Review Article

Recognizing the importance of ecoinformatics in agricultural entomology

Abstract

Ecology relies on collecting, integrating, and interpreting large volumes of diverse data, which remains a greater challenge for ecologists. Ecoinformatics could provide a solution for this which integrates agricultural and ecosystem sciences with computer sciences, geographic information sciences and quantitative methods. It provides tools and methods for organizing and converting ecological data into information and knowledge. The use of ecoinformatics in entomology progresses from documenting pest and disease patterns and their colonization, pest impact on crop yield to the food web and farmer's decision making. However, there are still numerous obstacles to overcome, particularly when it comes to integrating ecoinformatics methods into mainstream research and education. Several technical and sociocultural challenges like collaboration with statisticians, developing data sources, data privacy concerns and developing mobile platforms for pest management remain some of the major hurdles. Since it is akin to experimental methodology, ecoinformatics-based techniques when well integrated, give a more optimal solution to pest management problems than any single strategy employed independently.

Key words: bulk data, interpretation, pest management, decision making

Introduction

Ecology is the field of biology concerned with the interactions of organisms with their physical environments. It is evolving quickly and progressively changing to a more open, liable, collaborative, interdisciplinary and data-intensive discipline. Discovering, integrating and evaluating enormous amounts of heterogeneous information are central to ecology, which remains a greater challenge for ecologists. Ecoinformatics refers to ecological studies that use preexisting data. It suggests tools and methods for handling ecological data and transmuting the data into information and knowledge [1]. This emerging field integrates ecosystem and agricultural sciences with quantitative methods, computer sciences and geographic information sciences. Entomologists and ecologists have been using these approaches for the past 75 years, for example, locust outbreak patterns and colonization pathways to Europe were studied using historic maps, texts and museum records dating back to 300 CE [2,3]. Thus a question arises, is there anything new in this approach? or are there any new features added to it?

Ecoinformatics is a branch of big data research systems in which the data sets are characterized by high data variety, velocity and volume. In agricultural entomology, ecoinformatics can be used in recording pest and disease patterns, pest impact on crop yield and patterns in pesticide use, the efficacy of transgenic crops for pest control, landscape context effects on crop colonization by pests, beneficial insects, the efficacy of cultural controls and host-plant resistance, food web and farmer's decision making [4]. However, there are still numerous obstacles to overcome, particularly in terms of bringing ecoinformatics methods into mainstream research and education. Several technical and sociocultural challenges like collaboration with statisticians, developing data sources, data privacy concerns and developing mobile platforms for pest management remains as some of the major hurdles. Ecoinformatics is akin to an experimental methodology. Thus, when experimental, observational, and ecoinformatics-based techniques are combined, they give more effective answers to pest control problems than if they are employed alone [5].

Features of ecoinformatics

It mainly relies on the use of pre-existing data, which in the majority of cases are observational. Besides, it allows to access data with large temporal and spatial scales and to integrate data from multiple sources. Thus, a large amount of data will be generated and ecoinformatics provides new tools for managing and analyzing these data. The features of data informatics lie in the characteristics of data sources, data set construction, their statistical considerations and finally the acceptance of research results.

Data sources: Data required for studies could be collected from a variety of new and old information sources like private and public data repositories, passive surveillance and indirect sampling, citizen science and academic data [4]. Data can either be collected deliberately to solve a particular entomological question, or might be the by-product of former sampling programs, with data being re-used to answer novel queries. Several regional, state, national and international agencies gather and store a range of entomological, environmental and agricultural data regularly. Some of the examples are; http://traps.ncipmc.org/, http://sba.ipmpipe.org/cgi-bin/sbr/public.cgi and https://datcpservices.wisconsin.gov/pb. Insect monitoring is conducted at the quarantine stations, by surveying agricultural producers for recording the population trends, the occurrence of new insect/ pathogen [6] and the sampling data collected as a part of integrated pest management by private pest management consultants and agricultural cooperatives. As the data collection becomes gradually digital [7], the availability of data about insects in agriculture is expected to expand intensely. The passive surveillance methods are generally used in the case of medical entomology through the disease occurrences recorded either by health agencies [8] or through internet searches [9]. Insects are monitored indirectly by tracing the consequences of their activities, such as the presence of excreta and damaged plants which allows the researchers to understand insect phenology, distributions or abundances. Pest activity can also be assumed from the data on pesticide-use patterns acquired during farmer surveys [10].

Citizen science is the research partnership between scientists and the public to gather, discover and study the data about the natural world. The Internet has got a major role in connecting the amateur naturalist and the scientists and citizen science work well for those insects which can easily be identified. For example http://www.naba.org/butter_counts.html for butterflies, http://bugguide.net for insects and spiders, and https://www.bumblebeewatch.org for bumblebees. These portals enable an unprofessional naturalist to upload observations or images of animals or plants. The academic data already published by researchers stands as an additional source of information. Until recently it was difficult to obtain the data sets which were used in the analysis in published papers in it is raw form. Now several data sources are available like http://www.gbif.org, https://www.idigbio.org, http://vegbank.org, https://www.dataone.org and http://datadryad.org these would eventually extend the availability of experimental data rather than observational for ecoinformatics studies.

