Journal of Economic Entomology, 114(4), 2021, 1842–1846
doi: 10.1093/jee/toab127
Advance Access Publication Date 28 June 2021
Short Communication

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Sampling and Biostatistics

Evaluating the Quality of Ecoinformatics Data Derived From Commercial Agriculture: A Repeatability Analysis of Pest Density Estimates

Jay A. Rosenheim[1,]

Department of Entomology and Nematology, University of California Davis, One Shields Avenue, Davis, CA 95616, USA and [1]Corresponding author, e-mail: jarosenheim@ucdavis.edu

Subject Editor: Dominic Reisig

Received 21 February 2021; Editorial decision 3 June 2021

Abstract

Each year, consultants and field scouts working in commercial agriculture undertake a massive, decentralized data collection effort as they monitor insect populations to make real-time pest management decisions. These data, if integrated into a database, offer rich opportunities for applying big data or ecoinformatics methods in agricultural entomology research. However, questions have been raised about whether or not the underlying quality of these data is sufficiently high to be a foundation for robust research. Here I suggest that repeatability analysis can be used to quantify the quality of data collected from commercial field scouting, without requiring any additional data gathering by researchers. In this context, repeatability quantifies the proportion of total variance across all insect density estimates that is explained by differences across populations and is thus a measure of the underlying reliability of observations. Repeatability was moderately high for cotton fields scouted commercially for total *Lygus hesperus* Knight densities (R = 0.631) and further improved by accounting for observer effects (R = 0.697). Repeatabilities appeared to be somewhat lower than those computed for a comparable, but much smaller, researcher-generated data set. In general, the much larger sizes of ecoinformatics data sets are likely to more than compensate for modest reductions in measurement precision. Tools for evaluating data quality are important for building confidence in the growing applications of ecoinformatics methods.

Key words: data quality, data veracity, repeatability analysis, observer effect, *Lygus hesperus*

Analysis of large, observational data sets gathered in commercial agricultural production, termed "big data" or "ecoinformatics" research, is finding increasing application as a research methodology that can complement traditional small-plot experimentation (Jiménez et al. 2009, 2019; Cock et al. 2011; Rosenheim and Gratton 2017; Shakoor et al. 2019). These methods are also being adopted aggressively by the private sector, with the rapid growth of informatics-based companies working in agriculture and fueled by the increased availability of data from sensors mounted on tractors and harvesters, from remote sensing, and from other sources (e.g., The Climate Corporation, FBN, Trimble, Granular).

Ecoinformatics requires an abundant source of data; for research in agricultural pest management, abundant data often do exist, because farmers and consultants (pest control advisors) scout many fields to generate estimates of insect densities that drive real-time management decisions. This field scouting represents a vast, decentralized data-gathering effort. Marshaling these data can produce large, real-world data sets that, when analyzed properly, can generate powerful research insights (de Valpine et al. 2010, Frost et al. 2013, Cohen et al. 2017, Zhang et al. 2018, Emery and Mills 2019, Paredes et al. 2021).

Big data research methods have, however, been met with skepticism. Primary concerns include issues surrounding data privacy and the difficulties of inferring causal relationships from purely observational data (Rosenheim et al. 2011, Rosenheim and Gratton 2017). In this paper, I focus on another core concern regarding ecoinformatics methods in pest management research: are commercial field scouting data of a sufficiently high quality to be the foundation for robust research (Farley et al. 2018, Shakoor et al. 2019, Aubin et al. 2020)? Are we at risk of a "garbage in, garbage out" scenario, where no amount of sophisticated data analysis methodologies will be able to compensate for a flawed initial data set?

There are several possible ways to evaluate the quality of ecoinformatics data. One approach does not examine the raw data directly, but rather tests the 'downstream' products of ecoinformatics research, asking if the conclusions that emerge from ecoinformatics analyses can be confirmed by analysis of researcher-generated observational or experimental data sets (e.g., Rosenheim and Meisner 2013, Cass et al. 2019, 2021). Another approach is to assess directly the quality of raw data derived from commercial scouting. Researchers can visit fields immediately following a commercial scout, and repeat the same sampling procedures (e.g., Rosenheim et al. 2006). These methods can be quite labor-intensive, however, and professional consultants may, understandably, not always want to be evaluated in this way.

