

Methodological Difficulties of Conducting Agroecological Studies from a Statistical Perspective

ANDRÉ BIANCONI,^{1,2} TOMMY DALGAARD,³ BRYAN F. J. MANLY,⁴
JOSÉ S. GOVONE,² MICHAEL J. WATTS,⁵ PETER NKALA,⁶
GUSTAVO HABERMANN,⁷ YANBO HUANG,⁸
and ADRIANE B. S. SERAPIÃO²

¹*International Academy of Ecology and Environmental Sciences,
Hong Kong, P. R. China*

²*Universidade Estadual Paulista, Unesp, IGCE, DEMAC-Departamento de Estatística,
Matemática Aplicada e Computação, São Paulo, Brazil*

³*Department of Agroecology and Environment, Aarhus University, Tjele, Denmark*

⁴*Western EcoSystems Tech. Inc., Laramie, Wyoming, USA*

⁵*School of Earth and Environmental Sciences, The University of Adelaide,
Adelaide, Australia*

⁶*Institute of Development Studies (IDS), National University of Science and
Technology (NUST), Bulawayo, Zimbabwe*

⁷*Instituto de Biociências, Universidade Estadual Paulista, Unesp, Rio Claro-SP, Brazil*

⁸*United States Department of Agriculture (USDA), Agricultural Research Service, Crop
Production Systems Research Unit, Stoneville, Mississippi, USA*

Statistical methods for analyzing agroecological data might not be able to help agroecologists to solve all of the current problems concerning crop and animal husbandry, but such methods could well help them assess, tackle, and resolve several agroecological issues in a more reliable and accurate manner. Therefore, our goal in this article is to discuss the importance of statistical tools for alternative agronomic approaches, because alternative approaches, such as organic farming, should not only be promoted by encouraging farmers to deploy agroecological techniques, but also by providing

Address correspondence to André Bianconi, Universidade Estadual Paulista, Unesp, IGCE, DEMAC, Avenida 24-A, 1515, Bela Vista, Rio Claro, São Paulo State, 13506900, Brazil. E-mail: drebianconi@yahoo.com.br

agroecologists with robust analyses based on rigorous statistical procedures.

KEYWORDS *agroecology, multivariate statistics, scale of farming, sample size, ecological agriculture*

AGROECOLOGY AND AGROECOSYSTEMS

It is a well-known fact that intensive farming practices may have a deleterious effect on the structure of ecosystems (Zhang et al. 2011). Moreover, most conventional agricultural systems are dependent on large amounts of external inputs such as fertilizers, pesticides, and other agrochemicals in order to control environmental factors and variables that govern crop yield (Kizos et al. 2011; Zhang et al. 2011). This is because natural environmental conditions may not permit farmers to set up profitable agricultural systems, and environmental conditions that are suitable for the wildlife that is adapted to them might hamper high crop yield rates. Thus, from a large-scale food production standpoint, natural ecosystems need to be changed to more appropriate systems that are able to provide farmers with high yielding crops (Primavesi 1997). In this context, homogeneous landscape, low ecological diversity, and groundwater pollution, for example, may be regarded as inevitable side effects of intensive agricultural practices.

To a large extent, the well-known types of environmental harm derived from large-scale farming systems have stimulated researchers to devise alternative agricultural practices for more than a hundred years (Ehlers 1999; Wezel et al. 2009). Moreover, it would be fair to state that the Green Revolution and all of its inherent problems marked a watershed in agronomy, because after the Green Revolution researchers started to work at formalizing ways of obtaining high crop yields without reducing the natural biodiversity of farming systems (Ehlers 1999). That is, exhausting the natural biodiversity of farming systems, without replenishing such systems, cannot represent a sustainable long-term activity or a feasible approach to sustainable cropping (Primavesi 1997; Ehlers 1999; Gliessman 2007). This is because high yield, sustainable cropping requires that farmland areas be treated as either ecosystems or parts of larger ecosystems (Zhu et al. 2012).

