---

**Article:**
Pataky, Todd, Koskei, Michihiko and Cox, Philip Graham orcid.org/0000-0001-9782-2358 (2016) Probabilistic biomechanical finite element simulations:whole-model classical hypothesis testing based on upcrossing geometry. PeerJ. e96. ISSN 2167-8359

https://doi.org/10.7717/peerj-cs.96

---

# Probabilistic biomechanical finite element simulations: whole-model classical hypothesis testing based on upcrossing geometry

**Todd C. Pataky**[1], **Michihiko Koseki**[2], **and Philip G. Cox**[3,4]

[1]**Institute for Fiber Engineering, Shinshu University, Ueda, Japan**
[2]**Department of Bioengineering, Shinshu University, Ueda, Japan**
[3]**Centre for Anatomical and Human Sciences, Hull York Medical School, University of York, UK**
[4]**Department of Archaeology, University of York, UK**

## ABSTRACT

Statistical analyses of biomechanical finite element (FE) simulations are frequently conducted on scalar metrics extracted from anatomically homologous regions, like maximum von Mises stresses from demarcated bone areas. Advantages of this approach are numerical tabulability and statistical simplicity, but disadvantages include region demarcation subjectivity, spatial resolution reduction, and results interpretation complexity when attempting to mentally map tabulated results to original anatomy. This study proposes a method which abandons the two aforementioned advantages to overcome these three limitations. The method is inspired by parametric random field theory (RFT), but instead uses a non-parametric analogue to RFT which permits flexible model-wide statistical analyses through non-parametrically constructed probability densities regarding volumetric upcrossing geometry. We illustrate method fundamentals using basic 1D and 2D models, then use a public model of hip cartilage compression to highlight how the concepts can extend to practical biomechanical modeling. The ultimate whole-volume results are easy to interpret, and for constant model geometry the method is simple to implement. Moreover, our analyses demonstrate that the method can yield biomechanical insights which are difficult to infer from single simulations or tabulated multi-simulation results. Generalizability to non-constant geometry including subject-specific anatomy is discussed.

Keywords: Biomechanics, Finite element analysis, Computational statistics, Probabilistic simulation, Random field theory

## 1 INTRODUCTION

In numerical finite element (FE) simulations of biomechanical continua model inputs like material properties and load magnitude are often imprecisely known. This uncertainty arises from a variety of sources including: measurement inaccuracy, *in vivo* measurement inaccessibility, and natural between-subject material, anatomical and loading variability (Cheung et al., 2005; Ross et al., 2005; Cox et al., 2011, 2015; Fitton et al., 2012b). Despite this uncertainty, an investigator must choose specific parameter values because numerical simulation requires it. Parameters are typically derived from published data, empirical estimation, or mechanical intuition (Kupczik et al., 2007; Cox et al., 2012, 2013; Rayfield, 2011; Cuff et al., 2015).

It is also possible to perform multiple FE simulations using a spectrum of feasible model input values to generate a distribution of model outputs (Dar et al., 2002; Babuska and Silva, 2014). More simply, probabilistic model inputs yield probabilistic outputs, and continuum mechanics' inherent nonlinearities ensure that these input and output probabilities are nonlinearly related. Probing output distributions statistically therefore generally requires numerical simulation. Such analyses can require substantial computational resources: probabilistic FE outputs have been shown to converge to stable numerical values only for on-the-order of 1000 to 100,000 simulation iterations depending on model complexity (Dopico-

González et al., 2009). The advent of personal computing power has mitigated problems associated with this computational demand and has led to a sharp increase in probabilistic FE simulation in a variety of engineering fields (Stefanou, 2009) including biomechanics (Easley et al., 2007; Laz et al., 2007; Lin et al., 2007; Radcliffe and Taylor, 2007; Fitzpatrick et al., 2012).

Producing a probabilistic input-output mapping is conceptually simple: iteratively change input parameters according to a particular distribution and assemble output parameters for each iteration to yield an output distribution. The simplest method is Monte Carlo simulation which randomly generates input parameters based on given mean and standard deviation values (Dar et al., 2002). More complex methods like Markov Chain Monte Carlo can accelerate probabilistic output distribution convergence (Boyaval, 2012).

Once probabilistic inputs / outputs are generated they may be probed using a variety of statistical methods. A common technique is to extract scalars like maximum von Mises stress from anatomically demarcated regions of interest (Radcliffe and Taylor, 2007). Other techniques include Taguchi global model comparisons (Taguchi, 1987; Dar et al., 2002; Lin et al., 2007) to fuzzy set modeling (Babuska and Silva, 2014) and probability density construction for specific model parameters (Easley et al., 2007; Laz et al., 2007; McFarland and Mahadevan, 2008; Dopico-González et al., 2009).

The purpose of this paper is to propose an alternative method which conducts classical hypothesis testing at the whole-model level using continuum upcrossing geometry. An 'upcrossing' is a portion of the continuum that survives a threshold (Fig.1) like an island above the water's surface or a mountain top above clouds. Each upcrossing possess a number of geometrical features including maximum height, extent and integral, where integrals, for examples, are areas, volumes and hyper-volumes for 1D, 2D and 3D continua, respectively. Parametric solutions to upcrossing geometry probabilities exist for $n$-dimensional Gaussian continua in the random field theory (RFT) literature (Adler and Taylor, 2007), and non-parametric approximations have been shown to be equally effective (Nichols and Holmes, 2002). The method we propose follows the latter, non-parametric permutation approach because it is ideally suited to the iterative simulation which characterizes probabilistic FE analysis.

The method is inspired by hypothesis testing approaches in nonlinear modeling (Legay and Viswanatha, 2009) and in particular a label-based continuum permutation approach (Nichols and Holmes, 2002). It first assembles a large number of element- or node-based test statistic volumes through iterative simulation, then conducts inference using non-parametrically estimated upcrossing probabilities. These upcrossing distributions form a general framework for conducting classical, continuum-level hypothesis testing on FE models in arbitrarily complex experiments.

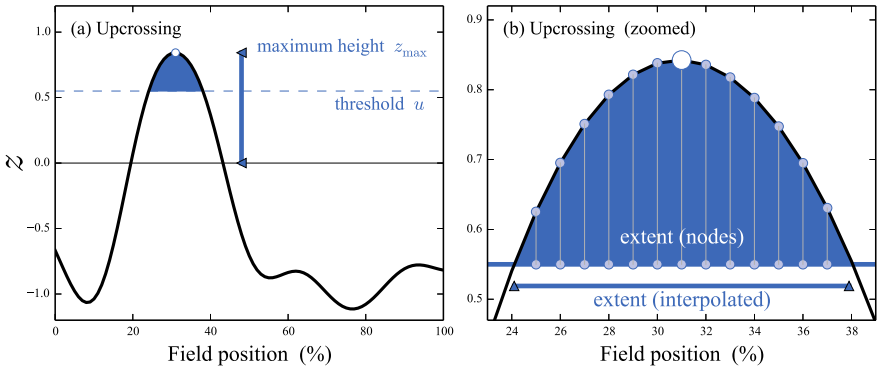


**Figure 1.** Example upcrossing in a 1D continuum. A thresholded continuum contains zero or more upcrossings, each with particular geometric characteristics including: maximum height, extent, integral, etc., each of which is associated with a different probability. The maximum height characteristic — across all upcrossings — can be used to conduct classical hypothesis testing as described in §2.

## 2 METHODS

All analyses were were implemented in FEBio v.2.4.2 and v.2.5.0 (Maas et al., 2012) and Python 2.7 (van Rossum, 2014). All partial differential equations underlying the models' numerical solutions are

described in the FEBio Theory Manual (Maas et al., 2015). Model files and analysis scripts are available in this project's GitHub repository (`github.com/0todd0000/probFEApy`).

## 2.1 Models

### *2.1.1 Model A: Simple anisotropic bone compression*

A single column of hexahedral elements (Fig.2a) with anisotropic stiffness (Fig.2b) was used to represent bone with local material inconsistencies. This simplistic model was used primarily to efficiently demonstrate the key concepts underlying the proposed methodology. Nodal displacements were fully constrained at one end, and a total compressive force of 8000 N was applied to the other end along the longitudinal axis. The bone material was linearly elastic with a Poisson's ratio of 0.3.