Perhaps most desirable are methods of assessing ecoinformatics data quality that do not require additional data to be collected, and that instead use internal features of an ecoinformatics data set to evaluate data quality. Here I capitalize on such a 'built-in' opportunity that is offered by an ecoinformatics dataset that includes a repeated sampling of the same cotton fields, on the same day (or rarely the following day), by different scouts to quantify densities of the western tarnished plant bug, *Lygus hesperus* Knight (Hemiptera: Miridae). Although the efficiency demands of commercial scouting usually preclude redundant sampling of fields, in some cases consultants may repeat sampling to confirm key observations before costly interventions are recommended. Formal analysis of the repeatability of pest density estimates provides a direct assessment of the reliability of commercially gathered data. As a point of comparison, I also conduct a repeatability analysis of a much smaller researcher-generated dataset on *L. hesperus* densities in cotton.

Methods

I analyzed a subset of an ecoinformatics data set on densities of *L. hesperus* gathered from commercial cotton production (three large ranches, each with many fields; field size ca. 61–182 ha) in the San Joaquin Valley, California. One independent pest control consultant hired one, or sometimes two, summer field scouts each year (some scouts returned for a second summer) to assist with sampling. Samples were taken with a standard sweep net (38.1 cm diameter) swung fifty times through the upper plant canopy; counts of nymphs and adults were recorded separately, and are reported throughout as numbers per sweep sample (per 50 net swings). Although most fields were sampled only once, approximately weekly, some fields were sampled both by the consultant and by a field scout on the same day (less commonly separated by one day), often as the field neared a threshold triggering a pesticide application. Sweep samples were typically taken at several locations, spread ca. evenly across the field; on average, the consultant took 10.6 ± 6.8 (SD) sweep samples per field (range 3–47) and the scout took 16. 2 ± 8.3 samples (range 3–52), with a larger number of samples typically taken in larger fields. The identity of the sampler was recorded with each sampling record. Pairs of consultant-scout repeat samplings (n = 253, with seven different scouts) were gathered from 1998 to 2007. *Lygus hesperus* densities in sampled fields averaged 1. 32 ± 1.74 (SD) nymphs (range 0–12.67) and 2. 33 ± 1.94 (SD) adults (range 0–12.86) per sweep sample.

Repeatability analyses were conducted using the *rptR* package (Stoffel et al. 2017, 2019) in R, which fits linear, mixed effect statistical models to the data using the *lme4* package (Bates et al. 2015). Both (1) 'consistency repeatabilities' and (2) 'adjusted repeatabilities,' which include statistical control for observer effects, were computed. Confidence intervals were calculated with parametric bootstrapping with 1,000 replicates, and P-values were computed with likelihood ratio tests. The repeatability metric, R, measures the proportion of total variance across all mean density estimates that is explained by differences between 'groups' (Nakagawa and Schielzeth 2010); for my analysis, each pair of observed mean *L. hesperus* densities recorded by the consultant and the scout for a given field on a given date forms a group. These groups are treated as a random effect, and the repeatability analysis quantifies how much of the total variance in *L. hesperus* mean density estimates is explained by this random effect. Total variance is equal to the sum of between-group differences and within-group differences. The repeatability value, R, can be compared with the coefficient of variation explained, r^2, in simple linear regression, as both reflect the proportion of variance explained. Within-group differences, which we hope to minimize for high-quality data, are generated by differences between different samplers (observer effects), within-field spatial heterogeneity in *L. hesperus* densities (the consultant and scout do not sample precisely the same locations within each field), stochasticity in the process of capturing nymphs with the sweep net, and errors in finding all captured insects. Small immatures are especially difficult to recover from sweep nets, as they are often lodged among leaves and other plant debris at the base of the net bag, whereas the larger nymphs and adults can usually quickly extricate themselves from the plant debris and climb up the sides of the net bag, where they are readily counted (Rosenheim et al. 2006). Observer effects were expected, because no two individuals swing a sweep net in precisely the same way, or have precisely the same skills in finding the insects within the net bag.