Agroecology, as a science and practice, was derived both from the awareness that farming systems should be regarded and characterized as ecosystems or parts of larger ecosystems, as well as from formal (academic research) and informal (farmers' practices) studies of the importance of natural biodiversity to a sustainable crop yield. There have been a large number of definitions of the term "agroecology" since the beginning of the 20th century, in different regions of the world, making a simple, precise, and concise definition difficult (Wezel and Soldat 2009; Wezel et al. 2009). However,

such definitions of agroecology convey (implicitly or explicitly) the basic meaning that large-scale or small-scale farming systems are integrated with the surrounding environment, and that animal and crop production may well be governed by several physical, chemical, biological, and socioeconomic variables and factors that may well be intertwined in an intricate fashion (Gliessman 2007). Hence, most farming methods are themselves complex methods.

Going into all of the possible definitions of agroecology lies beyond the scope of this article. However, Wezel et al. (2009), apart from providing an in-depth discussion of different definitions of agroecology, recommended that researchers should define the word in their articles in an explicit fashion. Therefore, it should be highlighted that, in this work, the fact that agroecology is devoted to analyzing farming systems in an integrated context, in which several variables and factors govern crop yield (Gliessman 2007; Zhu et al. 2012), is considerably more important than discussing different definitions of agroecology. Moreover, it is more relevant to distinguish between agroecological farming systems, in which the farmland is regarded as part of an ecosystem (Zhu et al. 2012), and conventional farming systems, in which ecology as a science is not employed.

In this article, farming systems that are handled by means of agroecology are regarded as agroecosystems. Thus, organic farming systems, for instance, will be described as agroecosystems, because it would not be possible to substitute, for example, manure or green manure for synthetic fertilizers if biological, ecological, and socioeconomic factors were not analyzed in a careful and thorough manner; that is, it would not be possible to regard organic farming as an effective method for sustainable food production at local and regional levels without assessing social, economic, and intrinsic environmental factors that control organic food production (Ronchi and Nardone 2003; Dantsis et al. 2009; Gabriel et al. 2010; Seufert et al. 2012; Sheahan et al. 2012).

Most farmers that manage agroecosystems usually have the twin goals of obtaining high crop yields (profitability) and maintaining farmland and landscape biodiversity (sustainability) (Dalgaard et al. 2006; Di Felice et al. 2012), which represents a complex problem (Di Felice et al. 2012). Overall, the dynamics of nutrients and pest management in agroecosystems, such as organic farming systems, is governed by several complex biological and ecological processes. Therefore, organic farming depends more on scientific knowledge and appropriate management techniques than conventional farming systems (Seufert et al. 2012). In this context, organic farming systems could be characterized as ecosystems (i.e., agroecosystems).

It should be emphasized that the terms “agroecology” and “organic farming” are not synonyms, but agroecology as a science includes organic farming, so that organic farming is regarded as an agroecological procedure for food production in this article. In addition, some researchers

have regarded organic farming as a synonym of sustainable agriculture or sustainability, but others have not (Rigby and Cáceres 2001). Ronchi and Nardone (2003), for example, regarded organic farming as a practice that could promote environmentally sustainable farming systems and, thus, these authors did not regard organic farming and sustainability as synonyms. In this article, both organic farming and sustainable agriculture were regarded as agroecological methods if they had been used in describing agricultural systems whereby farmland areas were considered to be ecosystems. Thus, both theoretical (such as Primavesi 1997) and descriptive works (such as Ehlers 1999) were considered to be agroecological publications. Even more mathematical articles (such as Schultz and Wieland 1997; Schultz et al. 2000; Dorigo et al. 2007) that were aimed at using complex analytical methods such as artificial neural networks (i.e., nonparametric nonlinear models) in modeling agroecosystems were regarded as agroecology papers in this article, as those papers considered farming systems to be agroecosystems that would require complex modeling techniques in order to model and describe them.

The development of agroecology would enhance subsistence farming, smallholders' incomes, and conservation of natural resources (Primavesi 1997; Gliessman 2007; Nkala et al. 2011). It has been suggested that organic farming techniques could minimize negative impacts of agricultural intensification on farmland biodiversity (Ponce et al. 2011). Nonetheless, there are a considerable number of agroecological studies that exhibit problems with methodology and rather subjective judgments (Ehlers 1999). For example, it has been suggested that organic farming methods would be able to increase crop yields in small farming systems in developing countries. In this context, Seufert et al. (2012) stated that they would not be able to negate such an assertion, but nor would they be able to support it owing to a lack of quantitative studies containing appropriate controls and methods.

