frontiers in Plant Science

TYPE General Commentary
PUBLISHED 14 April 2026
DOI 10.3389/fpls.2026.1776560

Check for updates

OPEN ACCESS

EDITED BY
Pe Rajasekharan,
Indian Institute of Horticultural Research (ICAR), India

REVIEWED BY
Yue Li,
Liaoning University, China
Pablo Tejero,
Spanish National Research Council (CSIC), Spain

*CORRESPONDENCE
Kenneth J. Elgersma
kenneth.elgersma@uni.edu

RECEIVED 28 December 2025
REVISED 24 March 2026
ACCEPTED 31 March 2026
PUBLISHED 14 April 2026

CITATION
Elgersma KJ, Aird S, Martina JP and Goldberg DE (2026) Commentary: Field-based measurement tools to distinguish clonal *Typha* taxa and estimate biomass: a resource for conservation and restoration.
Front. Plant Sci. 17:1776560.
doi: 10.3389/fpls.2026.1776560

COPYRIGHT
© 2026 Elgersma, Aird, Martina and Goldberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Commentary: Field-based measurement tools to distinguish clonal *Typha* taxa and estimate biomass: a resource for conservation and restoration

Kenneth J. Elgersma[1*], Sydney Aird[1], Jason P. Martina[2] and Deborah E. Goldberg[3,4]

[1]Department of Biology, University of Northern Iowa, Cedar Falls, IA, United States, [2]Department of Biology, Texas State University, San Marcos, TX, United States, [3]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, United States, [4]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, United States

KEYWORDS

clonal plants, experimental design, field identification, morphometrics, spatial dependence, *Typha* × *glauca*, *Typha latifolia*, *Typha angustifolia*

A Commentary on:

Field-based measurement tools to distinguish clonal Typha taxa and estimate biomass: a resource for conservation and restoration

By Ohsowski BM, Redding C, Geddes P and Lishawa SC (2024). *Front. Plant Sci.* 15:1348144. doi: 10.3389/fpls.2024.1348144

Introduction

Hybridization between North American native and non-native cattail species (*Typha latifolia* and *T. angustifolia*) has led to widespread dominance of *T.* × *glauca*, making the morphological identification of *Typha* critical for conservation and management. Several studies have identified morphological distinctions corresponding to molecular genetic identities, with varied success (Kuehn and White, 1999; Snow et al., 2010; Geddes et al., 2021; Calvo et al., 2025). Ohsowski et al. (2024) used a similar approach but added leaf count - a previously unexplored trait - that they suggest holds strong predictive power of *Typha* genetic ID. While we agree this finding warrants further investigation, we raise questions about the sampling design and analytical methods used in this and similar previous studies (including our own; Calvo et al., 2025). Specifically, we address limitations related to sample size and spatial dependence that influence generalizability, and we offer recommendations for future collaborative research to overcome these limitations.

Sample size and independence

The small sample size (SSS) problem (Sharma and Paliwal, 2015) is a long-standing problem in Linear Discriminant Analysis (LDA) methods used by Ohsowski et al. (2024) and similar efforts. When the number of potential predictors is large and sample size is limited, "useful" predictors have poor out-of-sample predictive power. In general, for accuracy estimations to be "reasonably close" to reality (Huberty and Olejnik, 2007 p. 309), the *smallest* group's sample size (n) should be 3–10 times greater than the number of potential predictors (k). Ohsowski et al. start with k=10, so n should be ≥30 for the smallest group–approximately 10-fold greater than their smallest groups (3 for advanced generation hybrids (AGHs); 5 for *T. latifolia*). The risk of SSS problems is amplified by their backward model selection procedure (RFE), which retains only the best predictors (Yates et al., 2023). This optimization process raises the concern that "optimization capitalizes on chance" (Mosteller and Tukey, 1977, p. 769).

To manage SSS problems, Ohsowski et al. use the common approach of cross-fold validation with training and testing datasets. Theoretically, this should safeguard against SSS problems when testing and training data are independent. Unfortunately, that critical assumption was not met due to the clonal nature of *Typha*. The 33 samples Ohsowski et al. used for taxonomic prediction came from 12 transects in 7 wetlands. This spatial structuring resulted in identical genetic IDs within transects for all but 3 samples. All other measured variables were also strongly correlated within transects; we found that "transect" explains most of the variance (typically ~75% and up to 97% of variance) for each. This strong spatial autocorrelation lowers effective sample size below the nominal n=33 and violates the assumption of independence between testing and training data, thereby inflating out-of-sample prediction accuracy (Xu and Goodacre, 2018).

