On the inference of function from structure using biomechanical modelling and simulation of extinct organisms

Biomechanical modelling and simulation techniques offer some hope for unravelling the complex inter-relationships of structure and function perhaps even for extinct organisms, but have their limitations owing to this complexity and the many unknown parameters for fossil taxa. Validation and sensitivity analysis are two indispensable approaches for quantifying the accuracy and reliability of such models or simulations. But there are other subtleties in biomechanical modelling that include investigator judgements about the level of simplicity versus complexity in model design or how uncertainty and subjectivity are dealt with. Furthermore, investigator attitudes toward models encompass a broad spectrum between extreme credulity and nihilism, influencing how modelling is conducted and perceived. Fundamentally, more data and more testing of methodology are required for the field to mature and build confidence in its inferences.

**Keywords:** musculoskeletal system; dinosaur; computer modelling; simulation; palaeontology; biomechanics

In an influential review, Lauder [1] outlined the hierarchical levels within organisms that obfuscate the relationship between structure and function. Neural control of muscles was adduced as one critical level that could give even anatomically similar muscles quite different functions. Lauder then cautioned that, while very general and very specific predictions about function using the structure of extinct taxa as primary evidence might sometimes be reliable, the more common predictions of intermediate specificity were the most problematic. His major points include that a structure–function relationship is a hypothesis that is better tested than presumed correct and that the inference of such relationships is more challenging for extinct taxa. But there are other subtleties in biomechanical modelling that include investigator judgements about the level of simplicity versus complexity in model design or how uncertainty and subjectivity are dealt with. Furthermore, investigator attitudes toward models encompass a broad spectrum between extreme credulity and nihilism, influencing how modelling is conducted and perceived. Fundamentally, more data and more testing of methodology are required for the field to mature and build confidence in its inferences.

Regardless of the modelling approach used, models can be essential simply because musculoskeletal function is typically complex. Indeed, Lauder [1] might have under-emphasized this complexity, because the complex dynamics of biomechanical systems are involved in a feedback loop with motor control and musculoskeletal structure. Segment inertial and gravitational properties, as well as interactions between linked segments, let alone environmental interactions such as limb–substrate or fluid dynamics [4], can cause muscles with similar structure or motor control to have different functions. Indeed, a compelling review [5] explained how dynamic coupling can cause muscles to influence the motions of joints they do not cross. Thus, even experimental studies or simple models of musculoskeletal systems, particularly those dichotomizing muscles into uni- versus bi-articular groups, may reach conclusions that more dynamic simulations would show to be mechanically inaccurate or even unreliable. One might then ask, how reliable is not doing computational analysis of musculoskeletal function? This is especially a tricky question for palaeontology, where most of these complex strata of data are missing.

However, just because models capture structural or mechanical complexity does not mean they are accurate or reliable. Care must be taken to avoid a dangerous attitude that, because the theoretical basis of a model may seem sound and realistic, modelling may be accurate or at least reliable. The danger lies in the assumption that we understand reality enough to model its complexity, including all the critical theoretical ingredients with no missing interactions between components that could cause surprising errors. For example, experimental and modelling studies of locomotor mechanics have provided a basic understanding of walking and running [6], but there is very little known about many of the detailed principles of these gaits (e.g. Why gait transitions occur? Why different footfall patterns are used? Or How important passive forces are?). As a result, a wide diversity of often contradictory theoretical models now exist [7]. Hence models must be carefully matched to the current state of biomechanical understanding. Where omissions are made, their implications for the results must be carefully and explicitly gauged. Here lies a trade-off, and investigator judgement call, between model reductionism and realism. Models must be complex enough but not too complex, yet where does one draw the line?

Two critical tools at the disposal of those using models to relate structure to function are ‘validation’
and 'sensitivity analysis'. A model's 'validity', or match to some form of empirical data (the higher the quality of those data, the better), must be checked. For example, computational models used to estimate body mass for extinct animals have been compared with estimates [8,9] and direct specimen-specific measurements [10] for extant animals. Those validation tests suggest approximately 50 per cent errors. Other studies [11,12] applied static modelling techniques to extant taxa to test if animals known to be proficient or poor bipedal runners would be predicted as such and found good qualitative support. However, it is a mistake to expect that the results of a validation test will match reality extremely well, because models only approximate reality, which is noisy, resulting in non-zero errors. Conversely, to approach a validation test with the bias that it cannot fail or that the purpose of validation is to prove a method is correct (to what standard of accuracy?) is just as naive. A healthier approach to both of these extremes is that the purpose of validation is to quantify how far an estimated value may deviate from empirical measurements. Or in other words, to find just how wrong the results of a model might be, given that all models are wrong to some degree. Thus, the term 'validation' is somewhat misleading [13] — the goal is to quantify just how much a model fails to replicate reality.

One reaction to this responsibility to validate models might be that palaeontologists are often not equipped in terms of tools or expertise to conduct validation tests, especially experiments with live animals. Is this an excuse from doing validation? This question is the cause for introspection—Are modelling approaches done for the right reasons if their quality is not assessed? How can those not doing validation studies determine if their models had some critical flaws? I contend that modelling alone is not enough. In structure–function analyses, one cannot dwell in the theoretical realm alone [1]. Performing validation is a win–win situation, because the process helps researchers contribute data, understand and improve methodological limits, break out-of-disciplinary pigeonholes and potentially develop new collaborations with those with the resources to conduct strong validation tests. Our models cannot be more reliable than the empirical, biological understanding that supports them. Adherence to this principle ultimately should lead to a reduction in ambiguity in modelling and greater scientific confidence in palaeobiology. The value of contributing methods and evidence with more lasting influence also should not be overlooked as a source of satisfaction—not to mention historical responsibility.