Such methodological problems hinder the full extrapolation and practical application of results derived from some agroecological studies. Moreover, some apparent benefits derived from agroecological research may turn out to be superficial (Giles 2004), especially from a statistical point of view. In this context, it could be argued that the use of statistical techniques might well help agroecologists to solve such problems in a more effective manner. Therefore, this article is devoted to describing, assessing, and highlighting the importance of statistical methods for analyzing agroecosystems.

STATISTICAL METHODS AND AGROECOLOGICAL PROCESSES

The analysis of the dynamics of agroecosystems entails complex interactions between plants, animals, humans, and the environment (Altieri 1995; Dalgaard et al. 2003; Gliessman 2007; Zhu et al. 2012), and agroecological

research has taken into consideration spatial scales that are larger than plot, field, or farmland areas (Wezel and Soldat 2009). This is because both the effects of agricultural practices on the environment and the influence of environmental conditions on farming areas might not constitute local phenomena from an agroecological standpoint, and, thus, they could well be analyzed on a regional scale (Dorigo et al. 2007). Additionally, even if a relatively small farming system were analyzed, crop and animal production in such a small system would be influenced by both local and regional environmental conditions; that is, even a small farmland area may well be regarded as a complex ecosystem from an agroecological perspective. Hence, multivariate and multifactorial statistical approaches could be necessary to model and describe a small farming system in an accurate fashion.

In such approaches that consider multiple interactions, it is widely acknowledged that the outcomes of several agroecological projects are able to provide new insights into the development of feasible methodological innovations that may contribute to clarifying the underlying principles governing such intricate farming systems (Nkala et al. 2011; Zhu et al. 2012). However, agroecology still lacks clear definition, and operational and analytical tools for agroecological modeling have not developed much (Wezel and Soldat 2009). In this context, statistical methods could well be utilized as analytical tools for assessing and describing agroecological systems.

For example, pest management is an issue for both farmers and researchers, because it does represent a complex problem; that is, several activities, factors, variables, and concepts should be assessed in a careful and integrated manner in order to overcome such a problem. Thus, integrated pest management (IPM) is usually regarded as an important approach to monitoring and controlling potentially harmful organisms (arthropods, weed species, pathogens, etc.) that may impair crop yield in a significant manner. This approach includes a number of integrated techniques that have been used in order to maintain the number of pests at reduced levels that are economically insignificant from an agronomic standpoint, either using no pesticides or using such agrochemicals at reduced levels, and, thus, IPM may well be characterized as a complex issue. Furthermore, it has been described that farmers usually have an insufficient amount of information both on the problem itself and on the methods for overcoming such a problem, which makes the implementation of IPM techniques even more difficult and complex (Hashemi and Damalas 2011). Therefore, it is clear that implementing IPM techniques in an effective manner should entail the use of multivariate or multifactorial statistical analyses.

With respect to organic farming, it has been stated that organic yields are highly dependent on nutrient recycling processes in soils, since only small amounts of authorized fertilizers have usually been used in such agroecological systems (Monokrousos et al. 2008). Additionally, intentional management of ecosystem processes can reduce nutrient surplus (nutrient

losses) in agroecosystems if it is conducted in an appropriate manner (Drinkwater and Snapp 2007). However, management of agroecosystems requires a large amount of expertise in ecology, since an agroecosystem is a kind of ecosystem. Therefore, attaining low rates of nutrient loss in agroecosystems should be thought of as a complex ecological problem, and the analysis of complex ecological problems involves implementing appropriate statistical methods such as statistical models and statistical hypothesis tests.

Drinkwater and Snapp (2007), for example, asserted that it would be worth assessing the role of intrinsic ecological processes in enhancing nutrient use efficiency and nutrient balance in order to obtain sufficient levels of productivity, while still maintaining natural ecological processes in agroecosystems and, thus, maintaining the ecological function of such systems. These authors discussed several concepts and aspects of nutrient management in agroecosystems, but they did not highlight the fact that it would be essential for agroecologists to utilize complex statistical methods in modeling nutrient losses in agroecosystems. This does not mean that the authors have deliberately ignored the usefulness of statistical techniques, but this indicates that statistical methods might have been relegated to less important roles in modeling the dynamics of soil nutrients in agroecosystems. That is, would it be possible for agroecologists to devise and assess methods for modeling the dynamics of nutrient losses in agroecosystems without the intensive and effective deployment of statistical methods?