Mean *L. hesperus* density estimates generated by the consultant and the scouts were based on quite variable numbers of replicate sweep samples (ranges for consultant 3–47 and for summer scouts 3–52). To see if the agreement between consultant and scout density estimates was improved when greater numbers of sweep samples were conducted, I tested if the absolute values of the residuals from a regression of scout total *L. hesperus* counts on consultant total counts were inversely related to the number of replicate sweep samples (1-tailed tests).

Some of the lack of repeatability of mean *L. hesperus* density estimates comes from the inherent variation seen across the multiple sweep net samples made by a single observer on a single date in a single field; this source of variation is ubiquitous (e.g., Sevacherian and Stern 1972). I performed simulations to produce an approximate quantification of the impact of this inherent between-sweep sample variation on values obtained in the broader repeatability analysis. I used non-parametric bootstrapping to create n = 400 replicate sets of sweep samples. Bootstrapping resamples with replacement from the collection of *L. hesperus* counts observed across the sweep samples taken in a particular focal field. I retained the same number of sweep samples as were actually taken for each field sampled, and used these to generate different mean density estimates. These were combined across all 253 sampled fields to create sets of bootstrapped mean estimates. These sets were then randomly paired to create 200 replicate pairs of mean density estimates that could be fed into the repeatability analysis. This was done separately for (1) the consultant data, and (2) the data generated by the group of summer field scouts. This analysis was designed to reveal the degree to which the inherent variation across replicate sweep samples taken by a single field scout within a single field contributed to the loss of repeatability observed when different field scouts sampled the same field.

To provide a point of comparison, I also present a repeatability analysis of a much smaller data set of *L. hesperus* densities, generated by three research assistants and me in nine commercial cotton fields in the San Joaquin Valley, California in 2004. Approximately 1–2 ha subplots of large commercial fields were sampled; thus, it is likely that only part of the spatial heterogeneity present in the field was sampled. One researcher sampled all nine fields (mean number of sweep samples per field 8.89 ± 1.54; range 5–10), and each field was simultaneously sampled by one of three other researchers (mean samples per field 9.11 ± 1.69; range 5–11). Methods were the same as those used by the commercial consultant; *L. hesperus* densities in sampled fields averaged 2.07 ± 1.50 (SD) nymphs (range 0–5.10) and 3.87 ± 2.01 (SD) adults (range 1.33–8.00) per sweep sample.

Results

Consultant Data Set

Consistency repeatability values, which excluded observer effects, were R = 0.597 ± 0.042 (SE) (95% CI 0.509–0.669, P < 0.001) for *L. hesperus* nymphs; R = 0.580 ± 0.042 (SE) (95% CI 0.492–0.654, P < 0.001) for adults; and R = 0.631 ± 0.038 (SE) (95% CI 0.547–0.703, P < 0.001) for total counts (all motile stages combined). Thus, 63.1% of the variation in total *L. hesperus* counts was explained by differences between fields, with the remaining 36.9% of the variation explained by differences between the two mean density estimates made for each field, reflecting effects of within-field spatial heterogeneity, observer effects, and sampling error.

A simple linear regression shows that the data were somewhat noisy, and also that the consultant tended to record greater *L. hesperus* counts than did the scouts (Fig. 1). Linear regressions examining relationships between mean density estimates generated by the consultant and each of the summer scouts considered individually suggested that each of the five scouts who sampled ≥20 fields showed the same pattern of producing lower *L. hesperus* counts than did the consultant (data not shown).

Including *observer* as a fixed effect in the analysis produced adjusted repeatability values that were moderately higher: R = 0.650 ± 0.035 (SE) (95% CI 0.585–0.727, P < 0.001) for *L. hesperus* nymphs; R = 0.629 ± 0.038 (SE) (95% CI 0.558–0.706, P < 0.001) for adults; and R = 0.697 ± 0.032 (SE) (95% CI 0.636–0.762, P < 0.001) for all motile stages combined. In the adjusted repeatability analysis for total counts, the observer effect itself had a repeatability of 0.065 ± 0.016 (95% CI: 0.043–0.104), consistent with the modest improvement in the adjusted repeatability value (R = 0.697) compared to the consistency repeatability value (R = 0.631).

