



## REVIEW PAPER

# Measurement accuracy and uncertainty in plant biomechanics

Nathanael Nelson<sup>1</sup>, Christopher J. Stubbs<sup>2</sup>, Ryan Larson<sup>1</sup> and Douglas D. Cook<sup>1,\*</sup> 

<sup>1</sup> Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA

<sup>2</sup> Department of Mechanical Engineering, New York University, Brooklyn, NY 11201, USA

\* Correspondence: [ddc971@byu.edu](mailto:ddc971@byu.edu)

Received 19 December 2018; Editorial decision 3 June 2019; Accepted 27 June 2019

Editor: Anja Geitmann, McGill University, Canada

## Abstract

**All scientific measurements are affected to some degree by both systematic and random errors. The quantification of these errors supports correct interpretation of data, thus supporting scientific progress. Absence of information regarding reliability and accuracy can slow scientific progress, and can lead to a reproducibility crisis. Here we consider both measurement theory and plant biomechanics literature. Drawing from measurement theory literature, we review techniques for assessing both the accuracy and uncertainty of a measurement process. In our survey of plant biomechanics literature, we found that direct assessment of measurement accuracy and uncertainty is not yet common. The advantages and disadvantages of efforts to quantify measurement accuracy and uncertainty are discussed. We conclude with recommended best practices for improving the scientific rigor in plant biomechanics through attention to the issues of measurement accuracy and uncertainty.**

**Keywords:** Best practices, error, measurement, repeatability, uncertainty, validation.

## Introduction

Although it seems paradoxical, science is based upon both skepticism and trust. As scientists, we are keenly skeptical, but can be convinced by data that are deemed to be trustworthy. In 2016, a *Nature* survey revealed that 90% of scientists felt that there existed either a slight or significant reproducibility crisis (Baker, 2015); that is, scientists cannot recreate another's experiment or results. Both engineering and biology were identified as troubled areas, with only 24% of respondents in physics and engineering reporting that they had taken steps to improve reproducibility in their lab. The scientists involved in this survey were also asked to provide suggestions for increasing scientific reproducibility. The most common suggestion was to develop more robust experimental designs. The *Nature* study was in part a reaction to the reproducibility crisis in the fields of psychology and biomedicine (Ioannidis, 2005; Collins and Tabak, 2014; Baker, 2015; Bustin and Nolan, 2016).

As a relatively small and developing field, plant biomechanics has a unique opportunity to act more quickly than older, more developed fields. A concerted effort to increase scientific rigor will have many long-term benefits (Huth, 1987), allowing our field to progress rapidly.

In the broadest terms, reproducible research can be conceptualized as arising from two factors: reliable methods and a thorough documentation of these methods (Goodman *et al.*, 2016). The current review focuses on methods for increasing the reliability of mechanical measurements and also touches on the documentation of measurement methods. It does not delve into the ways in which data can be misused, as this has been reported in a number of other previous studies, both in the popular press (Randall, 2018) and in the scientific literature (Kerr, 1998; Head *et al.*, 2015). Although this review focuses on

mechanical measurements, the principles and techniques described are readily applicable to other types of measurements.

In many fields, measurement standards and well-defined best practices help ensure that measurement results are accurate and reliable (Boone *et al.*, 1978; Atkinson and Nevill, 1998; Van Gheluwe *et al.*, 2002; Weir, 2005; Bartlett and Frost, 2008). Indeed, the purpose of measurement standards and best practices is to establish trust between the generators of information and the users of information (Taylor and Kuyatt, 1994). In fields such as physics, chemistry, and engineering, specimens are typically fabricated according to well-defined specifications and standards. Several standards organizations exist to curate these standards [e.g. ASTM International, the National Institute of Standards and Technology (NIST), the International Organization for Standardization (ISO), and the International Committee for Weights and Measures (CIPM)].

Obtaining accurate, reliable data from mechanical measurements of plant specimens is fraught with challenges (Huth, 1987). The irregular geometry of plant specimens cannot be fabricated in a controlled fashion (Sacks and Sun, 2003; Lim *et al.*, 2010). Also, even genetically identical individuals that are maintained in seemingly identical environments will eventually exhibit unique differences due to the accumulated influence of minor environmental differences (Johnson *et al.*, 1955; Allard and Bradshaw, 1964). These are just two of many challenges in this area.

Here we review several methods for quantifying the accuracy and reliability of scientific measurements. The purposes of this article are (i) to provide a brief review of methods for quantifying and reporting measurement error and uncertainty in the field of plant biomechanics; and (ii) to survey the field of plant biomechanics research in an attempt to describe the current state of our measurement and reporting practices. Although much of the article focuses on mechanical tissue measurements, the principles described herein are applicable to many other types of measurements in plant biomechanics. The review consists of four major sections: measurement theory; case studies; an examination of current practices in the plant biomechanics literature; and, finally, conclusions and recommendations.

## Measurement theory

### *Error and uncertainty*

Researchers collect data in order to test a hypothesis. However, all data are subject to both measurement error and uncertainty (Beckwith *et al.*, 1982). Error is the difference between a measured value and the specimen's true value (Croarkin *et al.*, 2006). Because all measurement methods are imperfect, the true value can never be perfectly measured (Beckwith *et al.*, 1982). Error is, therefore, an abstract idea—an extremely useful concept, but inherently unknowable. Although we cannot measure error directly, we can make informed estimates regarding error. The field of measurement theory was developed to provide methods for dealing with the uncertainty surrounding measurement error.

Errors arise from both random and systematic effects (Beckwith *et al.*, 1982). Random effects are often due to a

multitude of small factors that cannot be experimentally controlled and are difficult or impossible to identify and eliminate. Systematic effects are typically due to some kind of measurement process error such as a poorly calibrated measurement device or a sample preparation procedure that induces a systematic error. The term 'bias error' is sometimes used to refer to systematic effects because most systematic effects introduce some amount of bias into the data (Sanderson *et al.*, 2000). However, in certain circumstances, systematic errors can be introduced that may have no detectable bias; instead they may tend to increase the variance of collected data (Bland and Altman, 1996). Briefly stated, systematic effects can be attributed to some cause while random effects cannot.

### *Quantifying random effects: measurement repeatability*

Measurement repeatability is one way of quantifying measurement uncertainty. As will be shown in the 'Case studies' section, it can also provide valuable insights into the measurement process itself. Measurement repeatability has been defined as the degree of agreement between the results of successive measurements carried out under identical conditions (Taylor and Kuyatt, 1994). This concept is distinct from reproducibility: the ability to reproduce another scientist's method or results (Taylor and Kuyatt, 1994). Repeatability provides an estimate of the amount of uncertainty that should be attributed to the testing process itself.