STATISTICAL PROCEDURES FOR ANALYZING AGROECOLOGICAL DATA

If researchers were to look up “agroecology” on the Internet, using search engines such as Web of Knowledge, they would find a large number of papers on agroecology. In fact, we searched for the term, using this database and found more than 352 hits (<http://apps.webofknowledge.com/>, accessed on April 15, 2012). Nonetheless, such hits represented studies that do not highlight the importance of statistics in analyzing agroecological practices; that is, even in studies that appropriately applied statistics to agroecological approaches, statistical techniques were only applied as an analytical tool and were not regarded as an essential part of conducting agroecological projects.

Regarding the use of statistical approaches to assessing the effectiveness of organic farming, the outcomes of organic methods have been frequently compared with those derived from conventional agricultural practices (Kizos et al. 2011; Seufert et al. 2012). However, it would be impossible to go over all published papers on either organic farming or other agroecological farming techniques, because some (or most) of these papers might have been published in obscure journals and might not be in English. Others have been

published as book chapters. Hence, the following paragraphs contain information on some agroecology papers that were published by peer-reviewed journals, and such papers were regarded as representative of their respective fields of research. It is important to highlight that, in this article, describing the usefulness of statistical methods in analyzing agroecological data was regarded as more important than providing a complete list of every statistical procedure that have already been used in agroecology.

Uni- and Multivariate Statistical Methods for Analyzing Agroecological Processes

UNIVARIATE METHODS

Most agroecological approaches to analyzing agroecosystems could well require the implementation of complex statistical methods, since most agroecosystems might be regarded as complex farming systems. However, there may be some practical situations in which the outcomes of alternative agricultural methods such as organic farming could be effectively analyzed by means of simple statistical procedures. For example, classical statistical methods such as univariate analysis of variance and simple linear regression may help researchers to analyze the dynamics of agroecosystems in an effective manner. Tu et al. (2006), using analysis of variance ('split-split' ANOVA), stated that organic mulching may have beneficial effects on soil microbes, probably reducing the extreme fluctuations in soil moisture and temperature. Levin (2007), for instance, apart from assessing changes in landscape composition after farmland conversion to organic farming, investigated differences in landscape composition between conventional and organic farming systems at both sampling (local) level and national level in Denmark. Levin made effective use of analysis of variance in combination with linear regression (simple statistical procedures) in analyzing relationships between farm properties and landscape composition.

Dantsis et al. (2009) used simple descriptive statistics in order to assess the interaction between organic farming and the agro-industrial complex in Central Macedonia (Greece), and most of their conclusions were based on such simple statistics. In this context, it should be emphasized that researchers should draw their conclusions from the statistical methods that they have implemented; that is, it would be unreasonable and ineffective for agroecologists to employ simple descriptive statistics if they were interested in disentangling involved relationships between input and output variables.

In the context of alternative agricultural systems, Kizos et al. (2011) used face-to-face interviews and a semi-structured questionnaire, and they took a large number of factors and variables into consideration. However, they used simple univariate statistical procedures such as *t* test, chi-square test, and simple linear correlation in order to compare their data sets. To some

extent, it might have been useful to analyze these same data sets by means of multivariate statistical techniques in order to compare the outcomes of multivariate approaches with those derived from the univariate methods that they deployed.

Panneerselvam et al. (2011) compared organic farming systems with conventional agricultural systems using data derived from Indian farms, and these authors mentioned that they implemented a general linear model in order to assess farm production, crop yield, input costs, and income of such farming systems. A considerable amount of their discussion was based on evaluating the presence of significant differences, or the lack thereof, between organic and conventional farms, and the authors stated that organic farming might be comparable with conventional agricultural methods with respect to yield in some parts of India. Thus, it was important for them to utilize the general linear model to validate their statements and inferences. However, the authors did not rigorously describe the general linear model, as they only presented a single equation in their paper. This does not mean that the statistical method had been used in an inaccurate manner, but this does mean that statistical tests might well represent a less relevant part of the experimental design, and, thus, a careful description of the usefulness of a statistical method could well be set aside without affecting the basic information provided by an agroecology paper.