Yet, model validation alone is not enough. The level of error suggested by a validation test usually is the 'best-case' situation, in which more parameters are known than are known for models of extinct taxa. Because any model incorporates assumptions about unknown parameters, those assumptions need to be explicitly stated and their influences on model predictions need to be quantified in a sensitivity analysis [14]. This process addresses how sensitive the model's quantitative results, and thus the study's qualitative conclusions, are to the input parameters and which of those parameters are most critical to precisely quantify. In many models, this can be determined by varying one parameter at a time between minimal and maximal values (e.g. crouched and columnar limb poses) and evaluating the changes in model output (e.g. required leg muscle mass [11,12]). This task should be disconcerting because some subjective decisions (ideally backed up with reference to variation in extant taxa) are needed to circumscribe a plausible range of values. One such judgement is whether to be maximally inclusive (e.g. varying muscle maximal isometric stress 100–1000 kN m⁻²); potentially sacrificing plausibility and accuracy; or to favour a best-supported 'consensus' value (e.g. maximal stress of 200–300 kN m⁻² [15]). But not only does good modelling practice require sensitivity analysis, it is also a positive driver of later research because it identifies where future studies should focus their efforts in trying to reduce uncertainty. Again, this is a win–win situation.

Studies have shown the utility of palaeobiological sensitivity analysis [11,12] in identifying how some parameters were critical for determining running potential in large bipeds such as tyrannosaurs (figure 1). As a result, later studies developed techniques to better estimate muscle moment arms [17], centre of mass [18] and posture [16]. These studies also found that other parameters such as body mass and muscle fascicle pennation angle were relatively unimportant for those models. Muscle fascicle length remains a vexing unknown; the approach of estimating fascicle length from necessary joint range of motion [19] is one promising avenue. Such critical parameters were quite straightforward to identify in simple models [11,12]. As model complexity increases (e.g. highly dynamic models of whole stride cycles), there is a similar increase in unknowns, but this problem is not impossible to deal with using broad sensitivity analyses [20].

A useful distinction in palaeobiological modelling is to separate the issue of accuracy (how closely estimates match reality) from reliability (how robust qualitative conclusions are despite the unknown parameters are in a model used to formulate those conclusions). Ideally, validation tests quantify model accuracy through the use of statistical tests and address the repeatability of methods. Sensitivity analyses then can circumscribe the more slippery nature of reliability—which is less amenable to statistical analyses for palaeontological data because many parameters are unknowable—by bounding a range of possibility through exclusion of the impossible or at least implausible [16]. By making this distinction between accuracy and reliability, a tension becomes evident, which lies between the quantitative uncertainty that we have about palaeobiological data and the qualitative understanding we minimally seek to establish. One might argue that such uncertainty does not matter much—perhaps, theoretical models are simply phenomenological and thus do not need to be accurate to be reliable. This argument, however, is worrisome in its lack of inquisitiveness. We simply do not know how much uncertainty matters until we peer over the ledge of the known and survey the landscape of the unknown. The alternative is to blindly step across that chasm and hope for serendipitous results.

In structure–function analyses of extinct organisms, a course must be steered between excessive credulity
(overlooking the uncertainties in palaeontological methods and evidence) and nihilism (retreating to the defeatist view that the uncertainties are too daunting to make progress). Both extremes dance precariously on a precipice of unambition and antiscience. A major point of this review is that progress will be most sustainable where studies are particularly explicit when methods are difficult to reproduce objectively or evidence is ambiguous. In modelling and simulation, quantitative approaches maximize the benefits of explicitness. Although uncertainty and subjectivity may or may not matter, their extent and influence can be characterized with model validation and sensitivity analysis. We all can do better in maximizing the accuracy and reliability of palaeobiological models and simulations. The way forward is clear—first, contribute more data to shore up the foundations of evidence; second, be explicit and cautious with the tools used to analyse that evidence, thereby improving the quality of methods and assumptions.

This work was supported by a grant from the NERC (NE/G00711X/1) to J.R.H. It benefited from discussions with colleagues in the Structure and Motion Laboratory as well as Steve Gatesy, Emily Rayfield, Bill Sellers and Karl Bates. Julia Molnar assisted with figure 1.

John R. Hutchinson*
Structure and Motion Laboratory, Department of Veterinary Basic Sciences, The Royal Veterinary College, Hawkshead Lane, Hatfield AL9 7TA, UK

*jrhutch@rvc.ac.uk

Figure 1. *Tyrannosaurus* right hindlimb in five poses (nos 1–5); asterisks identify joints near their limits of muscle strength (modified from Gatesy et al. [16]). The graph shows those five poses and the ground reaction forces (GRFs; in body weights, BW) that limb muscles could sustain at different hip heights. Limb pose no. 5 (on right) shows three centre of mass (CoM) positions that the foot could be placed under (shaded vertical bars). The graph also shows variation in sustainable GRFs corresponding to the CoM positions (shaded curves). This sensitivity analysis reveals how limb pose and CoM influence walking/running ability.

10 Allen, V., Paxton, H. & Hutchinson, J. R. 2009 Variation in center of mass estimates for extant sauropods, and its

11 Hutchinson, J. R. & García, M. 2002 Tyrannosaurus was not a fast runner. Nature 415, 1018–1021. (doi:10.1038/4151018a)