In industrial practice, an assessment of measurement repeatability is often carried out on a number of specimens, with the assumption that the specimens are, for all intents and purposes, identical to each other. This approach is referred to as between-specimen repeatability. However, where variation between individual specimens is significant (as in plant biomechanics), within-specimen repeatability is more relevant. Within-specimen repeatability is the degree of agreement between repeated measurements that are performed with the same specimen (Atkinson, 1995; Gobbe *et al.*, 2002; D'Onofrio *et al.*, 2010; Al-Zube *et al.*, 2018). However, within-specimen repeatability cannot be assessed if the test causes damage to the specimen. In these cases, inter-specimen repeatability must be used, with the specimens chosen to be as similar to one another as possible.

To obtain a single repeatability value, one specimen is tested multiple times in relatively quick succession. When this is done, we assume that nothing about the process or specimen changes between testing cycles. For most mechanical tests, this requires the application of relatively low strain values so as to prevent tissue damage. The standard deviation of the resulting data is used to obtain a single estimate of repeatability for the test method. Repeatability is often reported as a percentage of the mean, also known as the coefficient of variation (the standard deviation divided by the mean). Because the repeatability can vary from specimen to specimen, this process can be repeated a number of times to assess the variation in repeatability values between samples. The set of repeatability values may then be averaged to obtain an aggregate repeatability value.

Mechanical measurements of plant tissues typically require the combination of multiple measurements using some kind of model. For example, the measurement of Young's modulus

using a compression test requires the measurement of the cross-sectional area, the applied force, and the resulting deformation of the specimen (Young and Budynas, 2002; Gibson, 2011). Repeatability values can be obtained for each step of the process independently, and combined using the law of propagation of uncertainty (Taylor and Kuyatt, 1994). This approach is advantageous because it allows the scientist to assess the repeatability of each step of the process independently. A disadvantage of this approach is that the law of propagation of uncertainty tends to exaggerate the predicted overall uncertainty (Hall, 2004). An alternative approach is to follow the measurement process from start to finish a number of times, each of which involves the collection of each individual measurement. All measurements for a single specimen are then combined in the normal fashion. The overall repeatability is then obtained by examining the repeatability of the final product of the measurement process. The repeatability of individual steps in the process can also be assessed from this data set.

### *Interpreting repeatability values*

Repeatability values provide information about the reliability and consistency of a measurement process. Because repeatability is highly context specific, we have chosen not to list (arbitrary) limits or ranges for 'acceptable' and 'unacceptable' values. However, several in-context repeatability values are provided in the 'Case studies' section later in this review. Under ideal conditions, individual repeatability values will be relatively small and, as a group, repeatability values will be consistent between specimens (i.e. the standard deviation of repeatability values will be small).

High repeatability values or a high variation between samples are indicators that the measurement process is adversely affected by unknown factors. For example, sample preparation processes often require subjective judgement and manual preparation. This can result in a wide variation of measured or calculated values between samples.

An assessment of each step of the measurement process as described in the previous section is an extremely beneficial means of identifying the largest contributors to measurement uncertainty. As will be shown in the 'Case studies' section, an

examination of these factors often leads to new insights for improving the measurement process.

However, repeatability values alone provide no information about the accuracy of the measurement process. In fact, it is entirely possible for a measurement process to have excellent repeatability while also being highly erroneous.

### *Accuracy and validation*

Accuracy is the degree of agreement between a measured quantity and the actual (i.e. exact, but unknown) value (BIPM *et al.*, 2012). Validation is the process of providing objective data to confirm that accuracy is appropriate for the purposes of a given study. One way to estimate measurement accuracy is through independent replication of the experiment. This is typically not practical or cost-effective. Another approach is to use a validation technique referred to as triangulation.

### *Quantifying accuracy: validation studies*

Validation by triangulation is performed by measuring the same quantity using two or more independent measurement methods (Lawlor *et al.*, 2016). The same set of specimens is used, and the average discrepancies between methods (if any) are used as an estimate of the overall error. Direct, quantitative validation can be achieved by showing that the data from two or more methods are statistically equivalent. The approach described above relies upon an assumption that different measurement methods are unlikely to be susceptible to the same sources and amounts of bias error. The statistical power is greater for paired testing, but this is not always possible. If the same set of specimens cannot be used for both tests, a single larger set of specimens can be randomly sorted into two groups.

A less rigorous form of validation is obtained when data from two or more different tests are statistically different, but the absolute numeric difference between tests is judged to be of no practical significance. For example, one type of test may be suspected of producing a slight bias in a predictable direction.

Validation efforts are informative, regardless of the outcome. If validation is achieved, we have higher confidence in the measured values as well as an estimate of the error. On the

#### **Box 1. Glossary of key terms with references**

**Accuracy:** the degree of agreement between a set of measurements and the actual value (i.e. average error) (BIPM *et al.*, 2012).

**Error:** the difference between a single measured value and the specimen's actual (i.e. exact, but unknown) value (Croarkin *et al.* (2006).

**Repeatability:** the degree of agreement between the results of successive measurements carried out under identical conditions (Taylor and Kuyatt (1994).

**Triangulation:** validation obtained by measuring the same quantity using two or more independent measurement methods. Validation is obtained when the results are sufficiently similar (Lawlor *et al.*, (2016).

**Validation:** the process of providing objective data to confirm that accuracy is appropriate for the purposes of a given study (BIPM *et al.* 2012).

other hand, if validation is not achieved, we are often led to insights that would not otherwise have been possible. Consider a situation in which two tests will be used to (hopefully) obtain validation. The researcher begins the study with an expectation that these two tests will produce results that are in agreement with each other. However, a discrepancy between the results of the two tests, along with further experimentation, may lead the researcher to the conclusion that the hierarchical nature of plant tissues causes the tissue to respond differently to these two different tests (Bidhendi and Geitmann, 2019).

The purpose of validation is to confirm that measurement processes are acceptably accurate, or to obtain an estimate of the accuracy of these methods. In many fields, researchers achieve validation through inter-study triangulation: by comparing new results with those previously reported in the literature. This approach can be problematic when dealing with biological specimens. This is because biological tissues are typically influenced by three major categories of factors: measurement techniques (including specimen preparation), genotype, and environment. Each of these categories has the potential to dramatically influence measurement results. A matching between genotype, method, and environment between studies can allow for inter-study triangulation, but, at this time, it is relatively rare to find another study with matching genotype and environmental data. When the goal is to assess the accuracy of a measurement process, major confounding factors such as genotype and environment must be controlled in some way.