Local anisotropy in Young's modulus (Fig.2b) was created using Gaussian pulses centered at 70% along the bone length with amplitudes and breadths of approximately 10% and 20%, respectively. The actual amplitudes and breadths of the stiffness increase were varied randomly to simulate an experiment involving $N$=8 randomly sampled subjects in which the bone's anisotropic stiffness profile was measured separately for each subject. Additionally, a small random signal was separately applied to each of the eight cases to ensure that variance was greater than zero, and thus that test statistic values were computable at all points in the continuum.

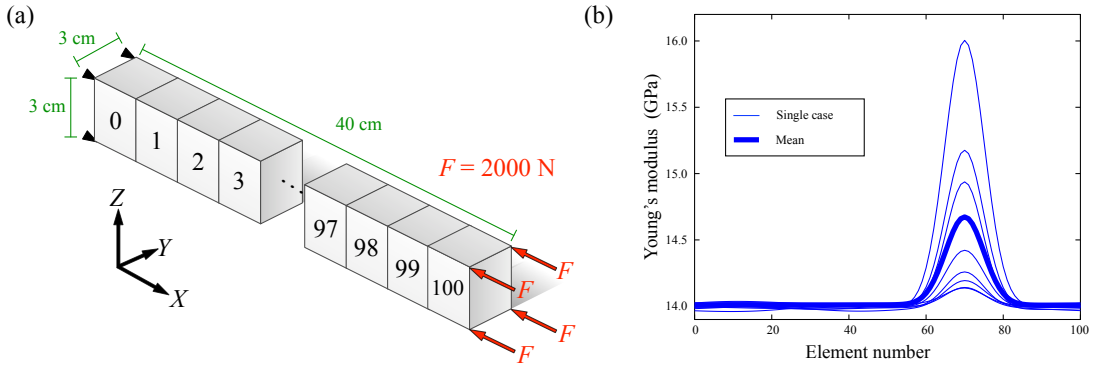


**Figure 2.** Model A. (a) Stack of cuboids representing a simplified bone. (b) Elemental Young's moduli representing local stiffness increase in $N$=8 cases.

### *2.1.2 Model B: Soft tissue indentation*

A rigid hexahedral block was compressed against soft tissue to a depth of 1 cm height as depicted in Fig.3. Nodal displacements on the soft tissue's bottom surface were fully constrained. The soft tissue was modeled as hyperelastic with the following Moony-Rivlin strain energy function (Maas et al., 2015) :

$$W = a(I-3) + \frac{k}{2}(\ln J)^2 \tag{1}$$

Here $a$ is the hyperelastic parameter, $k$ is the elasticity volume modulus, $I$ is the deformation tensor's first deviatoric invariant, and $J$ is the deformation Jacobian. The parameter $a$ was set to 100 and eight $k$ values (800, 817, 834, 851, 869, 886, 903, 920) were compared to a datum case of $k$=820.

Additionally, three different indenter face types were compared. The first indenter face was perfectly flat, and the other two were uneven but smooth as depicted in Fig.4. The uneven surfaces were generated by adding spatially smoothed Gaussian noise to the indenter face's z coordinates (i.e. the compression direction), then scaling to a maximum value of approximately 2.5 mm, or 1.7% the indenter's height.

### *2.1.3 Model C: Hip cartilage compression*

A separately-published model of hip cartilage compression (Maas et al., 2015) (Fig.5) was selected to demonstrate how the concepts from the simple models A and B above may extend to realistic biomedical applications. This model is available in the FEBio test suite (`febio.org`; model name: "hip_n10rb"), and the scripts we used to manipulate this model are available in this paper's GitHub repository (`github.com/0todd0000/probFEApy`).

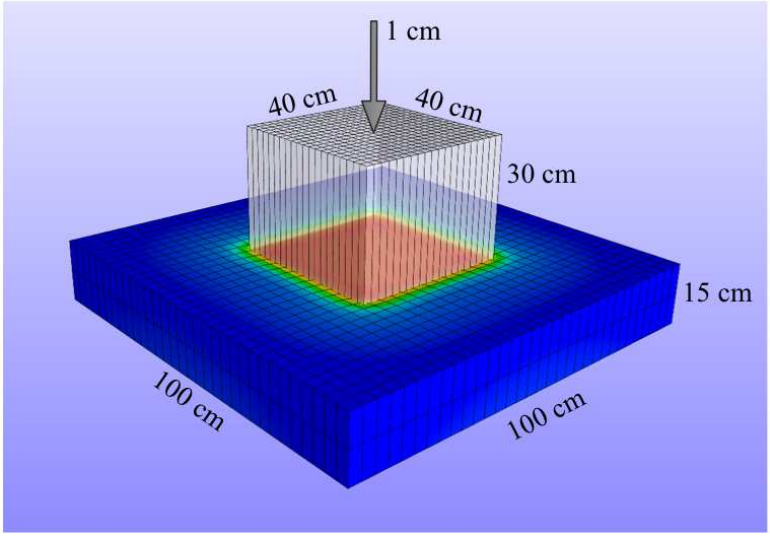


**Figure 3.** Model B: rigid block indentation on a hyperelastic material.

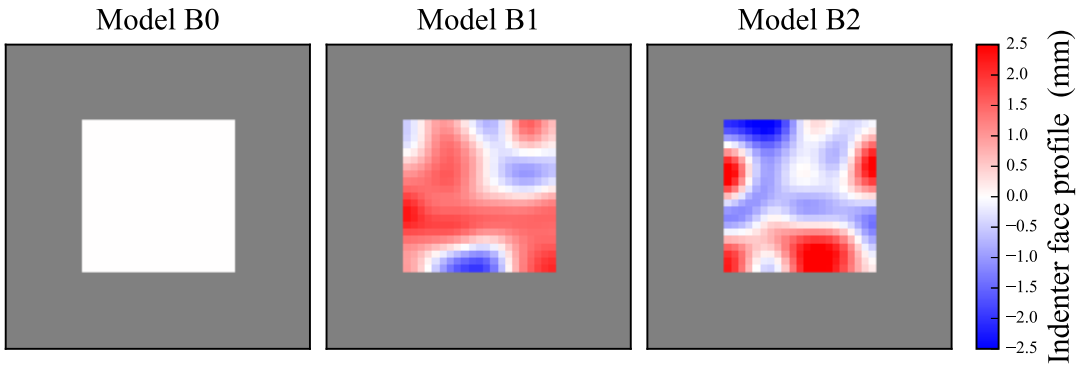


**Figure 4.** Model B indenter faces. The grey area depicts the compressed soft tissue.

The bones were rigid and the cartilage was modeled using the hyperelastic Mooney-Rivlin model above (Eqn.1) with a constant $a$ value of 6.817. Ten different values of $k$ were simulated for each of two hypothetical groups (Table 1) to mimick a two-sample experiment involving *in vivo* or *in vitro* material property measurements. The pelvis and acetabular cartilage were fixed and the femur was kinematically driven 1 mm in the upward direction.