Data set construction: Data collected through ecoinformatics differs from conventional research in quality, flexibility, privacy, quantity, and scale. Ecoinformatics data often show greater heterogeneity due to diverse sources and collection methods, whereas conventional research allows researchers superior control over data collection, quality, and uniformity, reducing bias. Additionally, conventional research provides flexibility for incorporating manipulations to study specific conditions of interest. This elasticity enables researchers to design experiments tailored to particular scenarios, making conventional methods advantageous for focused and controlled investigations [11]. On contrary, ecoinformatics data sets are mainly observational in nature and hence are limited to study only those practices which are already in use, not any novel methods or practices. In terms of data privacy, ecoinformatics studies are unattractive to the researchers because of the fear of getting scooped, spotting errors in the data, reduction in the number of publications and misinterpretation of data. Moreover, farmers might not be willing to share the data regarding the crop details or on the yield or about the pest management measures, as such information may be regarded as firmly proprietary [12]. Experimental studies are least affected by the issue of privacy.

While the data quality, privacy and elasticity remain as a drawback for ecoinformatics studies, the huge amount of data that can be collected stands as a major boon. Ecoinformatics studies can facilitate the collection of data in terms of terabytes or petabytes. In experimental studies, the amount of data generated is a limiting factor. The cost of data collection is also less in ecoinformatics studies, thus with the same amount more data could be procured [13]. Spatial and temporal scales of data also matter in data construction. Certain studies cannot be undertaken on a small area and for shorter periods like the effect of climate change on pests and natural enemies, the effect of crop rotations on pest densities and the impact of landscape on pest colonization of crops. Ecoinformatics studies cover an average of 18-22 years, along with several multidecadal data sets [4]. These are far lengthier than classic experimental studies in agricultural entomology [14]. The spatial scale of ecoinformatics studies may range from local [15], continental or even universal level [16].

Statistical considerations: Ecoinformatics studies with a huge amount of data set would create an impression of legitimacy and power, but these data are more vulnerable to selection bias, errors in measurement and unexplained confounding factors [17]. Similarly, the relation between two variables has to be construed with caution as ecoinformatics approaches depend primarily on observational data. Only because a data set is huge, the proverb “correlation does not imply causation” cannot be rejected. Some of the hurdles that encounter while working with ecoinformatics are statistical power, bias, the number of factors examined, mentioned, and correlation and causation.

The use of big data sets gives the advantage of identifying small effect proportions even though the original data sets are noisy. Statistical power studies indicate that large sample sizes are required in the case of pest management research for resolving significant effects fruitfully [18]. While comparing with the value of the crop, insecticides are commonly low-priced thus the farmers who are focused on profit-maximization will be interested in suppressing pest populations even under very low yield reduction. With conventional experimental studies, such small effects cannot be resolved normally[5] but can be considered via bigger ecoinformatics data sets [19]. However, care has to be taken in using a lesser ‘p’ value alone as adequate proof in rejection of a null hypothesis and establish a significant outcome [20]. Nonrandom selection of samples from larger population stands as a major difficulty in interpretation and creates a platform for bias in the response studied. Occurrence of bias in selection is common when pest management methods are applied non randomly across different plots or farmers. For example, in the case of cotton, progressive farmers would be the one to adopt new cultivars of Bt earlier as compared with others. However, such farmers would be those who were likely to produce higher yields even without the yield increment facilitated by Bt. Thus making a spurious correlation between higher yield and Bt cotton [21], makes it hard to find out the exact relationship between yield and genetically modified crops. These biases are not accounted for generally, thus giving an erroneous representation of the average response of population and affects the interpretation [22].

Experimental methods focus on a few manipulated factors, while ecoinformatics explores multiple factors and variables. However, most ecoinformatics studies limit responses to one variable. Including many predictors and responses poses challenges, especially with multicollinearity, which creates interpretational issues that are difficult to resolve solely through statistical methods [23]. False correlations could arise from several other unrecognized and recognized sources also. For example, both crop performance and pest densities could be influenced by variable weather conditions, building abundant opportunities for false correlations. Unrecognized (thus uncorrected) sources of false correlations are the biggest opponents of ecoinformatics studies, as they can result in serious errors of interpretation [4].

In agricultural entomology, the key goal of ecoinformatics is to improve research-mediated recommendations which allow the farmers to undertake management actions that result in preferred outcomes (e.g., pest management). Farmers will get aware of the likely consequences of a particular action only with the knowledge of causal relationships. The drawbacks of observational studies can be bypassed with the power of a well-designed manipulative experiment [24]. Thus, we can reject the proposal of most devoted exponents of big data approaches that "knowledge of correlation alone can fully replace knowledge of causation" as our major research goal [25]. These reasons make ecoinformatics most valued when used in close association with experimental studies.

