed only moderate agreement for a lot of the methods between animal studies and patient studies (Ohs et al., 2020). However, some studies have tried to use identical methods showing promising accuracy (Christen et al., 2016, 2013, 2012; MacNeil and Boyd, 2007). With the introduction of second generation HR-pQCT (61 μm), studies have started to utilize established animal study methods, given the increased resolution of second generation HR-pQCT. Nevertheless, so far, no one has addressed the quantification of mechanoregulation for this device generation.

There are additional challenges when investigating the mechanoregulation of fracture. Soft tissue in general cannot be identified in micro-CT or HR-pQCT images. For animal studies, novel algorithms were developed to identify the fracture region in a defect model (Tourolle né Betts et al., 2020). For HR-pQCT, segmentation algorithms have been developed for intact bones to reduce inter-operator variability of manual hand-drawn approaches (Buie et al., 2007; Burghardt et al., 2010). However, even smaller cortical interruptions can lead to broken segmentations with these approaches (Ang et al., 2020; Buie et al., 2007; Soltan et al., 2019; Zebaze et al., 2013). Lately, active contours have been proposed to be used to segment HR-pQCT images as they much better deal with cortical gaps (Caselles et al., 1997; Hafri et al., 2016a, 2016b; Kass et al., 1988; Marquez-Neila et al., 2014), but these methods have not been adapted to work with HR-pQCT images of fractures. There is a need for an algorithm which can consistently contour both the healthy and fractured distal radius.

Analysing only the mechanoregulation in bone can have the limitation that many treatments and therapies target the molecular level. Linking observed differences in mechanoregulation or morphology and the underlying biological complexity has proven to be difficult. In mice, two major approaches have been pursued. One is a combination of micro-CT images with

immunohistochemistry to spatially correlate changes in microstructure and local mechanical environments with cell locations (Scheuren, 2020; Trüssel, 2015). This approach, while powerful, is not applicable for clinical studies. The other approach is the use of computer-models. A multi-physics model has been recently proposed that models the key molecular interactions in bone (Tourolle, 2019). By tuning parameters to fit observed mechanoregulation profiles, chemical concentrations and cell counts can be inferred. However, this model is still in its early development phase. A simpler mechanostat-based (Frost, 1987) approach (Schulte et al., 2013) has already been shown to realistically simulate osteoporotic bone structures on cadaveric micro-CT images (Badilatti et al., 2016). While not offering the same level of detail as the multi-physics model, separate mechanostat parameters responsible for formation or resorption can still narrow down the list of potential pathways to investigate. However, while this *in silico* model has been used successfully on micro-CT images, in contrast to morphological analysis tools, it is not clear how the difference in resolution between HR-pQCT and micro-CT affects simulation accuracy.

Therefore, the goal of this thesis was to 1) provide a computational framework to study fracture healing and 2) investigate the mechanoregulation of fracture healing in a patient cohort. To achieve these aims, a Python framework was developed that encompasses existing validated methods as well as providing a basis for future computational research in the bone field. To demonstrate the capabilities of the framework, a proof-of-concept bone adaptation algorithm including biological markers was developed studying the effect of mechanical disuse on bone. Furthermore, within this framework, the established mechanostat-based bone adaptation simulation by Schulte et al was implemented. An upscaling pre-processing step was studied to address issues of resolution dependent convergence. Moreover, a regularization pre-processing step was developed to deal with initialization shocks of this model. This implementation was used to generate fully mechanically regulated time series of bone adaptation which were used to evaluate the ability of the established mechanoregulation quantification algorithm to work on images with resolution and noise levels identical to those produced by HR-pQCT. To address segmentation of HR-pQCT fracture data, a novel active-contour based approach was developed. The evaluated mechanoregulation analysis and the contouring approach were then used to study mechanoregulation in fracture patients, comparing their fractured to their contralateral arm.

5.2 Main findings and implications

The studies of this thesis have shown that second generation HR-pQCT images contain enough information to apply micro-CT methodologies to them, from bone adaptation simulations to mechanoregulation analyses in fracture data, all within one computational framework. Firstly, a novel Boolean network based bone remodelling algorithm was introduced to demonstrate the ease with which novel technologies can be incorporated into the framework. Simulating physical disuse, the model predicted bone loss in agreement with what was found in astronauts (Shackelford et al., 2004).