The most reliable approach is for each researcher to perform his/her own validation. Intra-study triangulation allows for a direct, quantitative assessment of measurement accuracy. This information is extremely valuable, and the attainment of this data need not be overly onerous. The great advantage of intra-study triangulation is the elimination of genetic and environmental factors. With these factors controlled, only the measurement processes themselves will influence the measured results. Ideally, triangulation is performed before the primary experiment is conducted and need not consist of a full replication of the entire study. Instead, a sample of specimens can be tested using two or more methods to obtain validation, and then the most advantageous measurement method can be used for the majority of data collection.

## Case studies

In this section, a number of case studies from the authors' previous research are reviewed in order to give concrete examples of the concepts introduced above. We also include examples of mistakes that were made, but not included in the associated scientific publications. We hope that these examples will be helpful in both illustrating the use of these techniques and in aiding future researchers to avoid similar mistakes.

### Repeatability

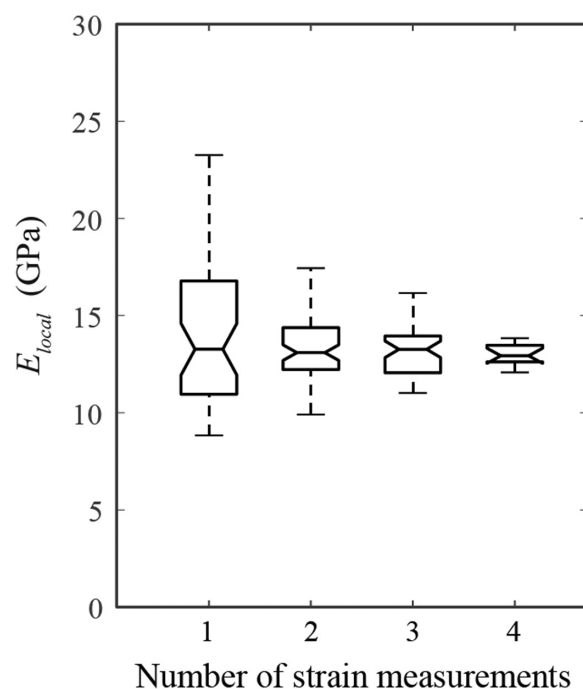
The assessment of repeatability was instrumental in revealing measurement limitations in our research group's recent study

on maize tissues (Al-Zube *et al.*, 2017). The purpose of that study was to quantify the longitudinal Young's modulus of maize rind tissues using a compression testing approach. Although best practices from ASTM standards were used (spherical platens and local measurement of strain, ASTM-E9, 2009; ASTM-D695, 2015), the initial repeatability of this test method was  $\pm 24\%$ . Based upon the authors' previous experience with this equipment and measurement techniques, this repeatability value suggested an error in the measurement process.

We began by re-examining the assumptions behind the test methodology. Two factors were identified as potential contributors. First, our initial testing method was based on the assumption that stress was evenly distributed within the specimen, but this assumption had not been confirmed. Secondly, if the cause was uneven strain distribution, the position of the specimen relative to the axis of rotation of the spherical platens could potentially exacerbate this problem.

Two small repeatability experiments were designed to test these hypotheses. The specimen position was found to have little effect on repeatability values, but the assumption of uniform strain was found to be erroneous. As shown in Fig. 1, the distribution of values obtained for a single specimen decreased dramatically as the number of strain measurements was increased (Al-Zube *et al.*, 2017). By taking multiple measurements of surface strain, average strain was assessed more accurately, leading to better repeatability values and a much improved test method. The final test method had an average repeatability value of  $\pm 5\%$  which represents a nearly 5-fold improvement over the initial repeatability value.

Another example of repeatability testing is available from a follow-up study (also from our research group). In this study,



**Fig. 1.** The distribution of single-specimen modulus of elasticity ( $E$ ) values obtained in a series of repeatability tests. Repeatability is represented in this figure by distribution widths. From Al-Zube *et al.* (2017).



we quantified the repeatability of four test types. We were surprised to learn that the three-point bending method had the lowest (best) average repeatability at 1.5%. The repeatability values for tension and compression test were found to be 1.9% and 3.8–3.9%, respectively (Al-Zube *et al.*, 2018). In the authors' experience, repeatability values of biological tissues >5% may be influenced by a methodological error, but repeatability values <5% are most likely to be due to random error. This threshold is not at all concrete, but rather depends upon the equipment, type of measurement, specimen type, etc.

It is worth noting here that one common pitfall when computing repeatability is to use a single input measurement repeatedly in a series of tests employing a single specimen. For example, when performing a compression test, the cross-sectional area of a test specimen is required to obtain the Young's modulus. It is tempting to collect this measurement once, and then perform five compression tests using the specimen. Indeed, we initially overlooked this issue. When the same cross-sectional area is used to compute the Young's modulus for each of these tests, the resulting repeatability value reflects only the repeatability of the compression test data—not the overall uncertainty of the entire measurement process. This approach is not necessarily incorrect, but any reported repeatability measurements should clearly specify the scope of the repeatability measurement process.

### Triangulation

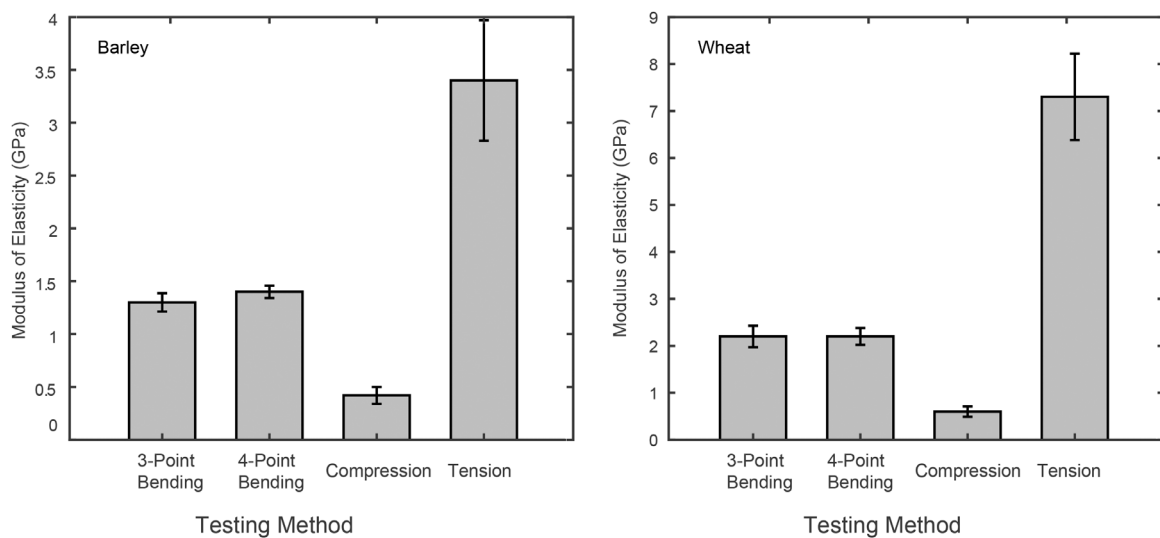
Intra-study triangulation was used in two closely related studies from our research group (Robertson *et al.*, 2014, 2015) to reveal weaknesses in bending test methods. In these studies, specimens were subjected to two types of three-point bending tests: a short-span test which has been used in numerous studies (Jenkins, 1930; Hondroyianni *et al.*, 2000; Tongdi *et al.*, 2011; Hu *et al.*, 2013; Gomez *et al.*, 2017, 2018), and a revised three-point bending test protocol that used a longer test