**Table 1.** Model C material parameters; see Eqn.1. SD = standard deviation.

| Group | Mooney-Rivlin $k$ values | Mean (SD) |
|---|---|---|
| 1 | [1200, 1230, 1260, 1290, 1320, 1350, 1380, 1410, 1440, 1470] | 1335 (90.8) |
| 2 | [1380, 1410, 1440, 1470, 1500, 1530, 1560, 1590, 1620, 1650] | 1515 (90.8) |

## 2.2 Analysis

We used a non-parametric permutation method from the Neuroimaging literature (Nichols and Holmes, 2002) to conduct classical hypothesis testing at the whole-model level. The technique employs observation permutation to generate non-parametric approximations to probabilities from (parametric) multi-dimensional

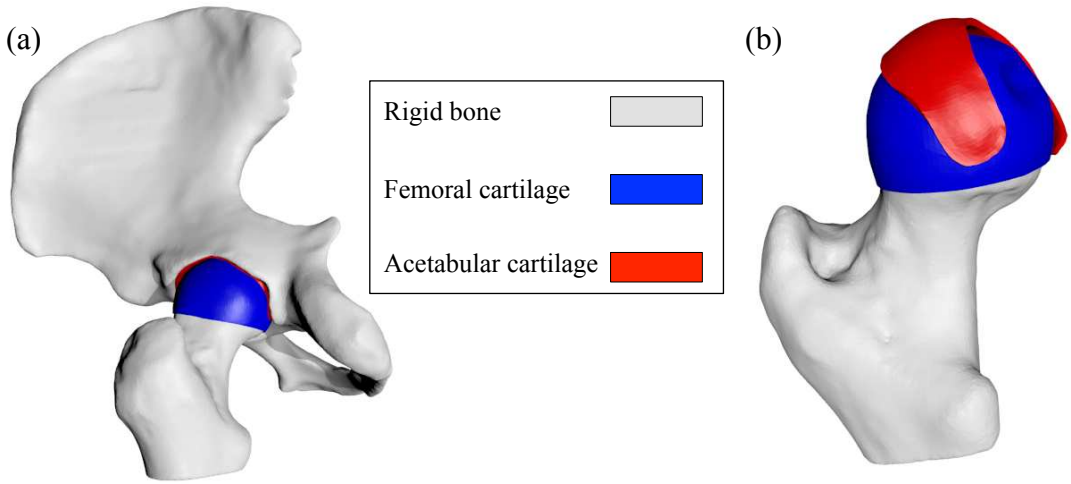


**Figure 5.** Model C; "hip_n10rb" from the FEBio test suite containing femoral and acetabular cartilage compressed via rigid bone displacement. (a) Full model. (b) Pelvis removed to expose the cartillage surface geometries.

Random Field Theory (Adler and Taylor, 2007). The method is described below and is depicted in Fig.6. All permutations described below were applied to pre-simulated FEA results.

#### *2.2.1 Model A*

The datum Young's modulus ($E$=14 GPa) was subtracted from the eight 1D Young's modulus continua (Fig.2b), and the resulting difference continua were sign-permuted (Fig.6a) to generate a number of artificial data samples. For each sample the $t$ continuum was computed according to the typical one-sample $t$ statistic definition:

$$t(\boldsymbol{q}) = \frac{\overline{y(\boldsymbol{q})} - \mu(\boldsymbol{q})}{s(\boldsymbol{q})/\sqrt{N}} \tag{2}$$

where $\bar{y}$ is the sample mean, $\mu$ is the datum, $s$ is the sample standard deviation, $N$ is sample size and $\boldsymbol{q}$ is continuum position. Repeating for all permutation samples produced a distribution of 1D $t$ continua (Fig.6b), whose maxima formed a 'primary' probability density function (PDF) (Fig.6c). This primary PDF represents the expected maximum difference (from the datum case of $E$ = 14 GPa) that smooth, purely random continua would be expected to produce if there were truly no effect.

We conducted classical hypothesis testing at $\alpha$=0.05 using the primary PDF's 95th percentile ($t^*$) as the criterion for null hypothesis rejection; if the $t$ continuum associated with original, non-permuted data (Fig.6a) exceeded $t^*$ the null hypothesis was rejected. In this example the original $t$ continuum failed to traverse $t^*$ (Fig.6e) so the null hypothesis was not rejected. Based on the primary PDF the exact probability value was: $p = 0.101$ in the depicted example.

We repeated this procedure for the effective strain and von Mises stress distributions associated with the eight Young's modulus continua. In cases where the original $t$ continuum exceeded the $t^*$ threshold, probabilities associated with the upcrossing(s) (Fig.1) were computed with a 'secondary' PDF (Fig.6d) which embodied the probability of observing upcrossings with particular volume (i.e. supra-threshold integral). Note that (i) (1-$\alpha$)% of the values in the secondary PDF are zero by definition, (ii) an upcrossing which infinitessimally exceeds $t^*$ has an integral of zero and a $p$ value of $\alpha$, and (iii) the minimum upcrossings $p$ value is 1/$n$, where $n$ is the total number of permutations. All integrals were computed using trapezoidal approximation.

#### *2.2.2 Model A, Part 2*

We conducted a secondary analysis of Model A to examine how additional probabilistic variables increase computational demand. For this analysis we considered load direction ($\theta$) to be uncertain,

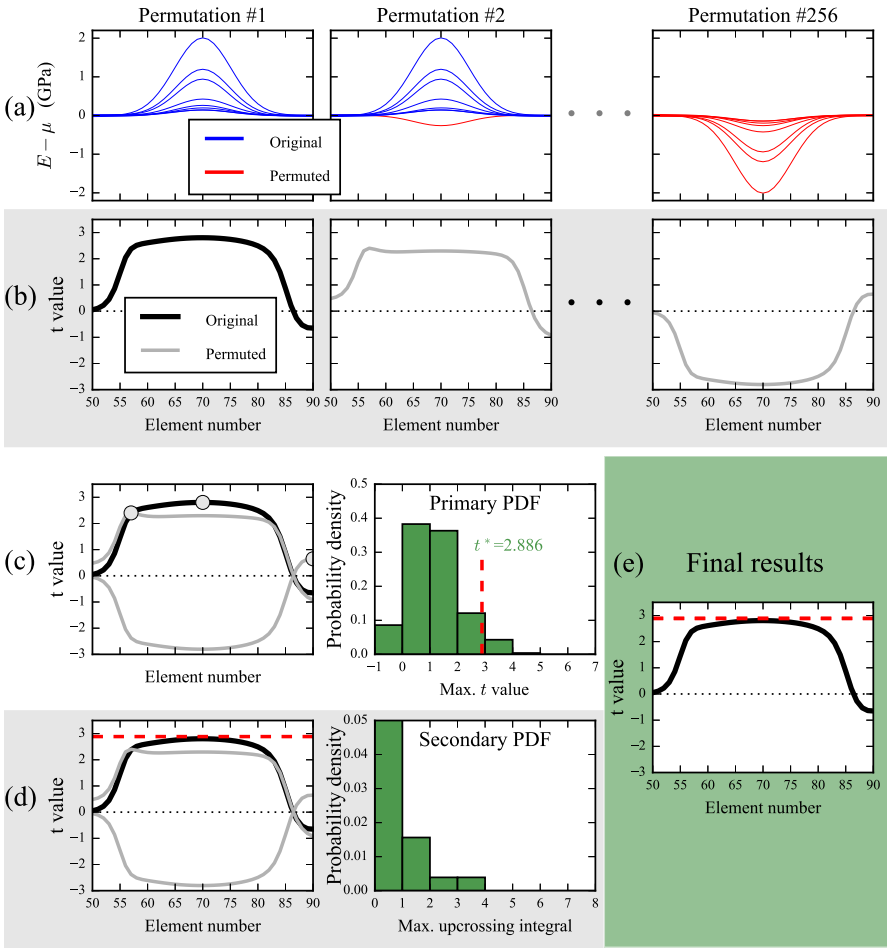


**Figure 6.** Depiction of non-parametric, permutation-based continuum-level hypothesis testing. This example uses five of the Young's modulus continua from Fig.2b and compares the mean continuum to the datum: $\mu$=14 GPa. (a) Original continua were sign-permuted by iteratively multiplying subsets by $-1$. (b) For each permutation a $t$ continuum was computed using Eqn.2 . (c) The maximum $t$ values from all permutations were assembled to form a primary probability density function (PDF) from which a critical test statistic ($t^*$) was calculated. (d) Thresholding all permuted test statistic continua at $t^*$ produced upcrossings (Fig.1) whose integral formed a secondary PDF from which upcrossing-specific $p$ values are computable. (e) Since the original test statistic continuum failed to traverse $t^*$ the null hypothesis was not rejected at $\alpha$=0.05 for this example.

with a mean of zero and a standard deviation of 3 deg (forces with $\theta$=0 deg are depicted in Fig.2a, and these forces were rotated about the depicted Y axis). For typical simulation of random variables hundreds or thousands of simulations are usually needed to achieve probability distribution convergence (Dopico-González et al., 2009) , but we aimed to show that computational increases may be minimal for the proposed hypothesis testing framework.