I found only mixed support for the idea that density estimates based on greater numbers of sweep samples would be more repeatable. The magnitudes of residuals from a regression of scout total *L. hesperus* counts on consultant total counts were not related to consultant sweep sample number (coefficient −0.0017 ± 0.0103 SE, df = 249, P = 0.43) and only marginally related to scout sample number (coefficient −0.0134 ± 0.0082 SE, df = 249, P = 0.052). This analysis suggests that the inherent variation seen across replicate sweep samples taken in a field contributes only modestly to depressing the repeatability of the mean estimates. This conclusion was reinforced by the simulation-based explorations of the consequences of variation across sweep samples. Bootstrapped resamplings of the replicate sweep samples made in each field yielded data sets with repeatabilities of 0.946 ± 0.009 (SD) for the consultant and 0.962 ± 0.006 (SD) for the summer field scouts.

Researcher Data Set

The researcher data set produced qualitatively similar results, but higher repeatability estimates (Fig. 2). Consistency repeatability

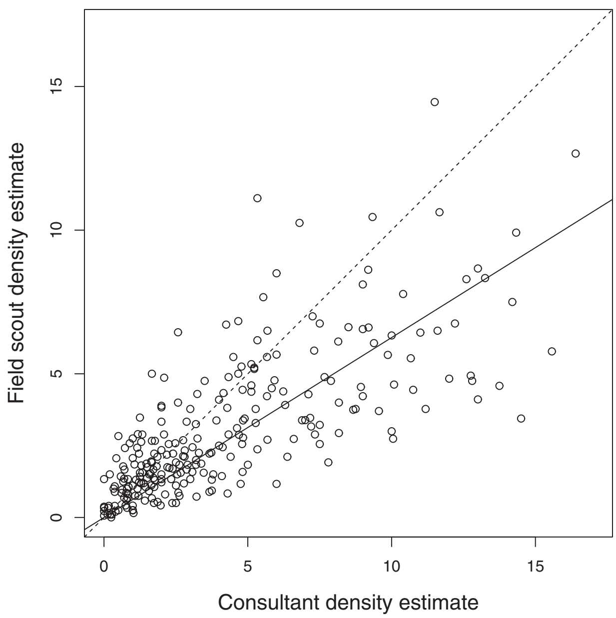

Fig. 1. Bivariate scatterplot showing the repeatability of mean density estimates of *L. hesperus* (nymphs plus adults) generated by averaging across multiple sweep net samples made in commercial cotton. Each field was sampled on the same day, or rarely on two successive days, by a consultant and a summer field scout employed by the consultant. The dashed line has a slope = 1 and includes the origin. The solid line is a linear regression, constrained to go through the origin, showing that summer scouts tended to under-estimate *L. hesperus* densities relative to the consultant (slope = 0.626 ± 0.020 (SE), r = 0.893, df = 252, P < 0.001).

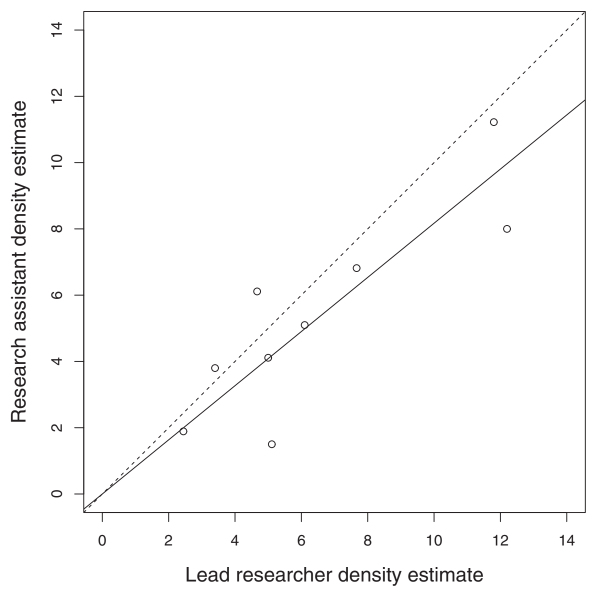

Fig. 2. Bivariate scatterplot showing the repeatability of mean density estimates of *L. hesperus* (nymphs plus adults) generated by averaging across multiple sweep net samples made in commercial cotton. Each field was sampled on the same day by two members of a university research team. The dashed line has a slope = 1 and includes the origin. The solid line is a linear regression, constrained to go through the origin (slope = 0.817 ± 0.073 (SE), r = 0.970, df = 8, P < 0.001).