The next achievement was the realisation of accurate bone adaptation simulations on HR-pQCT. Based on the work by Schulte and colleagues (2013) and motivated by the success of Badilatti and colleagues (Badilatti et al., 2016) to simulate osteoporotic structures on *ex vivo* human micro-CT scans, the bone adaptation algorithm was incorporated into the framework. Initialisation shocks, i.e. drastic changes in parameters during the first few iterations, were observed in alignment with previous studies (Badilatti et al., 2016; Schulte et al., 2013). A regularization pre-processing step was developed in which information from the grayscale image rather than from the binarized image was used for model initialisation. This reduced occurring initialisation shocks allowing for direct comparisons of simulation outcomes to clinically observed changes in bone microstructure. Furthermore, it was established that upscaling images with HR-pQCT resolution prior to running simulations led to accurate results in terms of morphometrics compared to simulations run on micro-CT images. A clear drop in accuracy was found without this additional upscaling showing a clear dependence on resolution as was also observed for other micro-CT methods (Christen et al., 2016; Müller et al., 1996). The drop in accuracy is in line with other studies that showed that no single threshold can be applied to HR-pQCT images to obtain both, accurate morphometric and accurate mechanics (Alsayednoor et al., 2018), the two key components of bone adaptation simulations. Thus, the groundwork for future simulation work has been laid with the findings suggesting that the information contained within images of HR-pQCT resolution is sufficient to recreate morphologically and mechanically accurate representations *in silico*. In particular, this questions the one to one correspondence between numerical mesh and image voxels primarily used with HR-pQCT images (Ohs et al., 2020).

The second achievement was the development of an accurate automated contouring approach for HR-pQCT images of distal radii. Observations of the current gold standard approach developed for intact radii failing to cope with larger cortical interruptions (Buie et al., 2007) and observations of inter-observer variability for hand-drawn contours motivated the development of an approach to use active contours. The developed approach was found to accurately contour images of fractured distal radii throughout the entire healing phase compared to hand-drawn contours. While no study has looked at fracture data so far, this result is in line with applications of active contours to ultra-distal scans in which cortical interruptions are also more prominent (Hafri et al., 2016b). Interestingly, the gold standard approach developed for intact radii was only successful on images one year post fracture, supporting evidence found that fracture healing could be a longer process than previously anticipated (de Jong et al., 2016). Furthermore, the gold standard approach was also found to struggle with contouring the contralateral images. This is in line with findings that radius fractures can be indicative of poor bone quality in general (Mallmin and Ljunghall, 1994).

Building on work performed by Schulte and Colleagues (Schulte et al., 2013, 2011) on bone adaptation simulations and quantification of mechanoregulation, bone adaptation simulations were used to demonstrate that mechanoregulation can be obtained from images with HR-pQCT resolution and noise levels. Such a form of validation is similar to the bone resorption based one by Salmon and colleagues (Salmon et al., 2015). Mechanoregulation was detected for bone remodelling (Christen et al., 2014) and fracture repair with levels similar to those found in mouse studies (Tourolle né Betts et al., 2020). This supports evidence that the healing process in distal radius fractures can be assessed by HR-pQCT (De Jong et al., 2014). Furthermore, an increased mechanoregulation was found during the early stage of fracture healing. This is in contrast to animal studies, however they studied a large size defect, which differs significantly from a conservatively treated radius fracture (Tourolle né Betts et al., 2020). Interestingly, the late reparative phase shows similar levels of mechanoregulation which would support the hypothesis that fracture healing and bone adaptation share similar mechano-regulatory processes (Huiskes et al., 2000; Isaksson et al., 2009).