span (Robertson *et al.*, 2014). Short-span tests were shown to induce premature failure of the stalk, producing bending test results that ranged from 1/2 to 1/4 of the values obtained when using the long-span test protocol. By using two types of test along with engineering analysis, we were able to demonstrate that both the span and the placement of the loading anvil contributed to the erroneous results obtained under the short-span method.

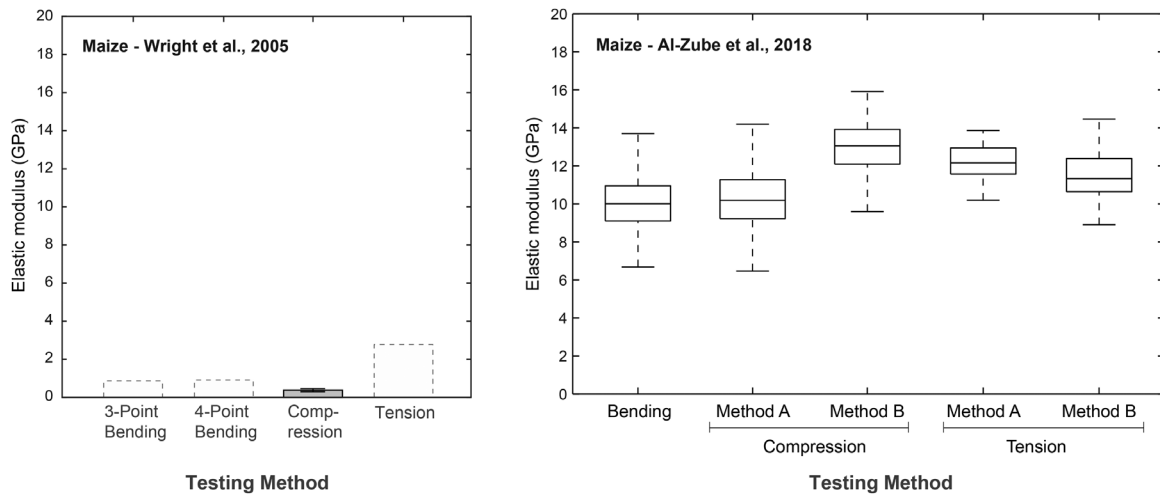
Another example of discrepancies between test methods is available in a study that examined differences in the Young's modulus of wheat, barley, and maize tissues (Wright *et al.*, 2005). This study was notable in that four different test methods were used: three-point bending, four-point bending, compression, and tension tests. The bending test results were in relatively close agreement, but compression and tension test results were quite different. Figure 2 depicts representative box plots based on the data reported by Wright *et al.* (2005). On average, the tension test results were 2.2 times higher than the bending results, and 6.3 times higher than the compression test results. Compression results were approximately half the value of bending results. The striking disagreement between test results indicates potential problems with the testing methods.

A similar study was conducted by Al-Zube *et al.* (2018). This study also used three-point bending, two types of compression test, and a tension test to assess the Young's modulus of maize. Although there were discrepancies between the results from these testing methods, the degrees of discrepancy were relatively moderate. For example, the largest discrepancy was between compression and bending results: 12.87 GPa and 10.1 GPa (a 27% discrepancy).

The average value for the Young's modulus of maize rind reported by Al-Zube *et al.* was 11.4 GPa, but the average value reported by Wright *et al.* was 0.38 GPa. This represents a discrepancy of >2900% (see Fig. 3). Although this discrepancy is notable, it is difficult to determine how much of the discrepancy should be attributed to testing method, genotype,



**Fig. 2.** Data from Wright *et al.* (2005), depicting differing modulus of elasticity values measured for barley and wheat, as assessed using four different methods. Bars represent mean values. As per the data in Wright *et al.*, whiskers represent the standard error of the mean.



**Fig. 3.** Reported modulus of elasticity values for maize rind tissues. Both charts use the same scaling to facilitate comparisons. Left: data from four varieties of maize as reported in [Wright \*et al.\* \(2005\)](#). In this study, maize tissues were measured only in compression. The solid bar represents these reported data while the dashed bars represent estimates that were imputed based on the data trends shown in [Fig. 2](#). Right: data from five varieties of maize reported in [Al-Zube \*et al.\* \(2018\)](#).

or environment. One difference between these studies is that Al-Zube *et al.* provided intra-study validation, whereas the measurements in Wright *et al.* were inconsistent.

Another instance of intra-study validation involved the measurement of turgor pressure. Tomato cell turgor pressure was measured using two methods: a pressure probe and force-sensing micromanipulation ([Wang \*et al.\*, 2006](#)). The reported values were found to be 3.3 bar and 3.2 bar, respectively. Because these values were statistically insignificant, the accuracy of each method could be reported as better than  $\pm 5\%$ .

## Literature meta-analysis

In preparing this article, the authors carefully examined 40 papers from the plant biomechanics literature that reported the measurement of mechanical tissue properties. These papers are indicated in the reference list by a dagger symbol. As a disclaimer, several of these papers ([Robertson \*et al.\*, 2014, 2015](#); [Al-Zube \*et al.\*, 2017, 2018](#)) originated from the last author's research group (The Crop Biomechanics Laboratory).