The fact is that analysis of variance, polynomial regression, and multiple linear regression, for example, are all general linear models. Hence, implementing a general linear model involves describing which linear model was used and why it was used. Such a description should be as important as describing the study area. In addition, in Panneerselvam et al. (2011), in which small-scale farming was taken into consideration, data collection entailed the use of semi-structured questionnaires. From a statistical standpoint, it would be worth mentioning what type of data might be derived from this kind of data collection; that is, would it be expected to obtain non-normal data? Are missing values common in this type of survey? If no information on these issues is available, readers will need to guess the answers.

In the context of IPM, Hashemi and Damalas (2011) stated that socio-economic characteristics of farmers such as personal beliefs, perceptions concerning the efficiency of pesticides, and personal preference have an important influence on the choice of IPM practices; that is, apart from being a complex agroecological problem, effective IPM methods need to take other non-ecological variables and factors into account. These authors also used data derived from face-to-face interviews (structured questionnaires) with 90 smallholding (small-scale) farmers in Iran. Apart from using descriptive statistics and a couple of nonparametric and parametric hypothesis tests (such as the Mann-Whitney test and t test), they implemented regression analysis. Moreover, they did describe the statistical assumption of normality in order to justify the use of t tests for one portion of their data set and

Mann-Whitney for the other, and such statistical procedures were used to corroborate their statements in a clear manner. That is, most of the conclusions provided by their paper would not have been meaningful if they had left statistical techniques out of their methods for assessing the information that they described.

To some extent, this kind of detailed description concerning the assumption of normality is not present in agroecology papers. However, the fact that specific data did not follow the normal distribution, for example, may be as important as other characteristics described by such data, since this kind of outcome may conceivably raise questions such as “What made the data depart from a normal distribution?”, “Does non-normality represent an intrinsic feature to these data?”, and “Are well-known tests for normality appropriate for agroecological data?”

MORE COMPLEX STATISTICAL PROCEDURES FOR ANALYZING AGROECOLOGICAL DATA SETS

Assessing a series of specific hypotheses, Drinkwater et al. (1995) used classical and well-known multivariate statistical approaches such as principal components analysis (PCA) and canonical discriminant analysis (see Manly 2005, for details concerning these statistical approaches) to identify relationships between agronomic and community level properties of organic and conventional systems. Monokrousos et al. (2008) assessed the influence of both seasonal management phase and management regime (i.e., organic and conventional agricultural practices) on soil variables, and they ascertained whether fields that had been under organic cultivation longer than other fields could be distinguished by their soil quality. These authors utilized PCA, classification trees, and artificial neural networks. PCA in combination with the other two analytical tools enabled the authors to state that the effect of management steps was much more pronounced than that derived from different management regimes. Furthermore, their study did not support the concept that soil quality improves as the duration of organic cultivation increases. Thus, the changes in soil biochemical variables with years of organic management were more complex than initially thought. In this context, Nautiyal et al. (2010) analyzed soil quality and microbial functional diversity in an organic farming system, and PCA was also used in an effective manner.

Di Felice et al. (2012) asserted that the relationship between cropping diversity and agroecosystem sustainability was connected with the presence of mixed farming (i.e., containing both animal and crop husbandry), or the lack thereof, without association with organic or conventional agricultural practices. They also stated that the economic performance of organic farming was similar to that derived from conventional farming systems. They used PCA in combination with canonical redundancy analysis (an extension of

the multiple linear regression that may contain several output variables), and it would not have been possible for them to reach such conclusions if they had not employed multivariate statistical procedures.

In addition, both univariate tests and multivariate methods have been used in evaluating intercropping designs by means of agro-economic indices. For instance, Bezerra Neto et al. (2012) used uni- and multivariate approaches (such as uni- and multivariate analysis of variance) to analyze such indices. Apart from highlighting the importance of multidimensional analysis (multidimensional vectors) for assessing the feasibility of intercropping designs, these authors described how they assessed the underlying statistical assumptions for the tests that they utilized in evaluating intercropping designs.

Generalized linear models (nonlinear regression models, see Kutner et al. 2004, for details), chi-square test, simple correlation analysis, and analysis of variance constituted the statistical techniques that were deployed in characterizing and assessing the effects of organic farming on biodiversity in a cereal farming system in central Spain (Ponce et al. 2011). The outcomes of organic farming were compared with those derived from conventional agricultural systems, using data on weed and arthropod communities (e.g., weed richness and arthropod biomass). The authors stated that organic farming techniques might contribute to preserving natural biodiversity in a dryland cereal agroecosystem. Their statements would have sounded highly subjective if they had set aside information on the statistical methods that they used.