5.3 Limitations and future research

A limitation of the model in chapter 3.2 is its inability to capture different concentrations for different molecular markers. Markers are either present in a voxel or not which does not

represent physiological conditions. However, the main purpose of this model was to demonstrate the capabilities of the implemented framework to work with bone data. This already has spawned several new projects utilizing the framework (Scheuren, 2020; Tourolle, 2019). The model by Tourolle and colleagues is a logical extension of the Boolean network approach in which molecular markers are stored as concentrations and molecular interactions happen via diffusion-reaction equations instead of Boolean logic. Such a model presents new research potential to simulate the effects of medication on morphological and mechanical properties of bone microstructure.

A limitation of the validation study in chapter 3.3 is the lack of experimental data. Accuracy was only determined from downscaling micro-CT images with respect to micro-CT. No HR-pQCT images of the same bone samples were taken. However, downscaling has already been used in several other studies to simulate HR-pQCT images (Christen et al., 2016; Liu et al., 2011; Manske et al., 2015). Furthermore, simulations run on micro-CT images have already been shown to yield realistic results (Badilatti et al., 2016), providing a reasonable ground truth. Furthermore, while HR-pQCT images can in principle be obtained from cadavers, they do not incorporate other imaging artefacts typically associated with HR-pQCT, such as motion artefacts. On the other hand, using HR-pQCT images of patients lacks the necessary micro-CT ground truth. Therefore, future effort should be put into capturing changes observed in HR-pQCT time-lapse images. These efforts, however, do not necessarily require the model used in chapter 3.3 (Schulte et al., 2013). Since one of the observations of chapter 3.3 was that HR-pQCT obtains enough information to generate a morphologically and mechanically accurate representation of a scanned bone *in silico*, more advanced models could also be tested, such as the model by Tourolle (Tourolle, 2019).

The major limitation of chapter 3.4, the development of an automatic fracture contouring algorithm for HR-pQCT images of distal radius fractures, is the variability found in the gold standard hand-drawn contours used for validation. While the approach was accurate within that variability, operators themselves did not always agree on the same contour lines. Especially fracture gaps in the cortex were sometimes identified as part of the fractured radius, and sometime as background. However, it is unclear, what the effect of this variability is on parameters of interest such as morphological indices or mechanical properties. Future studies should therefore investigate these effects and develop clear guidelines on how to draw these contours by hand for images of fractured radii based on these results. This way, automatic

contours can be combined with manual corrections such as is already the case for HR-pQCT of intact radii (Buie et al., 2007; Whittier et al., 2020) creating the best of both worlds.

With respect to chapter 4.1, the study of mechanoregulation has very strict selection criteria for HR-pQCT images that can be used for analysis. Several pre-conditions limited the number of patients that could be included in chapter 4.1. First of all, some patients did not complete the study limiting the database to 75 patients. Secondly, poor image quality, mainly on the contralateral arm, prohibited accurate image registration and would also prohibit accurate identification of sites of formation and resorption, which are required for the analysis of mechanoregulation. Thirdly, fracture fragment movement could not be compensated for by the commonly used rigid body registration. This led to poor registration results for some regions resulting again in incorrect detection of formation and resorption. Future studies can improve upon this by enhancing retention rates for studies can be challenging but strategies have been developed to ensure that most patients finish the study (Norvell et al., 2016). To compensate for poor image quality, future studies should account for re-scans (Cheung et al., 2013) in their study protocol in case poor image quality is detected. Since the image quality of the contralateral site was poor most of the times, patients should be clearly informed about the importance of the contralateral scan. Fracture fragment registration has already been attempted (De Jong et al., 2017) but with the currently proposed method, much of the scanned volume has to be discarded. Further studies should be conducted to improve the usable volume of such registered images, as other artefacts such as stack gaps (another form of movement artefact) already limit the usable volume in scanned HR-pQCT multi stack fracture images.

5.4 Conclusion

In conclusion, the development of a computational framework and its application to study fracture healing on longitudinal HR-pQCT images of fracture patients was successful. We demonstrated how and that HR-pQCT images can be used to run accurate bone adaptation simulations. We showed how upscaling and regularisation leads to better *in silico* representations of bone structures scanned with HR-pQCT with respect to morphology and mechanics. We proposed a novel approach to automatically contour HR-pQCT images of distal radius fractures, avoiding time intensive operator-biased manual contouring. Application of this approach to HR-pQCT images of intact distal radii showed that the novel approach is more robust than the current gold standard with respect to ultra-distal scans and poor bone quality.