Two of these papers (5% of the sample) provided measurement repeatability values for their measurement process ([Al-Zube \*et al.\*, 2017, 2018](#)). Validation was addressed in some form in 15 of 40 studies (35%) ([Mattheck, 1995](#); [Moulia and Fournier, 1997](#); [Henry and Thomas, 2002](#); [Ryden \*et al.\*, 2003](#); [Wright \*et al.\*, 2005](#); [Green \*et al.\*, 2006](#); [Wang \*et al.\*, 2006](#); [Onoda \*et al.\*, 2010](#); [Masselter \*et al.\*, 2011](#); [Milani \*et al.\*, 2013](#); [Sharma \*et al.\*, 2013](#); [Robertson \*et al.\*, 2014, 2015](#); [Leblicq \*et al.\*, 2015](#); [Al-Zube \*et al.\*, 2018](#)). Among these 15 studies, intra-study validation was discussed in four studies ([Green, 2006](#); [Wang \*et al.\*, 2006](#); [Sharma, 2013](#); [Al-Zube \*et al.\*, 2018](#)).

Based on this survey of the literature, it seems that a minority of studies from the plant biomechanics literature report on measurement repeatability or measurement validation. Also, the authors' research is not an exception to this trend: the majority of our papers have not included repeatability or validation data. However, it is certainly not the case that the

plant biomechanics community has simply ignored the issue of validation altogether. In reviewing the literature, we found that validation was most often addressed through two informal approaches: inter-study comparisons and justification by previously reported method.

### Inter-study comparisons

Inter-study triangulation is an important part of science. The purpose of these comparisons is to place each study in the broader scientific context, providing comparisons between data, trends, results, and conclusions of other studies ([Glass, 1976](#)). Inter-study comparisons are regularly performed in the plant biomechanics literature. For example, [Kokubo \*et al.\* \(1989\)](#) studied brittle barley culms and used existing research on non-brittle strains of barley as well as maize as points of comparison for their results. [Leblicq \*et al.\* \(2015\)](#) used bending tests on wheat and barley stems, comparing the results with other values found in the literature. Many other examples can be found in the literature ([Rüggeberg \*et al.\*, 2010](#); [Masselter \*et al.\*, 2011](#); [Al-Zube \*et al.\*, 2017, 2018](#)). Although inter-study triangulation is useful for many purposes, it should not be misinterpreted as measurement validation.

### Justification by prior method

Another common practice is the justification of a measurement technique via citation of a prior study that used the same method. This argument is based upon two important assumptions: (i) the accuracy of the cited method has been validated previously; and (ii) the method as described by the original authors has been followed precisely in subsequent studies.

As an exercise, we examined a set of 12 studies that used justification by method. For each study, we followed the citation trail back into the literature to determine if any prior study had reported direct, quantitative validation. None of the selected citations could be traced back to a fully validated measurement process.

Based on our experience and observations, it seems that published studies may sometimes be misinterpreted as ‘validation by peer review’. Subsequent researchers may therefore place undue trust in these prior studies, which causes the methods to be used repeatedly, but each time without validation. This pattern has been observed by the authors in the case of flawed bending methods (Evans *et al.*, 2007; Robertson *et al.*, 2014, 2015), omission of test set-up justification (Crook and Ennos, 1996; Abasolo *et al.*, 2009), and misapplication of measurement standards (Ampofo *et al.*, 2013). Similar patterns have been observed in other fields of biomechanics (Cook, 2009; Alipour *et al.*, 2011). Publication, even in a high-quality journal, does not guarantee freedom from measurement error.

## Discussion

### *A cognitive bias towards trusting data?*

Data are at the heart of the scientific endeavor. It is the raw material for new scientific insights and testing scientific hypotheses. However, even data that are carefully collected by experienced researchers can be erroneous.

Based on the authors’ experience and review of the literature, the community regularly asks important questions such as ‘Do the methods seem appropriate for the stated purpose?’ ‘Have these methods been used before in the literature?’ ‘Do the results agree with some other data in the literature?’

There exists an opportunity to improve the review process by asking additional questions such as, ‘Is there direct, quantitative evidence that these measurements are accurate?’ and ‘Is there quantitative evidence that the measurement uncertainty has been assessed?’ These questions have the potential to bring a healthy skepticism to the evaluation of data, a skepticism that acknowledges the difficulty of collecting accurate data in the field of plant biomechanics.

We also wish to address a counter-argument. Namely, is it necessary to rigorously validate every type of measurement? In our opinion, skepticism should be inversely proportional to the proven reliability of the measurement process. For example, it is probably not necessary to perform intra-study validation for an electronic mass balance purchased from a reliable manufacturer. This is because the mass balance represents a mature technology, and research-grade devices are typically certified to provide measurements with stated values of accuracy and uncertainty. In contrast, some measurements processes performed in the field of plant biomechanics represent emerging methods and technologies that have yet to be assessed in a similar fashion.

An argument could also be made that an assessment of measurement accuracy and uncertainty is not necessary when performing comparative studies. For example, comparisons between mutant and wild-type varieties are common in the plant biomechanics literature (Ryden *et al.*, 2003; Paul-Victor and Rowe, 2011; Park and Cosgrove, 2012). Studies of this kind are primarily interested in differences between the two varieties rather than precise physical values. However, if a measurement technique is inaccurate, there is the possibility that

the intended quantity and the measured quantity differ in important ways. For example, if an effort to assess the shear modulus of some tissue is in error by an unknown amount, should this measurement be interpreted to represent the behavior of the actual shear modulus? Without a validated measurement of shear modulus, we have no way of knowing if an erroneous shear modulus measurement follows the same trends as the actual shear modulus; there are many layers of uncertainty surrounding an unvalidated measurement. Although these data may be intended for comparative purposes, they may be used by future researchers for non-comparative purposes (e.g. computational modeling). The future misinterpretation of data is problematic even if inaccuracies in measurement do not directly affect the accuracy of intended comparisons.

### *Why bother to assess repeatability and accuracy?*

Some researchers may feel that the effort required to assess repeatability and validation outweighs the benefits. We frankly admit that these activities require additional effort. Here are several reasons why we think that the additional effort is justified:

First, repeatability can often be accomplished with a small amount of additional experimentation. A reasonable estimate of repeatability can be obtained from 5–10 repeated tests on 5–10 samples. Secondly, repeatability data often reveal weaknesses in a measurement process. In addition, improvements to the repeatability of a measurement process will have a direct effect on the statistical power of the test (i.e. all other factors held constant, lowering random error will increase statistical power). Thirdly, reporting repeatability values is extremely useful for future researchers, both as a metric for judging which method to use in an upcoming experiment, and as a diagnostic tool to confirm that a previously described process has been correctly implemented.