We randomly varied $\theta$ for an additional 400 FE simulations, 50 for each of the observations depicted in Fig.2b. We then qualitatively compared the permutation-generated distribution of $t$ continua after just 16 simulations (one extra FE simulation for each observation) to the distribution obtained after 400 FE simulations. To quantitatively assess the effects of the number of simulations $N$ on the distributions we examined the null hypothesis rejection rate for the $N$=16 and $N$=400 cases as a function of the number of post-simulation permutations.

### *2.2.3 Model B*

The goal of Model B analysis was to qualitatively assess the effects of imperfect contact geometry (Fig.4) on both mean FE simulation results and statistical interpretations. Nine simulations were conducted

for each of the three indenter faces (Fig.4): one datum ($k$=820) and then the eight other values of $k$ as described above. For each indenter we computed the mean von Mises stress distribution in the compressed soft tissue, then compared this mean to the datum ($k$=820) stress distribution through the one-sample test statistic (Eqn.2) .

#### *2.2.4 Model C*

The goal of Model C analysis was to demonstrate how the analysis techniques and results for Model A and Model B extend to realistic, complex models. The null hypothesis of equivalent von Mises stress distributions in each group (Table 1) was tested using a slight modification of the permutation approach described above (Fig.6). The only differences were that (i) the two-sample $t$ statistic was computed instead of the one-sample $t$ statistic, and (ii) group permutations were conducted instead of sign permutations. Group permutations were performed by randomly assigning each of the 20 continuum observations to one of the two groups, with ten observations in each group, then repeating for a total of 10,000 random permutations. Although the total number of possible permutations was 20! / (10! 10!) = 184,756, we found no qualitative effect of adding more than 10,000 permutations.

## 3 RESULTS

### 3.1 Model A

FE simulations of each of the eight cases depicted in Fig.2b yielded the stress/strain distributions and $t$ statistic distributions depicted in Fig.7. In this example Young's moduli only increased (Fig.7a) and strain only decreased (Fig.7b), but stress exhibited central increases (near element #70) and peripheral decreases (near elements #60 and #80) (Fig.7c), emphasizing the nonlinear relation between model inputs and outputs.

Maximum absolute $t$ values differed amongst the field variables (Fig.7d–f), with stress exhibiting the largest maximum absolute $t$ values. The null hypothesis was rejected for von Mises stresses but not for either Young's modulus or effective strain. Additionally, both stress increases and stress decreases were statistically significant (Fig.7f). These results indicate that statistical signal associated with the Young's modulus inputs was amplified in the von Mises stress field, but we note that strain would have been the amplified variable had the the model been displacement-loaded instead of force-loaded. More generally these results show that statistical conclusions pertaining to different model variables can be quite different, and that different continuum regions can respond in opposite ways to probabilistic inputs.

Although stiffness increased non-uniformly as a Gaussian pulse (Fig.7a) the test statistic magnitude was effectively uniform across that region (elements 60 – 80; Fig.7d). This suggests that mechanical and statistical magnitudes are not directly related, and thus that statistical conclusions mustn't be limited to areas of large mechanical signal unless one's hypothesis pertains specifically to those areas.

### 3.2 Model A, Part 2

Adding uncertainty to the load direction increased variability and thus caused absolute $t$ value decreases near element #70 (Fig.8a), but general loading environment changes caused increases to absolute $t$ values in other model areas, especially toward elements #50 and #90. The stress response was somewhat different , with absolute $t$ values increasing near element #70 but decreasing elsewhere (Fig.8c), re-emphasizing the complex relation amongst different field variables' response to probabilistic model features.

The $t$ distributions for stress and strain were not qualitatively affected by the number of additional FE simulations; 16 simulations, or one extra simulation per observation (Fig.8a,c) yielded essentially the same results as 400 simulations (Fig.8b,d). The reason is that permutation leverages variability in small samples to produce a large number of artificial samples, and thereby approximates the results of a large number of FE simulations.

To quantify $t$ continuum distribution stability as a function of the number of permutations we considered the null hypothesis rejection rate in both cases of 16 and 400 FE simulations (Fig.9). After approximately 200 permutation iterations the null hypothesis rejection rate was effectively identical for both 16 and 400 FE simulations. These results suggest that permutation, which is extremely fast compared to FE simulation, may be able to effectively approximate a large number of FE simulations using the results of only a few FE simulations.

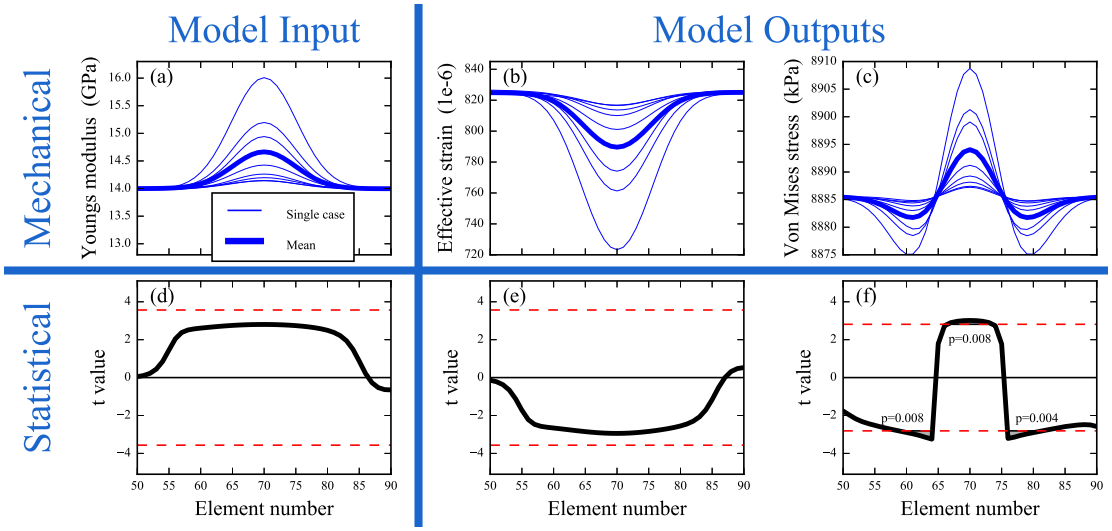


**Figure 7.** Model A results. (a–c) Young's modulus input observations and strain/stress continua associated with each observation. (d–f) Hypothesis testing results ($\alpha$=0.05); red dotted lines depict critical thresholds.

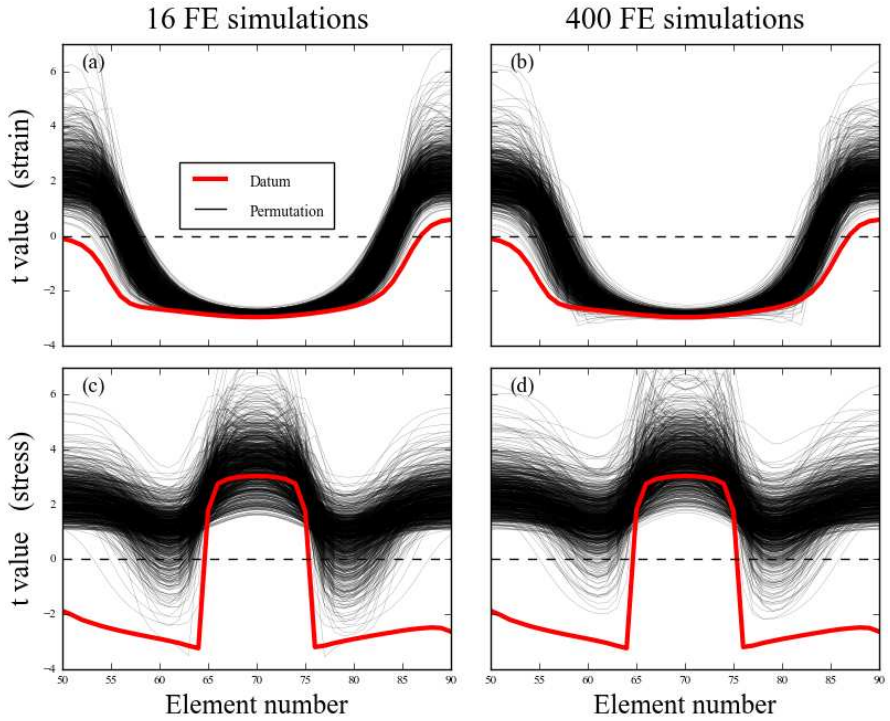


**Figure 8.** Model A $t$ distributions for strain (upper panels) and stress (lower panels) under a load direction uncertainty with a standard deviation of 3 deg.