Combining these technologies, we demonstrated how bone adaptation simulations can be a vital tool in method validation, by demonstrating that mechanoregulation can be obtained from images with HR-pQCT resolution and noise levels. Analysing a patient cohort of fracture patients, we confirmed that human bone remodelling is load driven and we showed for the first time *in vivo* that human fracture healing is mechanically regulated on the microstructural level.

The developed framework and the incorporated tools allow for comprehensive investigations of mechanoregulation in fracture patients. The successful utilization of HR-pQCT images for bone adaptation simulations indicate novel opportunities to study pharmaceutical interventions in humans *in vivo*. The successful application of animal-study tools used in the presented studies indicate that future HR-pQCT studies can employ such tools allowing for a stronger feedback cycle between labs and the clinics, helping to fully utilize both settings to unravel the mechanism governing fracture healing.

References

- 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. <https://doi.org/10.1007/s10237-018-1051-6>
- Ang, I.C., Fox, M., Polk, J.D., Kersh, M.E., 2020. An Algorithm for Automated Separation of Trabecular Bone from Variably Thick Cortices in High-Resolution Computed Tomography Data. *IEEE Trans. Biomed. Eng.* 1–1. <https://doi.org/10.1109/TBME.2019.2924398>
- Augat, P., Merk, J., Ignatius, A., Margevicius, K., Bauer, G., Rosenbaum, D., Claes, L., 1996. Early, Full Weightbearing With Flexible Fixation Delays Fracture Healing. *Clin. Orthop. Relat. Res.* 328, 194–202. <https://doi.org/10.1097/00003086-199607000-00031>
- Augat, P., Simon, U., Liedert, A., Claes, L., 2005. Mechanics and mechano-biology of fracture healing in normal and osteoporotic bone. *Osteoporos. Int.* 16, 36–43. <https://doi.org/10.1007/s00198-004-1728-9>
- Badilatti, S.D., Christen, P., Parkinson, I., Müller, R., 2016. Load-adaptive bone remodeling simulations reveal osteoporotic microstructural and mechanical changes in whole human vertebrae. *J. Biomech.* 49, 3770–3779. <https://doi.org/10.1016/j.jbiomech.2016.10.002>
- Boerckel, J.D., Kolambkar, Y.M., Stevens, H.Y., Lin, A.S.P., Dupont, K.M., Guldberg, R.E., 2012. Effects of in vivo mechanical loading on large bone defect regeneration. *J. Orthop. Res.* 30, 1067–1075. <https://doi.org/10.1002/jor.22042>
- Bouxsein, M.L., Boyd, S.K., Christiansen, B.A., Guldberg, R.E., Jepsen, K.J., Müller, R., 2010. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. *J. Bone Miner. Res.* 25, 1468–1486. <https://doi.org/10.1002/jbmr.141>
- Buie, H.R., Campbell, G.M., Klinck, R.J., MacNeil, J.A., Boyd, S.K., 2007. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* 41, 505–515. <https://doi.org/10.1016/j.bone.2007.07.007>

