Justin M Mcgrath[1,2], Matthew H Siebers[1,2], Peng Fu[3,4], Stephen P Long[4], and Carl J Bernacchi[1,2,4]

[1]Department of Plant Biology, University of Illinois at Urbana-Champaign
[2]USDA-ARS, Global Change and Photosynthesis Research Unit
[3]Geospatial Data Science, Harrisburg University
[4]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign

November 1, 2022

NAPPN Annual Conference Abstract: Well-found statistical tests for method comparisons instead of using Pearson's correlation coefficient

Justin M McGrath[1,2], *Matthew H. Siebers*[1,2], *Peng Fu*[3,4], *Stephen P. Long*[4], *Carl J. Bernacchi*[1,2,4]

[1] *Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

[2]*USDA-ARS, Global Change and Photosynthesis Research Unit, Urbana IL, USA*

[3]*Geospatial Data Science, Harrisburg University, Harrisburg, PA, USA*

[4]*Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

ORCiD: 0000-0002-7025-3906

Keywords: method comparison, Pearson, variance, bias

BodyText: A significant portion of plant phenotyping research involves development of new instruments and methodology. Thus, a common experiment is to compare new methods to established ones in order to assess the suitability of the new method. Pearson's correlation coefficient, *r*, is commonly calculated from the correlation between measurements of the two methods on the same subjects, and it is interpreted to assess whether the new method is a suitable replacement for the established one. However, *r* (and in this context *R*, and R^2) is not an appropriate statistic for this purpose, and it provides no meaningful information for comparing quality of methods. This is well established, and other alternatives are known. Here we present quantification and statistical tests of bias and variances of two methods that provide a well-founded approach to method comparison. Comparing newly developed methods to measure height and leaf area index (LAI) using lidar, we find that lidar estimates of height are more precise than established methods and lidar estimates of LAI are equivalent or slightly worse. Using *r* alone it is not possible to make these interpretations. These sorts of statements are possible due to clear, objective approaches to method comparison, which should be the standard for assessing new phenotyping methods.