values, which excluded observer effects, were $R = 0.898 \pm 0.093$ (SE) (95% CI 0.610–0.976, $P < 0.001$) for *L. hesperus* nymphs; $R = 0.623 \pm 0.219$ (SE) (95% CI 0.016–0.888, $P = 0.026$) for adults; and $R = 0.816 \pm 0.152$ (SE) (95% CI 0.392–0.951, $P = 0.001$) for total counts. Adjusted repeatability values, including observer effects, were moderately higher: $R = 0.910 \pm 0.071$ (SE) (95% CI 0.732–0.989, $P < 0.001$) for *L. hesperus* nymphs; $R = 0.744 \pm 0.157$ (SE) (95% CI 0.378–0.966, $P = 0.008$) for adults; and $R = 0.901 \pm 0.078$ (SE) (95% CI 0.731–0.987, $P < 0.001$), for total counts. In the adjusted repeatability analysis for total counts, the observer effect had a repeatability of 0.063 ± 0.080 (95% CI: 0.011–0.317).

Discussion

The repeatability of total *L. hesperus* density estimates in the ecoinformatics data set, a fundamental measure of data reliability, was moderately high. 63.1% of the total variation in mean *L. hesperus* densities was explained by between-field differences in *L. hesperus* numbers. Including the observer effect in the repeatability analysis increased the variance explained to 69.7%, with the remaining 30.3% of the variation that was unexplained produced by within-field spatial heterogeneity in *L. hesperus* densities and sampling error. Simulations suggested that variation across replicate sweep samples made by a single observer within a given field contributed only a small part of the total unexplained variation in mean *L. hesperus* density estimates; thus, the commercial field scouts did appear to be taking a sufficiently large number of replicate sweep samples to stabilize their estimate of mean *L. hesperus* density. Fields for which larger numbers of replicate sweep samples were made to produce a single mean density estimate were associated with only small improvements in repeatability. This may be because larger numbers of sweep samples were often taken in larger cotton fields, such that the expected improvement in density estimation was partially offset by the greater spatial heterogeneity in *L. hesperus* densities within larger fields. *Lygus hesperus* are notoriously difficult to sample, because they can be very patchily distributed within fields (Sevacherian and Stern 1972), because small nymphs are difficult to locate in sweep net bags (Rosenheim et al. 2006), and because *L. hesperus* can be very damaging even at low densities (1–2 per sweep sample, Rosenheim and Meisner 2013), forcing consultants to try to sample very low-density populations. Thus, this was a difficult test case for a study of ecoinformatics data quality. A parallel repeatability analysis of researcher-generated *L. hesperus* density estimates produced similar results, with somewhat higher repeatability values, although the small size of the researcher-generated data set meant that the confidence intervals for the repeatability values from the consultant-generated and researcher-generated datasets were broadly overlapping.

Improved repeatability values were observed in analyses that included observer effects for both the consultant- and researcher-generated data sets. This suggests that, whenever possible, including observer identity in statistical models of ecoinformatics data will enhance the quality of the resulting inferences.

Although the repeatability values for the researcher-generated data set appeared to be somewhat greater than those for the consultant-generated data set, this improvement would, in research practice, likely be offset by the larger sizes of most ecoinformatics data sets (Rosenheim and Gratton 2017). For example, researcher-generated observational data sets on *L. hesperus* population densities in California cotton have, in recent years, including a study of 18 fields (de Valpine and Rosenheim 2008), 21 fields (Rosenheim et al. 2006), and 136 fields (Carrière et al. 2012), whereas an ecoinformatics data set included 1118 fields (Rosenheim and Meisner 2013). Despite using what may have been slightly noisier data, the ecoinformatics data set had the statistical power to confidently resolve economically-important yield effects ($P < 0.0001$) that had been marginally non-significant in previous, experimental work (discussed in Rosenheim and Meisner 2013).