- Burghardt, A.J., Buie, H.R., Laib, A., Majumdar, S., Boyd, S.K., 2010. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. *Bone* 47, 519–528. <https://doi.org/10.1016/j.bone.2010.05.034>
- Carter, D.R., 1987. Mechanical loading history and skeletal biology. *J. Biomech.* 20, 1095–1109. [https://doi.org/10.1016/0021-9290\(87\)90027-3](https://doi.org/10.1016/0021-9290(87)90027-3)
- Carter, D.R., Beaupré, G.S., Giori, N.J., Helms, J.A., 1998. Mechanobiology of Skeletal Regeneration. *Clin. Orthop. Relat. Res.* 355S, S41–S55. <https://doi.org/10.1097/00003086-199810001-00006>
- Caselles, V., Kimmel, R., Sapiro, G., 1997. Geodesic Active Contours. *Int. J. Comput. Vis.* 22, 61–79. <https://doi.org/10.1023/A:1007979827043>
- Chen, Y., Dall'Ara, E., Sales, E., Manda, K., Wallace, R., Pankaj, P., Viceconti, M., 2017. Micro-CT based finite element models of cancellous bone predict accurately displacement once the boundary condition is well replicated: A validation study. *J. Mech. Behav. Biomed. Mater.* 65, 644–651. <https://doi.org/10.1016/j.jmbbm.2016.09.014>
- Cheung, A.M., Adachi, J.D., Hanley, D.A., Kendler, D.L., Davison, K.S., Josse, R., Brown, J.P., Ste-Marie, L.G., Kremer, R., Erlandson, M.C., Dian, L., Burghardt, A.J., Boyd, S.K., 2013. High-resolution peripheral quantitative computed tomography for the assessment of bone strength and structure: A review by the canadian bone strength working group. *Curr. Osteoporos. Rep.* 11, 136–146. <https://doi.org/10.1007/s11914-013-0140-9>
- Christen, P., Ito, K., Ellouz, R., Boutroy, S., Sornay-Rendu, E., Chapurlat, R.D., Van Rietbergen, B., 2014. Bone remodelling in humans is load-driven but not lazy. *Nat. Commun.* 5. <https://doi.org/10.1038/ncomms5855>
- Christen, P., Ito, K., Knippels, I., Müller, R., van Lenthe, G.H., van Rietbergen, B., 2013. Subject-specific bone loading estimation in the human distal radius. *J. Biomech.* 46, 759–766. <https://doi.org/10.1016/j.jbiomech.2012.11.016>
- Christen, P., Schulte, F.A., Zwahlen, A., van Rietbergen, B., Boutroy, S., Melton, L.J., Amin, S., Khosla, S., Goldhahn, J., Müller, R., 2016. Voxel size dependency, reproducibility

- and sensitivity of an in vivo bone loading estimation algorithm. *J. R. Soc. Interface* 13, 20150991. <https://doi.org/10.1098/rsif.2015.0991>
- Christen, P., Van Rietbergen, B., Lambers, F.M., Müller, R., Ito, K., 2012. Bone morphology allows estimation of loading history in a murine model of bone adaptation. *Biomech. Model. Mechanobiol.* 11, 483–492. <https://doi.org/10.1007/s10237-011-0327-x>
- Claes, L., Recknagel, S., Ignatius, A., 2012. Fracture healing under healthy and inflammatory conditions. *Nat. Rev. Rheumatol.* 8, 133–143. <https://doi.org/10.1038/nrrheum.2012.1>
- Claes, L.E., Heigele, C.A., Neidlinger-Wilke, C., Kaspar, D., Seidl, W., Margevicius, K.J., Augat, P., 1998. Effects of Mechanical Factors on the Fracture Healing Process. *Clin. Orthop. Relat. Res.* 355S, S132–S147. <https://doi.org/10.1097/00003086-199810001-00015>
- Court-Brown, C.M., Caesar, B., 2006. Epidemiology of adult fractures: A review. *Injury* 37, 691–697. <https://doi.org/10.1016/j.injury.2006.04.130>
- de Jong, J.J., Heyer, F.L., Arts, J.J.C., Poeze, M., Keszei, A.P., Willems, P.C., van Rietbergen, B., Geusens, P.P., van den Bergh, J.P., 2016. Fracture Repair in the Distal Radius in Postmenopausal Women: A Follow-Up 2 Years Postfracture Using HRpQCT. *J. Bone Miner. Res.* 31, 1114–1122. <https://doi.org/10.1002/jbmr.2766>
- De Jong, J.J.A., Christen, P., Plett, R.M., Chapurlat, R., Geusens, P.P., Van Den Bergh, J.P.W., Müller, R., Van Rietbergen, B., 2017. Feasibility of rigid 3D image registration of high-resolution peripheral quantitative computed tomography images of healing distal radius fractures. *PLoS One* 12. <https://doi.org/10.1371/journal.pone.0179413>
- De Jong, J.J.A., Willems, P.C., Arts, J.J., Bours, S.G.P., Brink, P.R.G., van Geel, T.A.C.M., Poeze, M., Geusens, P.P., van Rietbergen, B., van den Bergh, J.P.W., 2014. Assessment of the healing process in distal radius fractures by high resolution peripheral quantitative computed tomography. *Bone* 64, 65–74. <https://doi.org/10.1016/j.bone.2014.03.043>
- Frost, H.M., 1987. The mechanostat: a proposed pathogenic mechanism of osteoporoses and the bone mass effects of mechanical and nonmechanical agents. *Bone Miner.* 2, 73–85.