The assessment of measurement accuracy requires more investment as compared with the assessment of repeatability. However, there are distinct advantages to doing so. The primary reason to perform validation is to gain a quantitative understanding of the accuracy of one’s own measurement process. After all, if the measurement is erroneous, fallacious scientific conclusions may be reached. Secondly, even when validation is not achieved, the results can provide valuable insights. Thirdly, validation of this type accelerates research progress by contributing to a scientific literature that is based upon quantified accuracy instead of unknown amounts of accuracy in each study. This is important because incorrect results, once published, can be difficult to overturn (Lehrer, 2010). Finally, as validation and uncertainty quantification are practised, they quickly become more natural and less time-consuming. Thus, the effort to incorporate these techniques has a cost which diminishes over time, but the advantages to this approach remain constant.

### *Reporting reproducibility, accuracy, and methods*

It is commonly understood that a detailed description of one’s experimental methods is required for subsequent researchers to build on and perhaps replicate these studies (Baker, 2015).

However, it has also been shown in numerous recent studies that scientific publications often do not contain the level of detail that would be required for actual replication (Cook, 2009; Alipour *et al.*, 2011; Baker, 2015).

The plant biomechanics literature appears to be consistent with these trends from other fields. Specimen selection and preparation are typically well documented, but equipment information is sometimes incomplete. Based on our reading, finer details of the test procedure—such as are pre-conditioning, methods of fixation, applied strain rate, and load cycling—are sometimes not reported in sufficient detail. In addition, the results are sometimes reported only as derived quantities (mean, standard deviation, modulus, etc.) instead of as raw data.

In previous years, page limits imposed by journals may have inadvertently deterred researchers from fully describing their measurement methods and results. But most journals now allow authors to upload supplementary data files and descriptions of methods. With this option, there exists a new opportunity for more complete explanations of measurement techniques, including raw measurement data, valuable tips and pitfalls that the authors discovered in the course of research, etc. Both authors and reviewers can help improve this aspect of plant biomechanics research by including/requesting this information as part of the peer-review process.

### Further reading

This information presented herein contains only a small portion of the information in regards to the topic of experimental testing, repeatability, reliability, and validation. The following sources are recommended for further reading on these topics (Taylor and Kuyatt, 1994; Bell, 1999; Bartlett *et al.*, 2008; BIPM *et al.*, 2012).

## Conclusions

Authors and reviewers in the field of plant biomechanics can increasingly seek to apply principles of measurement theory in their work. In this review, we have discussed three best practices associated with data collection and reporting in plant biomechanics (i) the assessment of measurement repeatability; (ii) the use of intra-study triangulation to estimate the accuracy of measurement techniques; and (iii) detailed reporting of both measurement methods and the associated repeatability and triangulation data. Journal editors and reviewers can play a pivotal role by beginning to expect authors to use these best practices. Specifically, editors and reviewers can emphasize questions such as ‘Is there direct, quantitative evidence that the reported measurements are accurate?’, ‘Is there quantitative evidence that measurement repeatability has been assessed?’, and/or ‘Are the test methods described in sufficient detail for replication?’

The adoption of these practices has advantages for both individual scientists and the broader community. Specific advantages to the individual scientists include the following: (i) identification of flaws or limitations of a measurement processes; (ii) increased statistical power; and (iii) insurance against potentially erroneous conclusions that could be made when using unvalidated data

These best practices also have benefits to the broader scientific community, including the following: (i) increased confidence in reported results; (ii) reported data can be used as quantitative benchmarks for future researchers; (iii) encouragement of future use or replication of reported methods; and (iv) accelerated research progress.

As the field of plant biomechanics seeks to implement measurement best practices, the accuracy of and confidence in our scientific findings will increase. At the same time, studies that utilize these best practices will be more useful and informative to future researchers. Overall, the application of these methods will serve to accelerate research progress in the field of plant biomechanics

## Acknowledgements

The authors thank Dr Anja Geitmann (Faculty of Agriculture and Environmental Science, McGill University) for inviting this contribution to a special issue honoring the scientific contributions of Dr Hanns-Christof Spatz. They also thank two anonymous reviewers for their very helpful comments and suggestions.

## References

- <sup>†</sup>Abasolo W, Eder M, Yamauchi K, *et al.* 2009. Pectin may hinder the unfolding of xyloglucan chains during cell deformation: implications of the mechanical performance of Arabidopsis hypocotyls with pectin alterations. *Molecular Plant* **2**, 990–999.
- Alipour F, Brucker C, Cook DD, Gommel A, Kaltenbacher M, Mattheus W, Mongeau L, Nauman E, Schwarze R, Tokuda I. 2011. Mathematical models and numerical schemes for the simulation of human phonation. *Current Bioinformatics* **6**, 323–343.
- Allard RW, Bradshaw AD. 1964. Implications of genotype–environmental interactions in applied plant breeding 1. *Crop Science* **4**, 503–508.
- <sup>†</sup>Alméras T, Derycke M, Jaouen G, Beauchêne J, Fournier M. 2009. Functional diversity in gravitropic reaction among tropical seedlings in relation to ecological and developmental traits. *Journal of Experimental Botany* **60**, 4397–4410.
- <sup>†</sup>Al-Zube LA, Robertson DJ, Edwards JN, Sun W, Cook DD. 2017. Measuring the compressive modulus of elasticity of pith-filled plant stems. *Plant Methods* **13**, 99.
- <sup>††</sup>Al-Zube L, Sun W, Robertson D, Cook D. 2018. The elastic modulus for maize stems. *Plant Methods* **14**, 11.
- <sup>†</sup>Ampofo J, Ofori E, Ibrahim AA. 2013. Mechanical properties of plantain pseudostem and implications for susceptibility to lodging. *Journal of Agriculture and Environment for International Development* **107**, 3–11.
- ASTM-D695. 2015. Standard test method for compressive properties of rigid plastics. West Conshohocken, PA: ASTM International. <http://www.astm.org>. Accessed 1 August 2017.
- ASTM-E9. 2009. Standard test methods of compression testing of metallic materials at room temperature. West Conshohocken, PA: ASTM International.
- Atkinson G. 1995. A comparison of statistical methods for assessing measurement repeatability in ergonomics research. In: Atkinson G, Reilly T, eds. *Sport, leisure and ergonomics*. London: Chapman & Hall, 218–222.
- Atkinson G, Nevill AM. 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine* **26**, 217–238.
- Baker CJ. 1995. The development of a theoretical model for the windthrow of plants. *Journal of Theoretical Biology* **175**, 355–372.
- Baker M. 2015. Over half of psychology studies fail reproducibility test. *Nature News* doi:10.1038/nature.2015.18248.
- Bartlett JW, Frost C. 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics & Gynecology* **31**, 466–475.