Acceptance of research results

While working with ecoinformatics methodologies, the research could be integrated with outreach as the independent consultants or farmers acts as the information source of data sets and they could be involved in research activities right from the beginning of a project. Thus, farmers will get more confidence in recommendations arising from the evaluation of their own data, rather than the study undertaken in a university plot [12]. Another pitfall associated with experimental studies is that it is often executed under narrowly organized and agronomically optimal situations, whereas ecoinformatics works can cover the complete spectrum of commercial farming environments [26].

Data lifecycle

Knowledge is obtained by procurement of the data and by transforming these data into information that can be integrated into the body of scientific theories, facts and principles. The lifecycle of data can be explained through eight steps which may or may not be compulsory and that ultimately changes the data into information and further to knowledge. For example, if one is working with a project that focuses on data collection only, they can skip steps, discovery and integration. Steps need not be in order and the stages are not essentially exclusive [1].

1) Planning: In most projects, data planning is underutilized and not valued but it can augment research efficacy and save time. Explicit data organization plans are also required to satisfy the research proposal requirements of the sponsors. A data organization planning tool i.e. DMP Tool was developed by Digital Curation Center, the UK which helps a researcher in creating, revising and reviewing data.

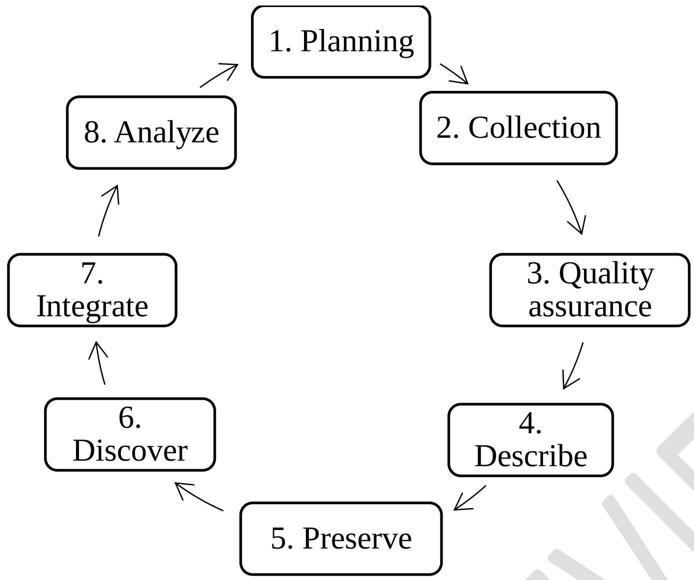

Figure 1: Data life cycle in ecoinformatics [1].

2) Collection: Collection and organisation of ecological data are mainly done by recording the field and laboratory observations manually, hand-held processors, programmed machines and with the help of satellites and sensor networks. Underwater, ground-based and aerial sensor networks encircling many sensors when pooled could provide access to tera to petabytes of data in a year. Examples of these networks are OOI (Ocean Observatories Initiative) [27] and NEON (New environmental observing systems)[28].
3) Quality Assurance: Quality control or quality assurance denotes the process for avoiding mistakes from entering a data set that is used as a precautionary measure to guarantee data of high quality before collection and to uphold data quality all through and after data collection. Quality of data can be assured before collection of data by defining criteria for measurement units, formats and metadata, also by allocating charge for data quality to an individual or a team. Integration of quality assurance with scientific workflow systems and data management systems has to be done to assess the data automatically for completeness of metadata as well as ensuring data quality [29,30]. One of the examples is the development of LTER NIS (Long Term Ecological Research Network Information System) to assess, treat and categorize data products into five levels, fluctuating from a level-0 (slightly modified data) to level-4 (data that are gap filled and semantically in sync to meet the requirements of precise data products)[29].
4) Describe: Adequate information regarding the content, arrangement and framework of a data product could be obtained from metadata. Metadata normally explain the author, reason for collecting data, date in which data was created or collected, the location from which data was collected, the date on which it got modified and the data size [31]. For ensuring reliability in the format and content of metadata, several tools and standards have been established.
5) Preserve: Preservation of data includes depositing data as well as metadata in a data repository or data center where the confirmation, multiplication and curation of data could be done on time [32,33]. A data center often assists a particular community of work that can be related to a researcher or sponsor or area or institution. Data centres provide services like peer review of products, help desk and Digital Object Identifiers (DOI) to the stakeholders for correct identification and citation of the data. For example, Oak Ridge National Laboratory Distributed

Active Archive Center for Biogeochemical Dynamics is providing DOI to facilitate tracking and using data products as a facility to research sponsors as well as data providers [34].