- Gardner, M.J., van der Meulen, M.C.H., Demetrakopoulos, D., Wright, T.M., Myers, E.R., Bostrom, M.P., 2006. In vivo cyclic axial compression affects bone healing in the mouse tibia. *J. Orthop. Res.* 24, 1679–1686. <https://doi.org/10.1002/jor.20230>
- Giannoudis, P. V., Einhorn, T.A., Marsh, D., 2007. Fracture healing: The diamond concept. *Injury* 38, S3–S6. [https://doi.org/10.1016/S0020-1383\(08\)70003-2](https://doi.org/10.1016/S0020-1383(08)70003-2)
- Goulet, R.W., Goldstein, S.A., Ciarelli, M.J., Kuhn, J.L., Brown, M.B., Feldkamp, L.A., 1994. The relationship between the structural and orthogonal compressive properties of trabecular bone. *J. Biomech.* 27. [https://doi.org/10.1016/0021-9290\(94\)90014-0](https://doi.org/10.1016/0021-9290(94)90014-0)
- Hafri, M., Jennane, R., Lespessailles, E., Toumi, H., 2016a. Dual active contours model for HR-pQCT cortical bone segmentation, in: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, pp. 2270–2275. <https://doi.org/10.1109/ICPR.2016.7899974>
- Hafri, M., Toumi, H., Boutroy, S., Chapurlat, R.D., Lespessailles, E., Jennane, R., 2016b. Fuzzy energy based active contours model for HR-PQCT cortical bone segmentation, in: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 4334–4338. <https://doi.org/10.1109/ICIP.2016.7533178>
- Hambli, R., 2013. Micro-CT finite element model and experimental validation of trabecular bone damage and fracture. *Bone* 56, 363–374. <https://doi.org/10.1016/j.bone.2013.06.028>
- Hildebrand, T., Laib, A., Müller, R., Dequeker, J., Rüegsegger, P., 1999. Direct Three-Dimensional Morphometric Analysis of Human Cancellous Bone: Microstructural Data from Spine, Femur, Iliac Crest, and Calcaneus. *J. Bone Miner. Res.* 14, 1167–1174. <https://doi.org/10.1359/jbmr.1999.14.7.1167>
- Hildebrand, T., Rüegsegger, P., 1997. A new method for the model-independent assessment of thickness in three-dimensional images. *J. Microsc.* 185, 67–75. <https://doi.org/10.1046/j.1365-2818.1997.1340694.x>
- Hill, P.A., 1998. Bone remodelling. *Br. J. Orthod.* 25, 101–107. <https://doi.org/10.1093/ortho/25.2.101>