- Beckwith TG, Buck NL, Marangoni RD. 1982. Mechanical measurements. Boston, MA: Addison-Wesley Publishing Company.
- Bell S. 1999. A beginner's guide to uncertainty of measurement. *Measurement Good Practice Guide* **11**, 1.
- Bidhendi AJ, Geitmann A. 2019. Methods to quantify primary cell wall mechanics. *Journal of Experimental Botany* **70**, 3615–3648.
- BIPM, IEC, IFCC, ILAC, ISO, IUPAP, OMIL. 2012. The international vocabulary of metrology—basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology 200, 2012. [https://www.bipm.org/utis/common/documents/jcgm/JCGM\\_200\\_2012.pdf](https://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf)
- Bland JM, Altman DG. 1996. Statistics notes: measurement error. *BMJ* **312**, 1654.
- Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L. 1978. Reliability of goniometric measurements. *Physical Therapy* **58**, 1355–1360.
- <sup>†</sup>Burgert I, Fröhmann K, Keckes J, Fratzi P, Stanzl-Tschegg SE. 2003. Microtensile testing of wood fibers combined with video extensometry for efficient strain detection. *Holzforschung* **57**, 661–664.
- <sup>†</sup>Burgert I, Keplinger T. 2013. Plant micro- and nanomechanics: experimental techniques for plant cell-wall analysis. *Journal of Experimental Botany* **64**, 4635–4649.
- Bustin SA, Nolan T. 2016. Improving the reliability of peer-reviewed publications: we are all in it together. *Biomolecular Detection and Quantification* **7**, A1–A5.
- Collins FS, Tabak LA. 2014. NIH plans to enhance reproducibility. *Nature* **505**, 612–613.
- Cook DD. 2009. Systematic structural analysis of human vocal fold models. PhD Thesis, Purdue University.
- Croarkin C, Tobias P, Filliben JJ, Hembree B, Guthrie W. 2006. NIST/SEMATECH e-handbook of statistical methods. NIST/SEMATECH, July. <http://www.itl.nist.gov/div898/handbook>.
- <sup>†</sup>Crook MJ, Ennos AR. 1996. Mechanical differences between free-standing and supported wheat plants, *Triticum aestivum* L. *Annals of Botany* **77**, 197–202.
- D'Onofrio M, Gallotti A, Mucelli RP. 2010. Tissue quantification with acoustic radiation force impulse imaging: measurement repeatability and normal values in the healthy liver. *American Journal of Roentgenology* **195**, 132–136.
- <sup>†</sup>Der Loughian C, Tadrist L, Allain J-M, Diener J, Moulia B, de Langre E. 2014. Measuring local and global vibration modes in model plants. *Comptes Rendus Mécanique* **342**, 1–7.
- <sup>†</sup>Evans LS, Kahn-Jetter Z, Marks C, Harmony KR. 2007. Mechanical properties and anatomical components of stems of 42 grass species. *Journal of the Torrey Botanical Society* **134**, 458–467.
- <sup>†</sup>Faisal TR, Rey AD, Pasini D. 2013. A multiscale mechanical model for plant tissue stiffness. *Polymers* **5**, 730–750.
- <sup>†</sup>Fournier M, Dlouhá J, Jaouen G, Almeras T. 2013. Integrative biomechanics for tree ecology: beyond wood density and strength. *Journal of Experimental Botany* **64**, 4793–4815.
- Gibson RF. 2011. Principles of composite material mechanics. Boca Raton, FL: CRC Press.
- Glass GV. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* **5**, 3–8.
- Gobbe M, Guillon M, Maissa C. 2002. Measurement repeatability of corneal aberrations. *Journal of Refractive Surgery* **18**, S567–S571.
- Gomez FE, Muliana AH, Niklas KJ, Rooney WL. 2017. Identifying morphological and mechanical traits associated with stem lodging in bioenergy Sorghum (*Sorghum bicolor*). *Bioenergy Research* **10**, 635–647.
- Gomez FE, Muliana AH, Rooney WL. 2018. Predicting stem strength in diverse bioenergy sorghum genotypes. *Crop Science* **58**, 738–751.
- <sup>†</sup>Goodman AM, Ennos AR, Booth I. 2002. A mechanical study of retting in glyphosate treated flax stems (*Linum usitatissimum*). *Industrial Crops and Products* **15**, 169–177.
- Goodman SN, Fanelli D, Ioannidis JP. 2016. What does research reproducibility mean? *Science Translational Medicine* **8**, 341ps12.
- <sup>††</sup>Green DW, Gorman TM, Evans JW, Murphy JF. 2006. Mechanical grading of round timber beams. *Journal of Materials in Civil Engineering* **18**, 1–10.
- Hall BD. 2004. On the propagation of uncertainty in complex-valued quantities. *Metrologia* **41**, 173.
- Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. 2015. The extent and consequences of p-hacking in science. *PLoS Biology* **13**, e1002106.
- <sup>†</sup>Henry HA, Thomas SC. 2002. Interactive effects of lateral shade and wind on stem allometry, biomass allocation, and mechanical stability in *Abutilon theophrasti* (Malvaceae). *American Journal of Botany* **89**, 1609–1615.
- <sup>†</sup>Hofstetter K, Gamstedt EK. 2009. Hierarchical modelling of microstructural effects on mechanical properties of wood. A review COST Action E35 2004–2008: wood machining—micromechanics and fracture. *Holzforschung* **63**, 130–138.
- Hondroyianni E, Papakosta DK, Gagianas AA, Tsatsarelis KA. 2000. Corn stalk traits related to lodging resistance in two soils of differing salinity. *Maydica* **45**, 125–133.
- Hu H, Liu W, Fu Z, Homann L, Technow F, Wang H, Song C, Li S, Melchinger AE, Chen S. 2013. QTL mapping of stalk bending strength in a recombinant inbred line maize population. *Theoretical and Applied Genetics* **126**, 2257–2266.
- Huth EJ. 1987. Needed: review articles with more scientific rigor. *Annals of Internal Medicine* **106**, 470–471.
- Ioannidis JP. 2005. Why most published research findings are false. *PLoS Medicine* **2**, e124.
- Jenkins MT. 1930. Experiments on stiffness of stalk. Iowa Agricultural Experiment Station Research Bulletin **1930**, 51.
- Johnson HW, Robinson HF, Comstock R. 1955. Estimates of genetic and environmental variability in soybeans 1. *Agronomy Journal* **47**, 314–318.
- Kerr NL. 1998. HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review* **2**, 196–217.
- <sup>†</sup>Köhler L, Spatz HC. 2002. Micromechanics of plant tissues beyond the linear-elastic range. *Planta* **215**, 33–40.
- <sup>†</sup>Kokubo A, Kuraishi S, Sakurai N. 1989. Culm strength of barley: correlation among maximum bending stress, cell wall dimensions, and cellulose content. *Plant Physiology* **91**, 876–882.
- <sup>†</sup>Kutschera U, Schopfer P. 1986. Effect of auxin and abscisic acid on cell wall extensibility in maize coleoptiles. *Planta* **167**, 527–535.
- Lawlor DA, Tilling K, Davey Smith G. 2016. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* **45**, 1866–1886.
- <sup>††</sup>Leblcq T, Vanmaercke S, Ramon H, Saeys W. 2015. Mechanical analysis of the bending behaviour of plant stems. *Biosystems Engineering* **129**, 87–99.
- Lehrer J. 2010. The truth wears off. *The New Yorker* **13**, 229.
- <sup>†</sup>Lemloh M-L, Pohl A, Weber E, Zeiger M, Bauer P, Weiss IM, Schneider AS. 2014. Structure–property relationships in mechanically stimulated *Sorghum bicolor* stalks. *Bioinspired Materials* **1**, 1–11.
- <sup>†</sup>Leroux O. 2012. Collenchyma: a versatile mechanical tissue with dynamic cell walls. *Annals of Botany* **110**, 1083–1098.
- Lim MD, Dickherber A, Compton CC. 2010. Before you analyze a human specimen, think quality, variability, and bias. Washington, DC: ACS Publications.
- <sup>†</sup>Liu L, Zhang W, Li L, Zhu X, Liu J, Wang X, Song Z, Xu H, Wang Z. 2018. Biomechanical measurement and analysis of colchicine-induced effects on cells by nanoindentation using an atomic force microscope. *Journal of Biomechanics* **67**, 84–90.
- <sup>†</sup>Ludwig F, de Bruyn G, Thielen M, Speck T. 2009. Plant stems as building material for living plant constructions. Sixth Plant Biomechanics Conference. 16–21 November 2009, Cayenne, French Guyana, France. 398.
- <sup>††</sup>Masselter T, Eckert S, Speck T. 2011. Functional morphology, biomechanics and biomimetic potential of stem–branch connections in *Dracaena reflexa* and *Freyinetia insignis*. *Beilstein Journal of Nanotechnology* **2**, 173–185.
- Mattheck C. 1995. Biomechanical optimum in woody stems. In: Gartner BL, ed. *Plant stems*. Amsterdam: Elsevier, 75–90.
- <sup>††</sup>Milani P, Braybrook SA, Boudaoud A. 2013. Shrinking the hammer: micromechanical approaches to morphogenesis. *Journal of Experimental Botany* **64**, 4651–4662.
- <sup>††</sup>Moulia B, Fournier M. 1997. Mechanics of the maize leaf: a composite beam model of the midrib. *Journal of Materials Science* **32**, 2771–2780.
- <sup>††</sup>Onoda Y, Richards AE, Westoby M. 2010. The relationship between stem biomechanics and wood density is modified by rainfall in 32 Australian woody plant species. *New Phytologist* **185**, 493–501.