6) Discover: Researchers face significant challenges in data discovery when extending the temporal and spatial scales of their work. Key issues include the reluctance of individuals or institutions to share valuable data stored offline and the overwhelming number of irrelevant results from simple searches, making relevant data difficult to identify. The first issue can be solved by making the researcher and institutions recognize data as a valuable product of the scientific domain and thus should make those data reachable for the rest of the community in order to extend its use [35,36]. The second problem could be solved with the help of projects like DataONE, which are refined and user-friendly search engines that save time for search as well as offer the facility to filter the search results or to pull out specific data according to one need [29].
7) Integrate: Integrating multi-disciplinary data for large-scale ecological studies is vital but challenging, requiring significant time, cost, and effort to address methodology variations, data conversion, and semantic compatibility before analysis. In majority of the cases, integration of data is done manually, which is a large time-consuming process [37]. Several approaches are developing which could help in overcoming these challenges by facilitating semi-automation of data integration process [38]. In contrast to the traditional data integration tools like Excel, which allows for manual data integration methods, semantic models offered by Extensible Observations Ontology (OBOE) and the Observations and Measurements could be used [39].
8) Analyze: Several geospatial and statistical analyses, as well as modelling tools, are required for distinguishing ecological processes from each other because of their interconnections and complex nature[1]. A wide range of statistical tools along with analytical and simulation models are being used by ecologists to find a conclusion for several ecological processes [40]. Unfortunately, data analysis tools and processes are rarely documented in detail, with articles often providing only an outline of methods. Ecoinformatics aims to address this by using new methods to comprehensively record processes leading to scientific interpretations [41–43]. Complete and executable picture of procedures used in the analysis could be obtained from scientific workflow systems such as Tavera, Pegasus, Kepler and VisTralis [42,44].

Ecoinformatics in Agricultural Entomology

Some of the agricultural entomology studies which have used ecoinformatics approaches are mentioned below.

Recording Pest Distribution Patterns: Pest outbreak studies that show heterogeneity in both time and space could effectively be tracked with ecoinformatics approaches. Insect distribution studies necessitate the gathering of a huge amount of data that too with larger temporal and spatial extents. Academic, state and central agencies have been involved in coordinated data collection, some examples are the National Ecological Observatory Network [45] and the National Science Foundation-funded Long-Term Ecological Research (LTER) sites [46]. Localized synchrony in insect populations [47], long-standing variations in pest densities [48] and temporal and spatial scales of variations in population [49] have been studied with ecoinformatics methods. Extensive spatial and temporal studies are vital while working with insect pests in the forest, for which various aerial surveys have been used [50].

Efficiency of Genetically Modified Crops in Pest Management: The majority of the studies related to genetically modified crops are skewed towards *Bt* crops. Long term spatial analysis of data sets showed that a noticeable local suppression of insect pest population has occurred as a result of large scale adoption of *Bt* crops, for example, *Ostrinia nubilalis* in the US [51], *Helicoverpa armigera* in China [52] and *Pectinophora gossypiella* in China [53] and US [54].

Effect of Landscape on Insect Pest Colonization on Crops: The influence of surrounding landscape on insect pests colonizing crops were studied using ecoinformatics methods [55]. Infestation and colonization of *Amyelois transitella* moths in pistachio orchards [56] and the effect of potato storage on colonization by *Premnotrypes* spp. weevils on potatoes were studied in Andes [57,58].

Impact of Pest on Crop Yield and Pesticide Use Pattern: Pest-yield relationships can be analyzed using expert data. One of the studies conducted on *Lygus hesperus* in cotton showed that the farmers of California were managing the pest sub-optimally. Early season yield loss occurred due to inadequate pest control, while unnecessary pesticide application in mid-season disrupted cotton's natural tolerance to pest damage [59,60]. The activity of pests in agricultural landscapes can be indirectly measured with the pesticide use data. The effectiveness of traditional pest management measures under field conditions could also be analyzed with ecoinformatics methods [57,58]. Several studies have pinpointed that variation in pesticide use patterns has aroused due to the variation among the farmers rather than regional differences [15,61]. However, cautious statistical modifications for resolving problems of spatial autocorrelation are required, that would otherwise interfere with the actual effects caused by explanatory variables[15].

Monitoring Beneficial Insects: The extensive monitoring of economically important insects, as well as natural enemies in agriculture, are being done since long back. The records in museum collections revealing the decline of particular groups of unmanaged bees were obtained [62,63]. Citizen science was used to gather continent-wide data on the distribution range of exotic coccinellids [64] and the effect of varying agricultural practices on species turnover patterns of ladybird beetles over 24 years were studied [65]. These studies reveal the importance of ecoinformatics in studying patterns of distribution and abundance which demands long term studies.

Food Webs: Tropical interactions occurring in different ecosystems remains extremely complex and understanding and such a system has always remained a major challenge for ecologists. An initiative to build agricultural food webs automatically has been undertaken [66]. The atrophic web was built with 72 nodes and 407 links by putting machine learning to already collected data on trophic interactions and by robotic text mining of available literature. Thus, revealing several novel and unpredictable trophic relations like intraguild predation of spiders and carabid beetles, which was later established with the support of molecular methods [67].

Efficiency of Host-Plant Resistance and Cultural Control Method: The studies on the efficiency of host plant resistance and cultural control measures in managing pest populations were also undertaken with the support of ecoinformatics. The level of orchard hygiene required in the almond orchard to reduce the pest (*Amyelois transitella*) damage below economic threshold levels was worked out [56]. The study on traits in cassava (*Manihot esculenta*) varieties that influence the level of resistance and susceptibility to three pests were conducted for several decades[68]. Effects of a single year and multi-year crop rotation on pest infestation levels and yield increment were studied with ecoinformatics approaches[69].