In this paper, I used repeatability analysis to evaluate noisiness in mean insect density estimates across observers, comparing consultant versus scout estimates. Not all ecoinformatics data sets will include a repeated sampling of the same fields on the same days, limiting the applicability of this approach. However, another more broadly applicable approach would be to examine the repeatability of the multiple sub-samples that are typically taken when large fields are sampled. Such an approach would be imperfect in some respects, as it would both underestimate some sources of error (by excluding across-observer differences) and overestimate other sources of error (because sub-samples are not used in isolation when making pest management decisions, but rather are averaged, which creates a more stable estimate of overall density). Nevertheless, this type of repeatability analysis would produce an objective measure of data reliability that could be compared with comparable measures from researcher-generated data.

In cases where researcher-generated insect density estimates differ from density estimates generated in the course of commercial farming, it may be tempting to prefer the researcher's data. However, a case can sometimes be made for preferring density estimates generated in commercial agriculture, even if they are noisier. If researchers produce management recommendations for farmers that are conditioned on pest densities, it will be sampling in the commercial setting that is most critical for the actual implementation of pest management. A barrier to the implementation of research-derived recommendations can be erected when researchers and commercial pest management advisors are 'speaking different languages.' Thus, working with commercial data, even if they are noisy, can in some cases be advantageous, because it allows researchers and farmers to work with fully comparable data.

Ecoinformatics studies will rightfully continue to be scrutinized for underlying data quality. It is therefore important for agricultural entomologists to develop efficient and flexible means for quantifying data quality; repeatability analyses are likely to be a useful tool in those efforts.

Acknowledgments

I thank the commercial consultant and summer field scouts who generated and then generously shared the data analyzed here. Kimberly Steinmann, Gail Langellotto, and Jill Hodgen assisted in gathering the researcher's data set on *L. hesperus* densities. Perry de Valpine provided very constructive feedback on an earlier draft of the manuscript, and Michael Culshaw-Maurer provided much-appreciated assistance with coding. This work was funded by United States Department of Agriculture—National Institute of Food and Agriculture (USDA-NIFA) grant 2015-70006-24164.

References Cited

Aubin, I., F. Cardou, L. Boisvert-Marsh, E. Garnier, M. Strukelj, and A. D. Munson. 2020. Managing data locally to answer questions globally: the role of collaborative science in ecology. J. Veg. Sci. 31: 509–517.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Package 'lme4', vers. 1.1–13. http://lme4.r-forge.r-project.org/.

Carrière, Y., P. B. Goodell, C. Ellers-Kirk, G. Larocque, P. Dutilleul, S. E. Naranjo, and P. C. Ellsworth. 2012. Effects of local and landscape factors on population dynamics of a cotton pest. PLoS One. 7: e39862.

Cass, B. N., E. E. Grafton-Cardwell, and J. A. Rosenheim. 2019. Resistance of fruits from a mandarin cultivar to feeding by fork-tailed bush katydids. J. Econ. Entomol. 112: 2861–2871.

Cass, B. N., H. M. Kahl, T. G. Mueller, X. Xi, E. E. Grafton-Cardwell, and J. A. Rosenheim. 2021. Profile of fork-tailed bush katydid (*Scudderia furcata* Orthoptera: Tettigoniidae) feeding on fruit of clementine mandarins. J. Econ. Entomol. 114: 215–224.

Cock, J., T. Oberthür, C. Isaacs, P. R. Läderach, A. Palma, J. Carbonell, J. Victoria, G. Watts, A. Amaya, L. Collet, *et al.* 2011. Crop management based on field observations: case studies in sugarcane and coffee. Agric. Syst. 104: 755–769.

Cohen, Y., I. Roei, L. Blank, E. Goldshtein, and H. Eizenberg. 2017. Spatial spread of the root parasitic weed *Phelipanche aegyptiaca* in processing tomatoes by using ecoinformatics and spatial analysis. Front. Plant Sci. 8: 973.

de Valpine, P., and J. A. Rosenheim. 2008. Field-scale roles of density, temperature, nitrogen, and predation on aphid population dynamics. Ecology. 89: 532–541.