Assessment with null models

We constructed two null models to assess the impacts of SSS and nonindependence: one assuming sample independence and one replicating the within-transect correlations observed in the original data. In both cases, we replaced the Ohsowski et al. predictor values with randomly-generated numbers (i.e., "noise") and used their methods (LDA via RFE with cross-fold validation) to classify the data. We recorded the resulting prediction accuracy and repeated this 1000 times to generate null distributions (Figure 1). Assuming sample independence, the observed accuracy of Ohsowski et al. was significantly better than chance ($p = 0.031$; Figure 1A). Although LDA via RFE correctly classified random "noise" 60.1% of the time on average, it rarely outperformed Ohsowski et al. However, when spatial dependence was preserved, the Ohsowski et al. observed accuracy was not significantly greater than chance ($p = 0.664$; Figure 1B). In fact, LDA via RFE correctly classified out-of-sample "noise" 83.3% of the time on average– slightly better than Ohsowski et al. (78.8%). This suggests that the observed spatial autocorrelation allows even random noise to be classified just as accurately as Ohsowski et al.

Our first null model assumes complete independence among samples, such that any observed similarity in morphology is attributable solely to shared taxonomic identity. In contrast, our second null model assumes that samples are spatially-correlated and therefore any similarity in morphology is due to being in the same transect. Reality likely lies somewhere between these two scenarios, but unfortunately it is difficult to determine precisely where because transect ID and taxonomic ID are nearly perfectly confounded. As a result, it is also not currently possible to know whether the Ohsowski et al. LDA model predicts taxonomic ID better than random chance or not.

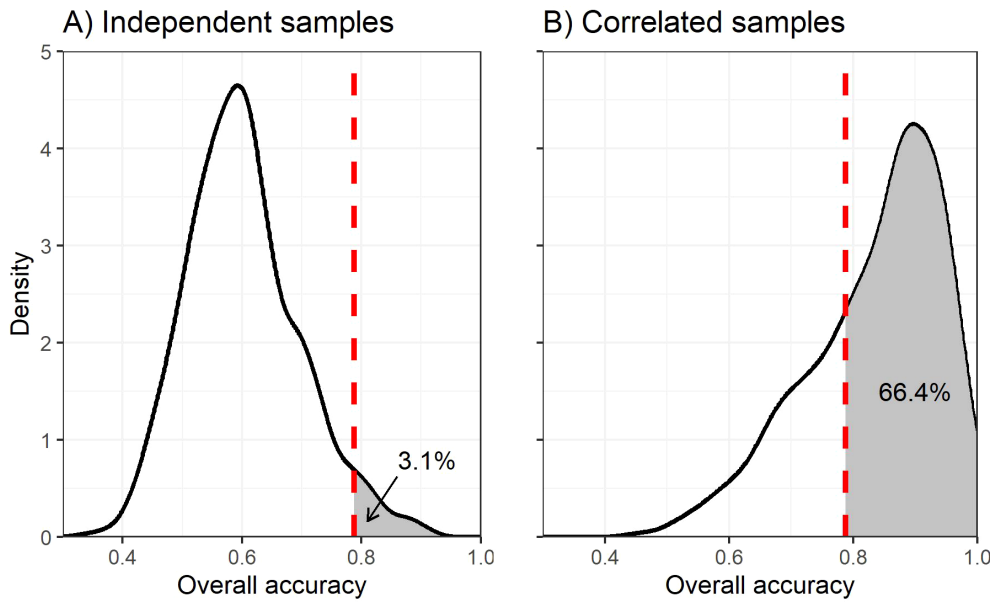


FIGURE 1
Null distribution for overall prediction accuracy when **(A)** samples are assumed independent or **(B)** spatially correlated within transects. The dashed line represents the overall prediction accuracy observed by Ohsowski et al. (2024), and shaded areas represent the proportion of null model simulations that outperformed Ohsowski et al. (2024).