Decision Making by Farmer: Experimental methods would not be sufficient to give a holistic view of the effectiveness of pest management. To get a complete picture, one should take care of the decisions taken from the farmer's side also, which can effectively be studied through ecoinformatics approach. These can be explained by a study conducted in China where they found that expectation of pest losses strongly influenced the pesticide use and the selection of pesticide was governed by its price and level of risk to workers[70]. The studies also unveiled the inefficiency of extension personal in spreading awareness about pesticide management measures.

Drawbacks

Even though various ecoinformatics measures has arouse to tackle many of the issues, several socio-cultural and technical problems still exist. Major socio-cultural issues which are faced while implementing ecoinformatics approaches are to escalate application literacy and awareness. Another one is the scarcity of funds for undertaking the research activity. Among the technical challenges, the difficulty in handling and processing such a huge (terabytes and petabytes) amount of data stands out as first. This issue can be solved by attaching high-performance software with the data resources and thereby processing the data sets before transporting. Another technical challenge is to develop innovative visualization methods that would reduce the labour, cost and time in analyzing such complex data at the same time reducing the error rate [71]. Also, more focus should be given to using workflows and algorithms in order to analyze, assure and visualize data.

Future

Ecoinformatics methods should be used along with experimental measures to test the hypothesis and establish the causal relations [12]. This demands a backbone of statistics and computer applications, for which ecologists should collaborate with biostatisticians, engineers and computer scientists. Cooperation with biostatisticians will strengthen the interpretation of data and thus solve the challenges related to analysis [72]. Collaboration with computer scientists and engineers facilitates automation and thereby reduce the time and labour especially in monitoring and detection [73]. Moreover, pest management and monitoring software could be spread rapidly through mobile platforms, which support uploading the observations as images and data and provide situation-specific management decisions [74]. Cyber infrastructure approaches could be tackled to facilitate the storage, retrieval and sharing of ecoinformatics data [75]. Data collection platforms with data collection protocols have to be developed in collaboration with the stakeholders in sub-disciplines to standardize the researcher-developed data sets [76].

Conclusion

Ecoinformatics is an evolving arena that integrates ecology and agricultural studies and improves the study with the application of GI sciences, quantitative techniques and computer sciences. It can represent dynamic aspects of change. On the other hand, many hurdles remain especially in bringing ecoinformatics practices into mainstream research and education. Experimental, observational, and ecoinformatics based approaches if used together, can provide more efficient solutions to problems than using a single approach.

References:

1. Michener, W. K. Jones, M. B. Ecoinformatics: Supporting ecology as a data-intensive science. Trends in Ecology and Evolution, 2012, 27: 85–93.
2. Carpenter, J. R. Insect Outbreaks in Europe. Journal of Animal Ecology, 1940. 9: 108.
3.Waloff, Z. V. The Distribution and Migrations of *Locusta* in Europe. Bulletin of Entomological Research, 1940, 31: 211–246.