De Valpine, P., K. Scranton, and C. P. Ohmart. 2010. Synchrony of population dynamics of two vineyard arthropods occurs at multiple spatial and temporal scales. Ecol. Appl. 20: 1926–1935.

Emery, S. E., and N. J. Mills. 2019. Sources of variation in the adult flight of walnut husk fly (Diptera: Tephritidae): a phenology model for California walnut orchards. Environ. Entomol. 48: 234–244.

Farley, S. S., A. Dawson, S. J. Goring, and J. W. Williams. 2018. Situating ecology as a big-data science: current advances, challenges, and solutions. Bioscience. 68: 563–576.

Frost, K. E., P. D. Esker, R. Van Haren, L. Kotolski, and R. L. Groves. 2013. Factors influencing aster leafhopper (Hemiptera: Cicadellidae) abundance and aster yellows phytoplasma infectivity in Wisconsin carrot fields. Environ. Entomol. 42: 477–490.

Jiménez, D., J. Cock, H. F. Satizábal, M. A. Barreto, A. Pérez-Uribe, A. Jarvis, and P. Van Damme. 2009. Analysis of Andean blackberry (*Rubus glaucus*) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Comp. Electron. Agric. 69: 198–208.

Jiménez, D., S. Delerce, H. Dorado, J. Cock, L. A. Muñoz, A. Agamez, and A. Jarvis. 2019. A scalable scheme to implemente data-driven agriculture for small-scale farmers. Glob. Food Secur. 23: 256–266.

Nakagawa, S., and H. Schielzeth. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev. Camb. Philos. Soc. 85: 935–956.

Paredes, D., J. A. Rosenheim, R. Chaplin-Kramer, S. Winter, and D. S. Karp. 2021. Landscape simplification increases vineyard pest outbreaks and insecticide use. Ecol. Lett. 24: 73–83.

Rosenheim, J. A., and C. Gratton. 2017. Ecoinformatics (big data) for agricultural entomology: pitfalls, progress, and promise. Annu. Rev. Entomol. 62: 399–417.

Rosenheim, J. A., and M. H. Meisner. 2013. Ecoinformatics can reveal yield gaps associated with crop-pest interactions: a proof-of-concept. PLoS One. 8: e80518.

Rosenheim, J. A., K. Steinmann, G. A. Langellotto, and A. G. Zink. 2006. Estimating the impact of Lygus hesperus on cotton: the insect, plant, and human observer as sources of variability. Environ. Entomol. 35: 1141–1153.

Rosenheim, J. A., S. Parsa, A. A. Forbes, W. A. Krimmel, Y. H. Law, M. Segoli, M. Segoli, F. S. Sivakoff, T. Zaviezo, and K. Gross. 2011. Ecoinformatics for integrated pest management: expanding the applied insect ecologist's tool-kit. J. Econ. Entomol. 104: 331–342.

Sevacherian, V., and V. M. Stern. 1972. Spatial distribution patterns of Lygus bugs in California cotton fields. Environ. Entomol. 1: 695–704.

Shakoor, N., D. Northrup, S. Murray, and T. C. Mockler. 2019. Big data driven agriculture: big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity. Plant Phenome J. 2: 180009. doi:10.2135/tppj2018.12.0009

Stoffel, M. A., S. Nakagawa, and H. Schielzeth. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol. Evol. 8: 1639–1644.

Stoffel, M., S. Nakagawa, and H. Schielzeth. 2019. Package 'rptR', vers. 0.9.22. http://rptR.r-forge.r-project.org/.

Zhang, W., Y. Lu, W. van der Werf, J. Huang, F. Wu, K. Zhou, X. Deng, Y. Jiang, K. Wu, and M. W. Rosegrant. 2018. Multidecadal, county-level analysis of the effects of land use, Bt cotton, and weather on cotton pests in China. Proc. Natl. Acad. Sci. U. S. A. 115: E7700–E7709.