- †Park YB, Cosgrove DJ.** 2012. A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. *Plant Physiology* **158**, 1933–1943.
- †Paul-Victor C, Rowe N.** 2011. Effect of mechanical perturbation on the biomechanics, primary growth and secondary tissue development of inflorescence stems of *Arabidopsis thaliana*. *Annals of Botany* **107**, 209–218.
- Randall D.** 2018. Opinion | a Cornell scientist's downfall. *Wall Street Journal*.
- ††Robertson D, Smith S, Gardunia B, Cook D.** 2014. An improved method for accurate phenotyping of corn stalk strength. *Crop Science* **54**, 2038–2044.
- ††Robertson DJ, Julias M, Gardunia BW, Barten T, Cook DD.** 2015. Corn stalk lodging: a forensic engineering approach provides insights into failure patterns and mechanisms. *Crop Science* **55**, 2833–2841.
- †Rodriguez M, Ploquin S, Moulia B, de Langre E.** 2012. The multimodal dynamics of a walnut tree: experiments and models. *Journal of Applied Mechanics* **79**, 044505.
- †Rüggeberg M, Burgert I, Speck T.** 2010. Structural and mechanical design of tissue interfaces in the giant reed *Arundo donax*. *Journal of the Royal Society, Interface* **7**, 499–506.
- ††Ryden P, Sugimoto-Shirasu K, Smith AC, Findlay K, Reiter WD, McCann MC.** 2003. Tensile properties of Arabidopsis cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II–borate complexes. *Plant Physiology* **132**, 1033–1040.
- Sacks MS, Sun W.** 2003. Multiaxial mechanical behavior of biological materials. *Annual Review of Biomedical Engineering* **5**, 251–284.
- Sanderson MJ, Wojciechowski MF, Hu JM, Khan TS, Brady SG.** 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* **17**, 782–797.
- ††Sharma B, Harries KA, Ghavami K.** 2013. Methods of determining transverse mechanical properties of full-culm bamboo. *Construction and Building Materials* **38**, 627–637.
- Taylor BN, Kuyatt CE.** 1994. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Note 1297. Washington, DC: National Institute of Standards and Technology.
- Tongdi Q, Yaoming L, Jin C.** 2011. Experimental study on flexural mechanical properties of corn stalks. 2011 International Conference on New Technology of Agricultural. IEEE, 130–134.
- Van Gheluwe B, Kirby KA, Roosen P, Phillips RD.** 2002. Reliability and accuracy of biomechanical measurements of the lower extremities. *Journal of the American Podiatric Medical Association* **92**, 317–326.
- ††Wang L, Hukin D, Pritchard J, Thomas C.** 2006. Comparison of plant cell turgor pressure measurement by pressure probe and micromanipulation. *Biotechnology Letters* **28**, 1147–1150.
- Weir JP.** 2005. Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research* **19**, 231–240.
- ††Wright CT, Pryfogle PA, Stevens NA, Steffler ED, Hess JR, Ulrich TH.** 2005. Biomechanics of wheat/barley straw and corn stover. *Applied Biochemistry and Biotechnology* **121**, 5–19.
- Young WC, Budynas RG.** 2002. Roark's formulas for stress and strain. New York: McGraw-Hill.
- †Zhang LX, Yang ZP, Zhang Q, Guo HL.** 2016. Tensile properties of maize stalk rind. *Bioresources* **11**, 6151–6161.