4.Rosenheim, J. A. Gratton, C. Ecoinformatics (Big Data) for Agricultural Entomology: Pitfalls, Progress, and Promise. Annual Review of Entomology, 2017, 62: 399–417.
5.Rosenheim, J. A. *et al.* Ecoinformatics for Integrated Pest Management: Expanding the Applied Insect Ecologist's Tool-Kit. Journal of Economic Entomology, 2011, 104: 331–342.
6.Brown, M. *et al.* Epidemiological Surveillance and Control of Bee Diseases. in Bee Health and Veterinarians (ed. Ritter, W.) Food and Environment Research Agency (Fera), National Bee Unit, Sand Hutton, UK. 2014, 193–213.
7.Teacher, A. G. F. Griffiths, D. J. Hodgson, D. J. Inger, R. Smartphones in Ecology and Evolution: A Guide for the Apprehensive. Ecology and Evolution, 2013, 3: 5268–5278.
8.Bisanzio, D. Mutuku, F. LaBeaud, A.D. Mungai, P. L. Muinde, J. Busaidy, H. Mukoko, D. King, C.H. Kitron, U. Use of Prospective Hospital Surveillance Data to Define Spatiotemporal Heterogeneity of Malaria Risk in Coastal Kenya. Malaria Journal, 2015, 14: 482.
9.Gluskin, R. T. Johansson, M. A. Santillana, M. Brownstein, J. S. Evaluation of Internet-Based Dengue Query Data. PLoS Neglected Tropical Diseases, 2014,8: e2713.
10.Larsen, A. E. Agricultural Landscape Simplification Does Not Consistently Drive Insecticide Use. Proceedings of the National Academy of Sciences, 2013, 110: 15330–15335.
11.Lawes, R. A. Lawn, R. J. Applications of Industry Information in Sugarcane Production Systems. Field Crops Research, 2005, 92: 353–363.
12.Cock, J. Oberthür, T. Isaacs, C. Laderach, P.R. Carbonell, J. Victoria, J. Watts, G. Amaya, A. Collet, L. Lema, G. Anderson, E. Crop Management Based on Field Observations: Case Studies in Sugarcane And Coffee. Agricultural Systems, 2011, 104: 755–769.
13.Gardiner, M. M. Allee, L.L. Brown, P.M. Losey, J.E. Roy, H.E. Smyth, R.R. Lessons from Lady Beetles: Accuracy of Monitoring Data from US and UK Citizen-Science Programs. Frontiers in Ecology and the Environment, 2012, 10: 471–476.
14.Schmitz, O. J. Hambäck, B. A. P. Trophic Cascades in Terrestrial Systems: A Review of the Effects of Carnivore Removals on Plants. The American Naturalist, 2000, 155: 141–53.
15.Meehan, T. D. Gratton, C. A Consistent Positive Association Between Landscape Simplification and Insecticide Use Across the Midwestern US from 1997 through 2012. Environmental Research Letters, 2015, 10: 114001.
16.Bebber, D. P. Ramotowski, M. A. T. Gurr, S. J. Crop Pests and Pathogens Move Polewards in a Warming World. Nature Climate Change, 2013, 3: 985–988.
17.Boyd, Danah. Crawford, K. Critical Questions for Big Data. Information, Community and Society, 2012, 15: 662–679.
18.van der Voet, H. Goedhart, P. W. The Power of Statistical Tests Using Field Trial Count Data of Nontarget Organisms, In Environmental Risk Assessment of Genetically Modified Plants. Agricultural and Forest Entomology, 2015, 17: 164–172.
19.Rosenheim, J. A. Meisner, M. H. Ecoinformatics Can Reveal Yield Gaps Associated with Crop-Pest Interactions: A Proof-of-Concept. PLoS One, 2013, 8: e80518.
20.Poole, C. Low P-Values or Narrow Confidence Intervals: Which Are More Durable? Epidemiology, 2001. 12(3):291-94.
21.Kathage, J. Qaim, M. Economic Impacts and Impact Dynamics of Bt (*Bacillus thuringiensis*) Cotton in India. Proceedings of the National Academy of Sciences, 2012, 109: 11652–11656.
22.Siontis, G. C. Ioannidis, J. P. Risk Factors and Interventions with Statistically Significant Tiny Effects. International Journal of Epidemiology, 2011, 40: 1292–1307.
23.Fieberg, J. Johnson, D. H. MMI: Multimodel Inference or Models with Management Implications? Journal of Wildlife Management, 2015, 79: 708–718.

24. Little, R. J. Rubin, D. B. Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes. Annual Review of Public Health, 2000, 21: 121–145.
25. Mayer-Schonberger V. Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. American Journal of Epidemiology, 2013, 179: 1143–1144.
26. Welch, J. R. Vincent, J.R. Aufthammer, M. Moya, P.F. Dobermann, A. Dawe, D. Rice Yields in Tropical/Subtropical Asia Exhibit Large but Opposing Sensitivities to Minimum and Maximum Temperatures. PNAS, 2010, 107: 14562–14567.
27. Cowles, T. Delaney, J. Orcutt, J. Weller, R. The Ocean Observatories Initiative: Sustained Ocean Observing Across a Range of Spatial Scales. Marine Technology Society Journal, 2010, 44: 54–64.
28. Keller, M. Schimel, D. S. Hargrove, W. W. Hoffman, F. M. A Continental Strategy for the National Ecological Observatory Network. The Ecological Society of America, 2008, 6: 282–284.
29. Michener, W. K. Porter, J. Servilla, M. Vanderbilt, K. Long Term Ecological Research and Information Management. Ecological Informatics, 2011, 6: 13–24.
30. Barseghian, D. Altintas, I. Jones, M. Crawl, D. Potter, N. Gallagher, J. Seabloom, E. Hosseini, P. Workflows and Extensions to the Kepler Scientific Workflow System to Support Environmental Sensor Data Access and Analysis. Ecological Informatics, 2010, 5: 42–50.
31. Michener, W. K. Meta-Information Concepts for Ecological Data Management. Ecological Informatics, 2006, 1: 3–7.
32. Jones, M. B. Schildhauer, M. P. Reichman, O. J. Bowers, S. The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. Annual Review of Ecology, Evolution, and Systematics, 2006, 37: 519–544.
33. Marcial, L. H. Hemminger, B. M. Scientific Data Repositories on the Web: An Initial Survey. Journal of the American Society, 2010, 61: 2029–2048.
34. Cook, R. Citations to published data sets. Flux Leifer, 2008, 1: 4–5.
35. Whitlock, M. C. McPeek, M. A. Rausher, M. D. Rieseberg, L. Moore, A. J. Data Archiving. The American Naturalist, 2010, 175: 145–146.
36. Vision, T. J. Open Data and the Social Contract of Scientific Publishing. Bioscience, 2010, 60: 330–331.
37. Kelling, S. Hochachka, W.M. Fink, D. Riedewald, M. Caruana, R. Ballard, G. Hooker, G. Data-intensive Science: A New Paradigm for Biodiversity Studies. Bioscience, 2009, 59: 613–620.
38. Buccella, A. Cechich, A. Fillottrani, P. Ontology-Driven Geographic Information Integration: A Survey of Current Approaches. Computers and Geosciences, 2009, 35: 710–723.
39. Madin, J. Bowers, S. Schildhauer, M. Krivov, S. Pennington, D. An Ontology for Describing and Synthesizing Ecological Observation Data. Ecological Informatics, 2007, 2: 279–296.
40. Kernighan, B. W. Ritchie, D. M. The C Programming Language Second. Prentice Hall,1988.
41. Hull, D. Wolstencroft, K. Stevens, R. Goble, C. Pocock, M. Li, P. Oinn, T. Taverna: A Tool for Building and Running Workflows of Services. Nucleic Acids Research, 2006, 34: 729–732.
42. Ludäscher, B. Altintas, I. Berkley, C. Higgins, D. Jaeger, E. Jones, M. Lee, E.A. Tao, J. Zhao, Y. Scientific workflow management and the Kepler system. Concurrency and Computation: Practice and Experience, 2006, 18: 1039–1065.
43. Taylor, I. Shields, M. Wang, I. Harrison, A. The Triana Workflow Environment: Architecture and Applications. In Workflows for e-Science, Springer London, London, 2007, 320–339.
44. Silva, C. T. Freire, J. Callahan, S. P. Provenance for Visualizations: Reproducibility and Beyond. Computing in Science and Engineering, 2007, 9: 82–89.