- Huiskes, R., Rulmerman, R., Van Lenthe, G.H., Janssen, J.D., 2000. Effects of mechanical forces on maintenance and adaptation of form in trabecular bone. *Nature* 405, 704–706. <https://doi.org/10.1038/35015116>
- Isaksson, H., Gröngröft, I., Wilson, W., Van Donkelaar, C.C., Ven Rietbergen, B., Tami, A., Huiskes, R., Ito, K., 2009. Remodeling of fracture callus in mice is consistent with mechanical loading and bone remodeling theory. *J. Orthop. Res.* 27, 664–672. <https://doi.org/10.1002/jor.20725>
- Johnell, O., Kanis, J.A., 2006. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos. Int.* 17, 1726–1733. <https://doi.org/10.1007/s00198-006-0172-4>
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *Int. J. Comput. Vis.* 1, 321–331. <https://doi.org/10.1007/BF00133570>
- Klein, P., Schell, H., Streitparth, F., Heller, M., Kassi, J.P., Kandziora, F., Bragulla, H., Haas, N.P., Duda, G.N., 2003. The initial phase of fracture healing is specifically sensitive to mechanical conditions. *J. Orthop. Res.* 21, 662–669. [https://doi.org/10.1016/S0736-0266\(02\)00259-0](https://doi.org/10.1016/S0736-0266(02)00259-0)
- Lambers, F.M., Kuhn, G., Weigt, C., Koch, K.M., Schulte, F.A., Müller, R., 2015. Bone adaptation to cyclic loading in murine caudal vertebrae is maintained with age and directly correlated to the local micromechanical environment. *J. Biomech.* 48, 1179–1187. <https://doi.org/10.1016/j.jbiomech.2014.11.020>
- Liu, X.S., Shane, E., McMahon, D.J., Guo, X.E., 2011. Individual trabecula segmentation (ITS)-based morphological analysis of microscale images of human tibial trabecular bone at limited spatial resolution. *J. Bone Miner. Res.* 26, 2184–2193. <https://doi.org/10.1002/jbmr.420>
- MacNeil, J.A., Boyd, S.K., 2007. Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Med. Eng. Phys.* 29, 1096–1105. <https://doi.org/10.1016/j.medengphy.2006.11.002>
- Mallmin, H., Ljunghall, S., 1994. Distal radius fracture is an early sign of general osteoporosis: Bone mass measurements in a population-based study. *Osteoporos. Int.* 4, 357–361. <https://doi.org/10.1007/BF01622198>

- Manske, S.L., Zhu, Y., Sandino, C., Boyd, S.K., 2015. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone* 79, 213–221. <https://doi.org/10.1016/j.bone.2015.06.006>
- Marquez-Neila, P., Baumela, L., Alvarez, L., 2014. A morphological approach to curvature-based evolution of curves and surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 2–17. <https://doi.org/10.1109/TPAMI.2013.106>
- Morgan, E.F., Mason, Z.D., Chien, K.B., Pfeiffer, A.J., Barnes, G.L., Einhorn, T.A., Gerstenfeld, L.C., 2009. Micro-computed tomography assessment of fracture healing: Relationships among callus structure, composition, and mechanical function. *Bone* 44, 335–344. <https://doi.org/10.1016/j.bone.2008.10.039>
- Müller, 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.
- Norvell, D.C., Dettori, J.R., Chapman, J.R., 2016. Enhancing Clinical Study Retention Rates to Avoid Follow-up Bias: How Do We Keep Our Study Participants from the Land of the Lost? *Glob. Spine J.* 6, 519–521. <https://doi.org/10.1055/s-0036-1584928>
- Odgaard, A., Gundersen, H.J.G., 1993. Quantification of connectivity with special emphasis on 3D reconstructions. *Bone* 14, 173–182.
- Ohs, N., Collins, C.J., Atkins, P.R., 2020. Validation of HR-pQCT against Micro-CT for Morphometric and Biomechanical Analyses: A Review. *Bone Reports* 100711. <https://doi.org/10.1016/j.bonr.2020.100711>
- Rho, J.Y., Kuhn-Spearing, L., Zioupos, P., 1998. Mechanical properties and the hierarchical structure of bone. *Med. Eng. Phys.* 20, 92–102. [https://doi.org/10.1016/S1350-4533\(98\)00007-1](https://doi.org/10.1016/S1350-4533(98)00007-1)
- Robling, A.G., Turner, C.H., 2009. Mechanical signaling for bone modeling and remodeling. *Crit. Rev. Eukaryot. Gene Expr.* 19, 319–338. <https://doi.org/10.1615/CritRevEukarGeneExpr.v19.i4.50>