### 3.3 Model B

The mean stress distributions associated with the three indenter faces (Fig.10) closely followed indenter face geometry (Fig.3). Variation in material parameters was associated with stress distribution variability (Fig.11a). Nevertheless, $t$ values were effectively constant across all elements and all three models (Fig.11b). This suggests that test statistic continua are more robust to model geometry imperfections than are stress/strain continua.

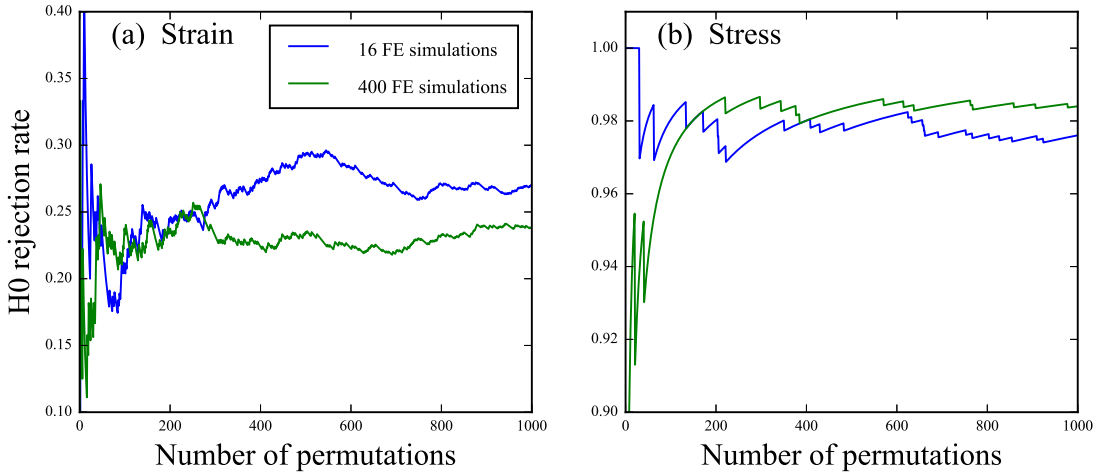


**Figure 9.** Model A null hypothesis (H0) rejection rate as a function of the number of permutations for both 16 and 400 FE simulations.

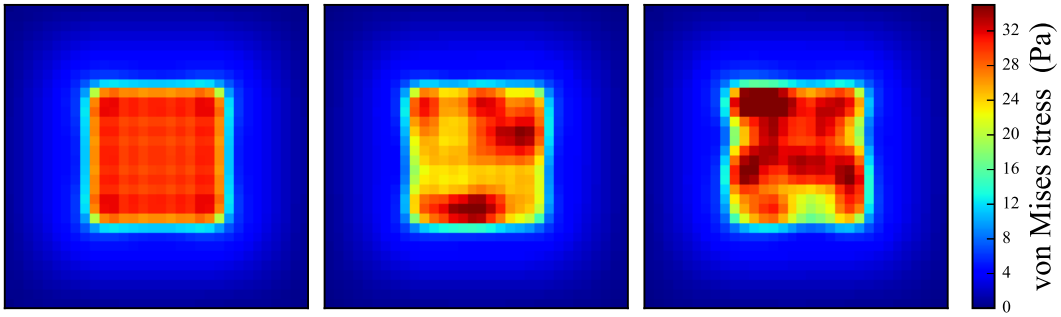


**Figure 10.** Model B mean stress distributions for the three indenter faces. Note that these patterns closely follow the indenter face geometry depicted in Fig.4.

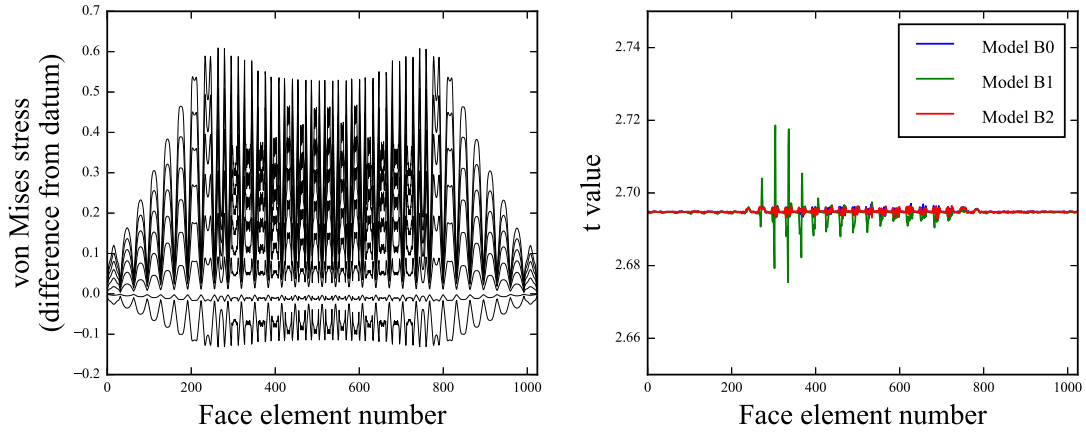


**Figure 11.** Model B statistical results. (a) Large variation was present amongst the individual continua (only Model B0 results shown). (b) Test statistic continua were effectively constant in all elements and across all three indenters.

### 3.4 Model C

A two-sample $t$ test regarding the material parameters (Table 1) yielded $t$=5.17, $p$<0.001 and thus a rejection of the null hypothesis of equal group means. These probabilistic material parameters produced

mean stresses which were generally higher in Group B vs. Group A (Fig.12), where a stress distribution difference plot clarified that inter-group differences were generally confined to areas of large stress (Fig.13). The inter-group statistical differences were much broader, covering essentially the entire femoral cartilage (Fig.14). Moreover, relatively broad regions of the cartilage exhibited significant stress *decreases*, similar to the result observed in the simple bone model (Fig.7f).

These results reiterate many of the aforementioned methodological points. In particular, changes in probabilistic model inputs (in this case: material parameter values) can have statistical effects on output fields (in this case: von Mises stresses) which are not easily predicted. Additionally, the visual advantages of full-field analyses are somewhat clearer in this more anatomically correct model; tabulated stresses from different regions of the femoral cartilage would be more difficult to interpret in terms of the original anatomy. Last, mechanical (Fig.13) and statistical (Fig.14) results can be quite different.