45.Kao, R. H. Gibson, C. Gallery, R. Meier, C. Barnett, D. Docherty, K. Blevins, K. Travers, P. Azuaje, E. Springer, Y. Thibault, K. Mc Kenzie, V. Keller, M. Alves, L. Hinckley, E. Parnell, J. Schimel, D. NEON Terrestrial Field Observations: Designing Continental-Scale, Standardized Sampling. Ecosphere, 2012, 3: 1–17.
46.Bahlai, C. A. vander Werf, W. O'Neal, M. Hemerik, L. Landis, D. A. Shifts in Dynamic Regime of an Invasive Lady Beetle are Linked to the Invasion and Insecticidal Management of its Prey. Ecological Applications, 2015, 25: 1807–1818.
47.Valpine, P. de, Scranton, K. Ohmart, C. P. Synchrony of Population Dynamics of Two Vineyard Arthropods Occurs at Multiple Spatial and Temporal Scales. Ecological Applications, 2010, 20: 1926–1935.
48.Wang, L. Hui, C. Sandhu, H. S. Li, Z. Zhao, Z. Population Dynamics and Associated Factors of Cereal Aphids and Armyworms Under Global Change. Scientific Reports, 2015, 5: 18801.
49.Brooks, D. R. Perry, J. N. Clark, S. J. Heard, M. S. Firbank, L. G. Holdgate, R. Mason, N. S. Shortall, C. R. Skellern, M. P. Woiwod, I. P. National-Scale Metacommunity Dynamics of Carabid Beetles in UK Farmland. Journal of Animal Ecology, 2008, 77: 265–274.
50.Allstadt, A. J. Haynes, K. J. Liebhold, A. M. Johnson, D. M. Long-term Shifts in the Cyclicity of Outbreaks of a Forest-Defoliating Insect. Oecologia, 2013, 172: 141–151.
51.Hutchison, W. D. Burkness, E. C. Mitchell, P. D. Moon, R. D. Leslie, T. W. Fleischer, S. J. Abrahamson, M. Hellmich, R. L. Kaster, L. V. Hunt, T. E. Wright, R. J. Pecinovsky, K. Rabaey, T. L. Flood, B. R. Raun, E. S. Areawide Suppression of European Corn Borer with Bt Maize Reaps Savings to Non-Bt Maize Growers. Science, 2010, 330: 222–225.
52.Wu, K.-M. Lu, Y.-H. Feng, H.-Q. Zhao, J.-Z. Suppression of Cotton Bollworm in Multiple Crops in China in Areas with Bt Toxin–Containing Cotton. Science, 2008, 321: 1676–1678.
53.Wan, P. Huang, Y. Tabashnik, B. E. Huang, M. Wu, K. The Halo Effect: Suppression of Pink Bollworm on Non-Bt Cotton by Bt Cotton in China. PLoS One, 2012, 7: e42004.
54.Carrière, Y. Ellers-Kirk, C. Sisterson, M. Antilla, L. Whitlow, M. Dennehy, T. Tabashnik, B. Long-term Regional Suppression of Pink Bollworm by *Bacillus thuringiensis* Cotton. PNAS, 2003, 100: 1519–1523.
55.Stack Whitney, K. Meehan, T. D. Kucharik, C. J. Zhu, J. Townsend, P. A. Hamilton, K. Gratton, C. Explicit Modeling of Abiotic and Landscape Factors Reveals Precipitation and Forests Associated with Aphid Abundance. Ecological Applications, 2016, 26: 2600–2610.
56.Higbee, B. Siegel, J. New Navel Orange Worm Sanitation Standards Could Reduce Almond Damage. California Agriculture, 2009, 63: 24–28.
57.Parsa, S. Ccanto, R. Rosenheim, J. A. Resource Concentration Dilutes a Key Pest in Indigenous Potato Agriculture. Ecological Applications, 2011, 21: 539–546.
58.Parsa, S. Ccanto, R. Olivera, E. Scurrah, M. Alcäzar, J. Rosenheim, J.A. Explaining Andean Potato Weevils in Relation to Local and Landscape Features: A Facilitated Ecoinformatics Approach. PLoS One, 2012, 7: e36533.
59.Meisner, M. Zaviezo, T. Rosenheim, J. Landscape Crop Composition Effects on Cotton Yield, *Lygus hesperus* Densities and Pesticide Use. Pest Management Science, 2017, 73: 232–239.
60.Rosenheim, J. A. Costs of Lygus Herbivory on Cotton Associated with Farmer Decision-Making: An Ecoinformatics Approach. Journal of Economic Entomology, 2013, 106: 1286–1293.
61.Andert, S. Burger, J. Gerowitt, B. On-farm pesticide use in four Northern German regions as influenced by farm and production conditions. Crop Protection, 2015, 75: 1–10.