- Rüegsegger, P., Koller, B., Müller, R., 1996. A Microtomographic System for the Nondestructive Evaluation of Bone Architecture. *Calcif. Tissue Int.* 58, 24–29. <https://doi.org/10.1007/s002239900006>
- Salmon, P.L., Ohlsson, C., Shefelbine, S.J., Doube, M., 2015. Structure model index does not measure rods and plates in trabecular bone. *Front. Endocrinol. (Lausanne)*. 6, 1–10. <https://doi.org/10.3389/fendo.2015.00162>
- Scheuren, A.C., 2020. Longitudinal assessment of frailty and osteosarcopenia in an in vivo model of premature aging.
- Schulte, F.A., Lambers, F.M., Kuhn, G., Müller, R., 2011. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. *Bone* 48, 433–442. <https://doi.org/10.1016/j.bone.2010.10.007>
- Schulte, F.A., Zwahlen, A., Lambers, F.M., Kuhn, G., Ruffoni, D., Betts, D., Webster, D.J., 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. <https://doi.org/10.1016/j.bone.2012.09.008>
- Shackelford, L.C., LeBlanc, A.D., Driscoll, T.B., Evans, H.J., Rianon, N.J., Smith, S.M., Spector, E., Feedback, D.L., Lai, D., 2004. Resistance exercise as a countermeasure to disuse-induced bone loss. *J. Appl. Physiol.* 97, 119–129. <https://doi.org/10.1152/jappphysiol.00741.2003>
- Soltan, N., Kawalilak, C.E., Cooper, D.M., Kontulainen, S.A., Johnston, J.D., 2019. Cortical porosity assessment in the distal radius: A comparison of HR-pQCT measures with Synchrotron-Radiation micro-CT-based measures. *Bone* 120, 439–445. <https://doi.org/10.1016/j.bone.2018.12.008>
- Tourolle, D.C., 2019. A micro-scale multiphysics framework for fracture healing and bone remodelling. <https://doi.org/10.3929/ethz-b-000364637>
- Tourolle né Betts, D.C., Wehrle, E., Paul, G.R., Kuhn, G.A., Christen, P., Hofmann, S., Müller, R., 2020. The association between mineralised tissue formation and the mechanical local in vivo environment: Time-lapsed quantification of a mouse defect healing model. *Sci. Rep.* 10, 1–10. <https://doi.org/10.1038/s41598-020-57461-5>

- Trüssel, A.J., 2015. Spatial mapping and high throughput microfluidic gene expression analysis of osteocytes in mechanically controlled bone remodeling.
- United Nations, 2020. World Population Ageing 2019 (ST/ESA/SER.A/444) [WWW Document]. Dep. Econ. Soc. Aff. Popul. Div. URL www.unpopulation.org
- van Rietbergen, B., Weinans, H., Huiskes, R., Odgaard, A., 1995. A new method to determine trabecular bone elastic properties and loading using micromechanical finite-element models. *J. Biomech.* 28, 69–81. [https://doi.org/10.1016/0021-9290\(95\)80008-5](https://doi.org/10.1016/0021-9290(95)80008-5)
- Wachter, N.J., Augat, P., Krischak, G.D., Mentzel, M., Kinzl, L., Claes, L., 2001. Prediction of cortical bone porosity in vitro by microcomputed tomography. *Calcif. Tissue Int.* 68, 38–42. <https://doi.org/10.1007/s002230001182>
- Wehrle, E., Tourolle né Betts, D.C., Kuhn, G.A., Scheuren, A.C., Hofmann, S., Müller, R., 2019. Evaluation of longitudinal time-lapsed in vivo micro-CT for monitoring fracture healing in mouse femur defect models. *Sci. Rep.* 9, 1–12. <https://doi.org/10.1038/s41598-019-53822-x>
- Whitehouse, W.J., 1974. The quantitative morphology of anisotropic trabecular bone. *J. Microsc.* 101, 153–168. <https://doi.org/10.1111/j.1365-2818.1974.tb03878.x>
- Whittier, D.E., Mudryk, A.N., Vandergaag, I.D., Burt, L.A., Boyd, S.K., 2020. Optimizing HR-pQCT workflow: a comparison of bias and precision error for quantitative bone analysis. *Osteoporos. Int.* 31, 567–576. <https://doi.org/10.1007/s00198-019-05214-0>
- Zebaze, R., Ghasem-Zadeh, A., Mbala, A., Seeman, E., 2013. A new method of segmentation of compact-appearing, Transitional and trabecular compartments and quantification of cortical porosity from high resolution peripheral quantitative computed tomographic images. *Bone* 54, 8–20. <https://doi.org/10.1016/j.bone.2013.01.007>