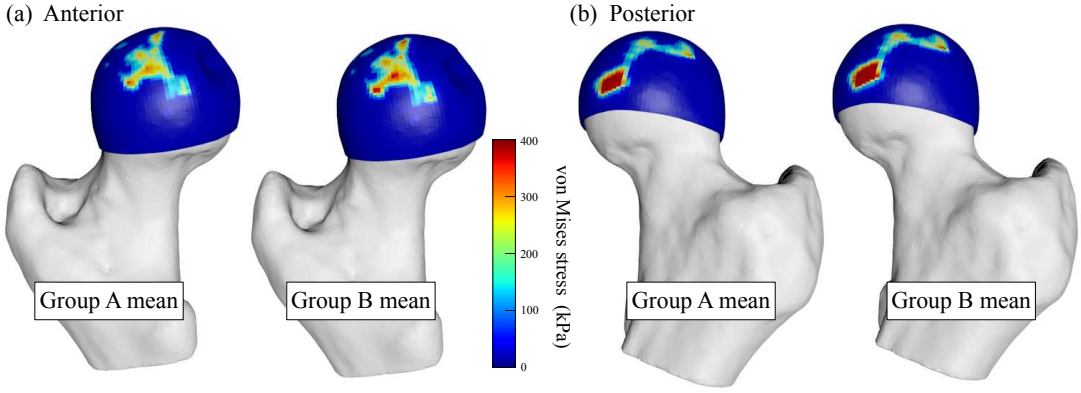


**Figure 12.** Model C: mean stress distributions.

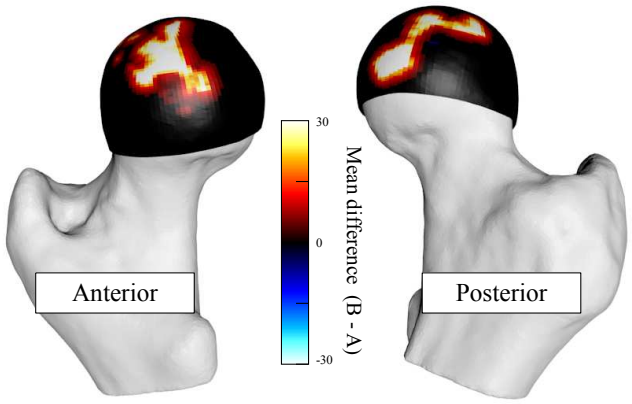


**Figure 13.** Model C: mean stress difference.

## 4 DISCUSSION

This paper demonstrated how a non-parametric permutation technique from Neuroimaging (Nichols and Holmes, 2002) can be used to conduct classical continuum-level hypothesis testing for finite element (FE) models. It's main advantages are:

1. Easy implementation. As demonstrated in this project's software repository (`github.com/0todd0000/probFEApy`), non-parametric hypothesis testing for FE models can be implemented using relatively compact scripts.

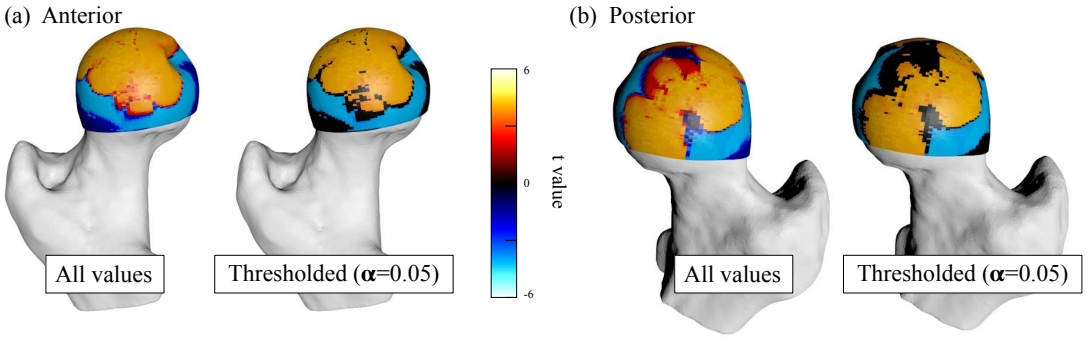


**Figure 14.** Model C: statistical results.

2. Computational efficiency. After simulating subject-specific results — which is usually necessary in arbitrary multi-subject studies — no additional FE simulations are needed; permutation can operate on pre-simulated small-sample results to approximate large-sample probabilities (Fig.6). Producing the main Model A results (Fig.7) required a total of only 1.3 s to execute on a desktop PC, including both FE simulations and permutation-based probability computation.

3. Non-measured uncertainty capabilities. Adding uncertainty in the form of random model parameters does not necessarily require large increases in computational demands; results suggest that with respect to an original dataset with $N$ simulations, it may be possible to robustly accommodate additional uncertainty with just $N$ additional simulations (Figs.8–9).

4. Visual richness and tabulation elimination. Continuum-level hypothesis testing results can be presented in the same geometric context as commonly visualized field variables like stress and strain (Fig.7b,e and Fig.7c,f), which eliminates the need to separately tabulate statistical results.

5. Arbitrarily complex experiments. While only one- and two-sample designs were considered here, $t$ statistic continua generalize to $F$ and all other test statistic continua, so arbitrarily complex designs ranging from regression to MANCOVA can be easily implemented using permutation.

6. Robustness to geometric imperfections. Small geometric changes can have qualitatively large effects on stress/strain continua, but have comparably little-to-no effect on test statistic continua (Fig.**??**), implying that continuum-level hypothesis testing may be more robust than commonly employed procedures which analyze local maxima. This potential danger is highlighted in the more realistic Model C, whose mean differences (Fig.13) exhibited high focal stresses whereas the statistical continuum was much more constant across the contact surface (Fig.14).

### 4.1 Mechanical vs. statistical interpretations

Mechanical and statistical continua are generally different. For example, for Model A it is clear that each stiffness increase (Fig.2b) has *mechanical* effects on the strain/stress continuum, but the *statistical* effects are less clear because there is relatively large uncertainty regarding the true nature of the stiffness increase in the population that this sample represents. For classical hypothesis testing, mechanical meaning is irrelevant because all mechanical effects must be considered with respect to their uncertainty. Further emphasizing the tenuous relation between mechanical and statistical meaning are regions of small mechanical signals (for Model A: near the periphery of the stiffness increase region) which can be accompanied by relatively large statistical signals.

To objectively conduct classical hypothesis tests on FEA results it is therefore essential to explicitly identify the hypothesis prior to conducting simulations. If limiting analyses to only areas of large mechanical signal can be justified in an *a priori* sense, then those, and only those areas should be analyzed without any theoretical problem. If, however, one's *a priori* hypothesis pertains to general stress / strain distribution changes, and not specifically to areas of high mechanical signal, it may be necessary to consider the entire model because maximal mechanical and maximal statistical signals do not necessarily coincide.

## 4.2 Comparison with common techniques

In the literature, FE-based classical hypothesis testing is typically conducted via scalar analysis of local extrema (Radcliffe and Taylor, 2007). Applying that approach to the local mechanical change extrema in Model A (Fig.7a–c) yielded the results in Table 2. The null hypothesis (of no mean change with respect to the 14 GPa case) was rejected at $\alpha = 0.05$ for all three mechanical variables: Young's modulus, effective strain and von Mises stress.

While the test statistic magnitudes are the same for both the proposed whole-model approach (Fig.7) and these local extremum analyses, the critical threshold at $\alpha$=0.05 is different because the spatial scope is different. The broader the spatial scope of the hypothesis, the higher the threshold must be to avoid false positives (Friston et al., 2007); in other words, random processes operating in a larger volume have a greater chance of reaching an arbitrary threshold.

The proposed model-wide approach (Fig.7) and the local extremum (scalar) approach have yielded contradictory hypothesis testing conclusions for both Young's modulus and strain distributions, so which approach is correct? The answer is that both are correct, but both cannot be simultaneously correct. The correct solution depends on the *a priori* hypothesis, and in particular the spatial scope of that hypothesis. If the hypothesis pertains to only the local extremum, then the local extremum approach is correct, and whole-model results should be ignored because they are irrelevant to the hypothesis. Similarly, if the hypothesis pertains to the whole model, then the whole model results are correct and local extrema results should be ignored because they are irrelevant to the hypothesis. We would argue that all FE analyses implicitly pertain to the whole model unless otherwise specified, and that focus on specific scalar metrics is appropriate only if justified in an *a priori* manner.

**Table 2.** Model A results. Analyses of local extrema (at element 70) using a non-parametric permutation-based two-sample *t* test. SD = standard deviation.

| Variable | Mean | SD | *t* | *p* |
|---|---|---|---|---|
| Young's modulus (GPa) | 14.665 | 0.670 | 2.804 | 0.026 |
| Effective strain (1e-6) | 789.6 | 33.9 | -2.946 | 0.022 |
| von Mises stress (kPa) | 8894.0 | 8.0 | 3.014 | 0.020 |

Historically in biomechanical FEA, low sample sizes (frequently n = 1 for each model) permitted nothing more than qualitative comparisons of stress or strain maps, and/or numerical comparison of output parameters at single nodes. Nevertheless conventional FEA can concurrently and ironically suffer from an excess of data when results are tabulated over many regions, often in a non-standardized manner across studies.