62. Bartomeus, I. Ascher, J. S. Gibbs, J. Danforth, B. N. Wagner, D. L. Hedtke, S. M. Winfree, R. Historical Changes in Northeastern US Bee Pollinators Related to Shared Ecological Traits. Proceedings of the National Academy of Sciences, 2013, 110: 4656–4660.
63. Cameron, S. A. Lozier, J. D. Strange, J. P. Koch, J. B. Cordes, N. Solter, L. F. Griswold, T. L. Patterns of Widespread Decline in North American Bumble Bees. Proceedings of the National Academy of Sciences, 2011, 108: 662–667.
64. Smyth, R. R. Allee, L. L. Losey, J. E. The Status of *Coccinella undecimpunctata* (L.) (Coleoptera: Coccinellidae) in North America: An Updated Distribution from Citizen Science Data. The Coleopterists Bulletin, 2013, 67: 532–535.
65. Bahlai, C. A. Colunga-Garcia, M. Gage, S. H. Landis, D. A. The Role of Exotic Lady Beetles in the Decline of Native Ladybeetle Populations: Evidence from Long-Term Monitoring. Biological Invasions, 2015, 17: 1005–1024.
66. Bohan, D. A. Caron-Lormier, G. Muggleton, S. Raybould, A. Tamaddoni-Nezhad, A. Automated Discovery of Food Webs from Ecological Data Using Logic-Based Machine Learning. PLoS One, 2011, 6: e29028.
67. Davey, J. S. Vaughan, I. P. Andrew, King, R. Bell, J. R. Bohan, D. A. Bruford, M. W. Holland, J. M. Symondson, W.O. Intraguild Predation in Winter Wheat: Prey Choice by a Common Epigeal Carabid Consuming Spiders. Journal of Applied Ecology, 2013, 50: 271–279.
68. Parsa, S. Medina, C. Rodríguez, V. Sources of Pest Resistance in Cassava. Crop Protection, 2015, 68: 79–84.
69. Chavas, J. Shi, G. Lauer, J. The Effects of GM Technology on Maize Yield. Crop Science, 2014, 54: 1331–1335.
70. Huang, J. Hu, R. Pray, C. Qiao, F. Rozelle, S. Biotechnology as an Alternative to Chemical Pesticides: A Case Study of Bt Cotton in China. Agricultural Economics, 2003, 29: 55–67.
71. Fox, P. Hendler, J. Changing the Equation on Scientific Data Visualization. Science, 2011, 331: 705–708.
72. Isaac, N. J. B. van Strien, A. J. August, T. A. de Zeeuw, M. P. Roy, D. B. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods of Ecology and Evolution, 2014, 5: 1052–1060.
73. Chen, Y. Why, A. Batista, G. Mafra-Neto, A. Keogh, E. Flying Insect Classification with Inexpensive Sensors. Journal of Insect Behaviour, 2014, 27: 657–677.
74. Cohen, Y. Cohen, A. Hetzroni, A. Alchanatis, V. Bro- day, D. Gazit, Y. Timar, D. Spatial decision support system for Medfly control in citrus. Computers and Electronics in Agriculture, 2008, 62: 107–117.
75. Hampton, S. E. Strasser, C. A. Tewksbury, J. J. Gram, W. K. Budden, A. E. Batcheller, A. L. Duke, C. S. Porter, J. H. Big data and the future of ecology. Frontiers in Ecology and the Environment, 2013, 11: 156–162.
76. Kitchin, R. Lauriault, T. P. Small Data in the Era of Big Data. GeoJournal, 2015, 80: 463–475.