Discussion

Our null models show the importance of sample independence for LDA and other multivariate techniques. Sampling clonal plants creates at least two sources of spatial non-independence: genetic and environmental. Travis et al. (2011) showed that ramets are frequently genetically identical up to 20 m apart (and occasionally up to 90 m), suggesting that samples should be collected at least 20m and ideally ~100 m apart. Additionally, non-independence can arise for traits that show environmental phenotypic plasticity, e.g. wider, longer, and more numerous leaves in nutrient-enriched environments (Knops and Reinhart, 2000). Together, these observations suggest it is better to sample few individuals from many wetlands, rather than many individuals in few wetlands. In fact, the ideal sampling design may involve collecting one individual per taxonomic group per wetland environment, across as wide a range of environmental conditions as possible.

This idealized sampling scheme is costly, labor-intensive, and would likely reduce sample sizes. However, the relatively positive outcomes from our first null model demonstrate that small sample size is less of a concern than non-independence. We recommend that future research prioritizes sample independence over sample size. Such a design is also ideally suited for a broad collaborative effort. Individual researchers using identical sampling protocols could collect a few samples from just a few locations, then combine samples to assess the ability of morphological traits to distinguish *Typha* taxa. A broad collaboration can overcome SSS and sample independence problems to seek generalizable, useful morphological traits.

Author contributions

KE: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. SA: Formal analysis, Writing – review & editing. JM: Writing – review & editing. DG: Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

We thank Dr. Doug Wilcox for helpful comments on this manuscript. The authors declare no conflicts of interest.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Calvo, C., Elgersma, K. J., Goldberg, D. E., Currie, W. S., Batzer, E., and Martina, J. P. (2025). Biotic resistance across a nutrient gradient in experimental wetland mesocosms. *Ecol. Appl.* 35, e70029. doi: 10.1002/eap.70029

Geddes, P., Murphy, L., Astudillo-Scalia, Y., Blasini, D., Nugent, S., Rios, M. J., et al. (2021). Microsatellite markers reveal unprecedented high frequencies of hybridization among *Typha* species in the Midwestern US. *Wetlands* 41, 24. doi: 10.1007/s13157-021-01429-2

Huberty, C. J., and Olejnik, S. (2007). *Applied MANOVA and discriminant analysis. 2nd ed* (Hoboken, NJ, USA: John Wiley & Sons).

Knops, J. M., and Reinhart, K. (2000). Specific leaf area along a nitrogen fertilization gradient. *Am. Midland Nat.* 144, 265–272. doi: 10.1674/0003-0031(2000)144[0265:SLAAAN]2.0.CO;2

Kuehn, M. M., and White, B. N. (1999). Morphological analysis of genetically identified cattails Typha latifolia, Typha angustifolia, and *Typha* × *glauca*. *Can. J. Botany.* 77, 906–912. doi: 10.1139/b99-037

Mosteller, F., and Tukey, J. W. (1977). *Data analysis and regression* (MA, USA: Addison-Wesley; Reading).

Ohsowski, B. M., Redding, C., Geddes, P., and Lishawa, S. C. (2024). Field-based measurement tools to distinguish clonal *Typha* taxa and estimate biomass: a resource for conservation and restoration. *Front. Plant Sci.* 15, 1348144. doi: 10.1111/j.1365-2494.1964.tb01175.x

Sharma, A., and Paliwal, K. K. (2015). Linear discriminant analysis for the small sample size problem: an overview. *Int. J. Mach. Learn. Cyber.* 6, 443–454. doi: 10.1007/s13042-013-0226-9

Snow, A. A., Travis, S. E., Wildová, R., Fér, T., Sweeney, P. M., Marburger, J. E., et al. (2010). Species-specific SSR alleles for studies of hybrid cattails (*Typha latifolia* × *T. angustifolia*; Typhaceae) in North America. *Am. J. Bot.* 97, 2061–2067. doi: 10.3732/ajb.1000187

Travis, S. E., Marburger, J. E., Windels, S. K., and Kubátová, B. (2011). Clonal structure of invasive cattail (*Typhaceae*) stands in the upper Midwest region of the. *Wetlands* 31, 221–228. doi: 10.1007/s13157-010-0142-7

Xu, Y., and Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* 2, 249–262. doi: 10.1007/s41664-018-0068-2

Yates, L. A., Aandahl, Z., Richards, S. A., and Brook, B. W. (2023). Cross validation for model selection: a review with examples from ecology. *Ecol. Monogr.* 93, e1557. doi: 10.1002/ecm.1557