With the continued increase of computer power and processing speed, FE models comprising over one million elements are becoming more and more common (Moreno et al., 2008; Bright and Rayfield, 2011a; Cox et al., 2013, 2015; Cuff et al., 2015) (e.g. Moreno et al, 2008; Bright & Rayfield, 2011a; Cox et al, 2013, 2015; Cuff et al, 2015). Yet, typically stress and strain values are only reported and analysed from just a few elements (Porro et al., 2013; Fitton et al., 2012a). Alternatively average or peak stress or strain values can be computed for whole models (Dumont et al., 2011; Cox et al., 2012; Parr et al., 2013; Sharp and Rich, 2016) or selected regions (Wroe et al., 2007a,b; Nakashige et al., 2011). The recent application of geometric morphometrics to FEA results (Cox et al., 2011; Fitton et al., 2012b; O'Higgins and Milne, 2013) has gone some way to providing a method of analysing whole models rather than individual elements, but is limited to the analysis of deformations. The approach outlined here enables, for the first time, the analysis of all stresses or strains in a single hypothesis test.

Another major benefit of the technique outlined here is its ability to take in consideration input parameters that are only imprecisely known. When modelling biological structures, the material properties of the model, and the magnitude and orientations of the muscle loads cannot always be directly measured. This is an especially acute problem in studies dealing with palaeontological taxa. Previous research has addressed this issue principally by the use of sensitivity analyses which test the sensitivity of a model to changes in one or more unknown parameters (Kupczik et al., 2007; Bright and Rayfield, 2011a; Cox

et al., 2011, 2015; Reed et al., 2011; Wood et al., 2016; Toro-Ibacache et al., 2016). The models are identical save for the unknown parameters, which are then varied between extremes representing likely biological limits or the degree of uncertainty. In such studies, the number of different models is usually quite low, with each parameter only being tested at a maximum of five different values. Our method takes this approach to its perhaps logical extreme – the unknown parameter is allowed to vary randomly within defined limits over a large number of iterations (usually on the order of 10,000). These iterations produce a distribution of results that can be statistically compared with other such distributions.

A final advantage is that statistical continua may be less sensitive to geometric mesh peculiarities than stress / strain continua. In Fig.10 and Fig.13, for example, it is clear from the oddly shaped regions of stress difference that these effects were likely caused by mesh irregularities and that remeshing would likely smooth out these areas of highly localized stress changes. The test statistic continuum, on the other hand, appeared to be considerably less sensitive to localization effects (Fig.11) and (Fig.14). This may imply that one needn't necessarily develop an ideal mesh, because statistical analysis may be able to mitigate mesh peculiarity-induced stress distribution irregularities.

### 4.3 Limitations

The major limitation of the proposed method as it currently stands is that only models of identical geometry can be compared. Thus, while the technique can be readily used to address sensitivity-like questions regarding material properties, boundary conditions and orientations, the method cannot readily address geometry-relevant questions, such as are created by varying mesh density (Bright and Rayfield, 2011b; Toro-Ibacache et al., 2016), or are found in between-taxa analyses (Dumont et al., 2005, 2011; Oldfield et al., 2012; Cox et al., 2012; Wroe et al., 2007a; Sharp, 2015). Nevertheless, through three-dimensional anatomical registration (Friston et al., 2007) and also potentially intra-model spatial interpolation to common continuum positions $\boldsymbol{q}$ (Eqn.2), it may be possible to apply the technique to arbitrary geometries even in cases of large deformation and/or geometrical disparity (Schnabel et al., 2003).

A second limitation is computational feasibility. Although our results suggest that incorporating a single additional uncertain parameter into the model may not greatly increase computational demand, this may not be true for higher dimensional parameter spaces. In particular, given $N$ experimental measurements, our results show that $2N$ simulations are sufficient to achieve probabilistic convergence (Fig.9). However, this result may be limited to cases where the uncertainty is sufficiently small so that it fails to produce large qualitative changes in the underlying stress/strain continua. Moreover, the feasibility for higher-dimensional parameter spaces is unclear. In particular, a sample of $N$ observations is likely unsuitable for an $N$-dimensional parameter space, or even an $N/2$-dimensional parameter space. The relation between uncertainty magnitude, number of uncertain parameters, the sample size and the minimum number of FE simulations required to achieve probabilistic convergence is an important topic that we leave for future work.

A third potential limitation is that both upcrossing features and the test statistic continuum can be arbitrary. In this paper we restricted analyses to the upcrossing maximum and integral due to the robustness of these metrics with respect to other geometric features (Zhang et al., 2009). Other upcrossing metrics and even arbitrary test statistic continua could be submitted to a non-parametric permutation routine. This is partly advantageous because arbitrary smoothing can be applied to the continuum data, and in particular to continuum variance (Nichols and Holmes, 2002), but it is also partly a disadvantage because it increases the scope of analytical possibilities and thus may require clear justification and/or sensitivity analyses for particular test statistic and upcrossing metric choices.

A final limitation is that the both the test statistic and probability continua are directly dependent on the uncertainty one selects via model parameter variance. This affords scientific abuse because it allows one to tweak variance parameters until the probabilistic results support one's preferred interpretation. We therefore recommend that investigators both clearly justify variance choices and treat variance itself as a target of sensitivity analysis.

### 4.4 Summary

This paper has proposed a probabilistic finite element simulation method for conducting classical hypothesis testing at the continuum level. The technique leverages probability densities regarding geometric features of continuum upcrossings, which can be rapidly and non-parametrically estimated using iterative permutation of pre-simulated stress/strain continua. The method yields test statistic continua which are visually rich, which may eliminate the need for tabulated statistical results, which may reveal unique

biomechanical information, and which also may be more robust to mesh and other geometrical model peculiarities than stress/strain continua.

## REFERENCES


Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer-Verlag.

Babuska, I. and Silva, R. S. (2014). Dealing with uncertainties in engineering problems using only available data. *Computer Methods in Applied Mechanics and Engineering*, 270(C):57–75.

Boyaval, S. (2012). A fast Montee-Carlo method with a reduced basis of control variates applied to uncertainty propagation and Bayesian estimation. *Computer Methods in Applied Mechanics and Engineering*, 241-244(C):190–205.

Bright, J. and Rayfield, E. (2011a). Sensitivity and ex vivo validation of finite element models of the domestic pig cranium. *Journal of Anatomy*, 219:456–471.

Bright, J. and Rayfield, E. (2011b). The response of cranial biomechanical finite element models to variations in mesh density. *Anatomical Record*, 294:610–620.

Cheung, J.-M., Zhang, M., Leung, A.-L., and Fan, Y.-B. (2005). Three-dimensional finite element analysis of the foot during standing – a material sensitivity study. *Journal of Biomechanics*, 38:1045–1054.

Cox, P., Fagan, M., Rayfield, E., and Jeffery, N. (2011). Finite element modelling of squirrel, guinea pig and rat skulls: using geometric morphometrics to assess sensitivity. *Journal of Anatomy*, 219:696–709.

Cox, P., Kirkham, J., and Herrel, A. (2013). Masticatory biomechanics of the Laotian rock rat, Laonastes aenigmamus, and the function of the zygomaticomandibularis muscle. *PeerJ*, 1:e160.

Cox, P., Rayfield, E., Fagan, M., Herrel, A., Pataky, T., and Jeffery, N. (2012). Functional evolution of the feeding system in rodents. *PLOS ONE*, 7(4):e36299.

Cox, P., Rinderknecht, A., and Blanco, R. (2015). Predicting bite force and cranial biomechanics in the largest fossil rodent using finite element analysis. *Journal of Anatomy*, 226:215–223.

Cuff, A., Bright, J., and Rayfield, E. (2015). Validation experiments on finite element models of an ostrich (Struthio camelus) cranium. *PeerJ*, 3:e1294.

Dar, F. H., Meakin, J. R., and Aspden, R. M. (2002). Statistical methods in finite element analysis. *Journal of Biomechanics*, 35:1155–1161.

Dopico-González, C., New, A. M., and Browne, M. (2009). Probabilistic analysis of an uncemented total hip replacement. *Medical Engineering and Physics*, 31(4):470–476.

Dumont, E., Davis, J., Grosse, I., and Burrows, A. (2011). Finite element analysis of performance in the skulls of marmosets and tamarins. *Journal of Anatomy*, 218:151–162.

Dumont, E., Piccirillo, J., and Grosse, I. (2005). Finite-element analysis of biting behaviour and bone stress in the facial skeletons of bats. *Anatomical Record Part A*, 283:319–330.

Easley, S. K., Pal, S., Tomaszewski, P. R., Petrella, A. J., Rullkoetter, P. J., and Laz, P. J. (2007). Finite element-based probabilistic analysis tool for orthopaedic applications. *Computer Methods and Programs in Biomedicine*, 85(1):32–40.

Fitton, L., Proa, M., Rowland, C., Toro-Ibacache, V., and O'Higgins, P. (2012a). The impact of simplifications on the performance of a finite element model of a Macaca fascicularis cranium. *Anatomical Record*, 298:107–121.

Fitton, L., Shi, J., Fagan, M., and O'Higgins, P. (2012b). Masticatory loadings and cranial deformation in Macaca fascicularis:a finite element analysis sensitivity study. *Journal of Anatomy*, 221:55–68.

Fitzpatrick, C. K., Baldwin, M. A., Clary, C. W., Wright, A., Laz, P. J., and Rullkoetter, P. J. (2012). Identifying alignment parameters affecting implanted patellofemoral mechanics. *Journal of Orthopaedic Research*, 30(7):1167–1175.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, London.

Kupczik, K., Dobson, C., Fagan, M., Crompton, R., Oxnard, C., and O'Higgins, P. (2007). Assessing mechanical function of the zygomatic region in macaques: validation and sensitivity testing of finite element models. *Journal of Anatomy*, 210:41–53.

Laz, P. J., Stowe, J. Q., Baldwin, M. A., Petrella, A. J., and Rullkoetter, P. J. (2007). Incorporating uncertainty in mechanical properties for finite element-based evaluation of bone mechanics. *Journal of Biomechanics*, 40(13):2831–2836.

Legay, A. and Viswanatha, M. (2009). Simulation + Hypothesis Testing for Model Checking of Proba-

bilistic Systems. In *2009 Sixth International Conference on the Quantitative Evaluation of Systems (QEST)*, pages 3–3. IEEE.

Lin, C.-L., Chang, S.-H., Chang, W.-J., and Kuo, Y.-C. (2007). Factorial analysis of variables influencing mechanical characteristics of a single tooth implant placed in the maxilla using finite element analysis and the statistics-based Taguchi method. *Eur J Oral Sci*, 115:408–416.

Maas, S., Ellis, B., Ateshian, G., and Weiss, J. (2012). FEBio: Finite Elements for Biomechanics. *Journal of Biomechanical Engineering*, 134(1):011005–10.

Maas, S., Rawlins, D., Weiss, J., and Ateshian, G. (2015). FEBio theory manual version 2.4.

McFarland, J. and Mahadevan, S. (2008). Multivariate significance testing and model calibration under uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 197(29-32):2467–2479.

Moreno, K., Wroe, S., Clausen, P., McHenry, C., D'Amore, D., and Rayfield, EJ amd Cunningham, E. (2008). Cranial performance in the Komodo dragon (Varanus komodoensis) as revealed by high-resolution 3-D finite element analysis. *Journal of Anatomy*, 212:736–746.

Nakashige, M., Smith, A., and Strait, D. (2011). Biomechanics of the macaque postorbital septum investigated using finite element analysis: implications for anthropoid evolution. *Journal of Anatomy*, 218:142–150.

Nichols, T. and Holmes, A. (2002). Nonparametric Permutation Tests for Functional Neuroimaging: A Primer with Examples. *Human Brain Mapping*, 15(1):1–25.

O'Higgins, P. and Milne, N. (2013). Applying geometric morphometrics to compare in size and shape arising from finite elements analyses. *Hystrix, the Italian Journal of Mammalogy*, 24:126–132.

Oldfield, C., McHenry, C., Clausen, P., Chamoli, U., Parr, W., Stynder, D., and Wroe, S. (2012). Finite element analysis or ursid cranial mechanics and the prediction of feeding behaviour in the extinct giant Agriotherium africanum. *Journal of Zoology*, 286:163–170.

Parr, W., Chamoli, U., Jones, A., Walsh, W., and Wroe, S. (2013). Finite element micro-modelling of a human ankle bone reveals the importance of the trabecular network to mechanical performance: New methods for the generation and comparison 3D models. *Journal of Biomechanics*, 46:200–205.

Porro, L., Metzger, K., Iriarte-Diaz, J., and Ross, C. (2013). In vivo bone strain and finite element modelling of the mandible of Alligator mississipiensis. *Journal of Anatomy*, 223:195–227.

Radcliffe, I. A. J. and Taylor, M. (2007). Investigation into the affect of cementing techniques on load transfer in the resurfaced femoral head: A multi-femur finite element analysis. *Clinical Biomechanics*, 22(4):422–430.

Rayfield, E. (2011). Strain in the ostrich mandible during simulated pecking and validation of specimen-specific finite element models. *Journal of Anatomy*, 218:47–58.

Reed, D., Porro, L., Iriarte-Diaz, J., Lemberg, J., Holliday, C., Anapol, F., and Ross, C. (2011). The impact of bone and suture material properties on mandibular function in Alligator mississipiensis: testing theoretical phenotypes with finite element analysis. *Journal of Anatomy*, 218:59–74.

Ross, C., Patel, B., Slice, D., Strait, D., Dechow, P., Richmond, B., and Spencer, M. (2005). Modeling masticatory muscle force in finite element analysis: sensitivity analysis using principal coordinates analysis. *Anatomical Record Part A*, 283:288–299.

Schnabel, J. A., Tanner, C., Castellano-Smith, A. D., Degenhard, A., Leach, M. O., Hose, D. R., Hill, D. L. G., and Hawkes, D. J. (2003). Validation of nonrigid image registration using finite-element methods: application to breast MR images. *IEEE Transactions on Medical Imaging*, 22(2):238–247.

Sharp, A. (2015). Comparative finite element analysis of the cranial performance of four herbivorous marsupials. *Journal of Morphology*, 276:1230–1243.

Sharp, A. and Rich, T. (2016). Cranial biomechanics, bite force and function of the endocranial sinuses in Diprotodon optatum, the largest known marsupial. *Journal of Anatomy*, in press:doi: 10.1111/joa.12456.

Stefanou, G. (2009). Comput. methods appl. mech. engrg. *Computer Methods in Applied Mechanics and Engineering*, 198(9-12):1031–1051.

Taguchi, G. (1987). *System of Experimental Design*. UNIPUB Kraus International Publications, New York.

Toro-Ibacache, V., Fitton, L., Fagan, M., and O'Higgins, P. (2016). Validity and sensitivity of a human cranial finite element model: implications for comparative studies of biting performance. *Journal of Anatomy*, 228:70–84.

van Rossum, G. (2014). The python library reference release 2.7.8.

Wood, S., Strait, D., Dumont, E., Ross, C., and Grosse, I. (2016). The effects of modelling simplifications

on craniofacial finite element models: The alveoli (tooth sockets) and periodontal ligaments. *Journal of Biomechanics*, 44:1831–1838.

Wroe, S., Clausen, P., McHenry, C., Moreno, K., and Cunningham, E. (2007a). Computer simulation of feeding behaviour in the thylacine and dingo as a novel test for convergence and niche overlap. *Proceedings of the Royal Society B*, 274:2819–2828.

Wroe, S., Moreno, K., Clausen, P., McHenry, C., and Curnoe, D. (2007b). High-resolution three-dimensional computer simulation of hominid cranial mechanics. *Anatomical Record*, 290:1248–1255.

Zhang, H., Nichols, T. E., and Johnson, T. D. (2009). Cluster mass inference via random field theory. *NeuroImage*, 44(1):51–61.