Regarding the assessment of relationships between crop species and weeds in agroecosystems, it should be noted that if weed species were removed from agroecosystems, essential ecological webs such as food webs would be modified and, thus, the biodiversity of such farming systems might be impaired. On the other hand, weed species could reduce the economic output of an agroecosystem in a significant manner (Debeljak et al. 2011). Hence, there should be an ecological compromise between crop species and weeds in agroecosystems. Debeljak et al. (2011) highlighted that occasional and intermittent evaluations of the relationships between crops and weeds would not be appropriate, because there might be interactions between these types of vegetation that would stand out only at particular points in a time series, and that would have effects on other developmental stages of crop species, but such interactions might not be conspicuous at subsequent time points. Hence, statistical methods for assessing time series analysis can represent an important means of modeling relationships between crops and weeds in agroecosystems.

However, it should be emphasized that quantifying and comparing different time series may well be a nontrivial task, and both the implementation of methods for analyzing time series data and the assessment of

their outcomes are considerably more complex than implementing most general linear statistical models and hypothesis tests. This is because ecological data such as agroecological data might be derived from intermittent measurements, and might well contain, for example, different lengths of time series, missing values, and both numeric and discrete attributes. On top of that, each time point could be different from the others in a nonlinear fashion (Debeljak et al. 2011). This all makes the implementation of time series analysis a nontrivial task.

In this context, time series analysis may well clarify involved relationships between crop species and weeds, and owing to the fact that agroecological data are themselves complex and that time series analysis is also complex, agroecologists should have been using methods for analyzing time series on a regular basis to improve and facilitate their implementation. However, such methods have not been used in describing agroecological phenomena as frequently as they should have been. Even though Debeljak et al. (2011) assessed the feasibility of predictive clustering trees (i.e., an approach to analyzing time series data that combines clustering and prediction) in analyzing large agroecological data sets derived from crop and weed species (percentage crop and weed cover), it could be stated that agroecologists have not been assessing the effectiveness of complex methods for analyzing and modeling agroecological time series data.

As for meta-analysis, it is important to note that the outcomes derived from independent agroecological studies can be combined into one single statistical analysis; that is, in order to compare organic farming with conventional farming systems, agroecologists can use data sets derived from farming systems that may be located in different regions. Meta-analysis represents the use of statistical methods for both combining data sets that have been derived from different experiments into one single experiment, and analyzing them as if they had been collected from the same site. Obviously, it may be difficult both to account for variance values derived from different sites and to allow for biases in an accurate manner. Therefore, an accurate and effective selection criterion for including samples should be used, including a solution to potential problems concerning scale of farming.

For example, Seufert et al. (2012) utilized meta-analysis in an effective manner. Based on a comprehensive literature search, they compared the yield performance of organic farming with the outcomes derived from conventional agricultural systems on a global scale. In outline, they found that organic yields were usually lower than conventional yields, but differences between organic and conventional yields were highly contextual. Additionally, although they were able to identify some factors that drive variations in organic yields, they were not able to identify several other potentially important factors owing to a lack of appropriate research.

WOULD IT BE POSSIBLE FOR AGROECOLOGISTS TO COPE WITH EXTREMELY LARGE NUMBERS OF VARIABLES AND FACTORS?

The relationships and interactions between output (target) and input (explanatory) variables may represent agroecological phenomena that are both multivariate and nonlinear (Delmotte et al. 2011). Furthermore, whether or not researchers regard a specific farmland area as an intricate agricultural ecosystem, such a farming area could well be an intrinsically complex ecological system, in which several variables and factors often directly influence food production (Drinkwater et al. 1995; Schultz and Wieland 1997; Schultz et al. 2000; Nelson 2004; Dorigo et al. 2007; Best 2008). Thus, researchers may need to use multivariate statistical methods and nonlinear statistical models for analyzing agroecological phenomena, and these techniques are considerably more complex than classical statistical procedures such as analysis of variance and multiple linear regression.

Although crop yield in agroecosystems can be governed by several factors and variables, it would be impractical to think that an extremely large number of input variables could be used in fitting one single mathematical or statistical model in a simple fashion. For example, if agroecologists were to implement one single model containing a large number of input variables (e.g., 50 variables), fitting such a model to the multivariate data set would be a nontrivial task. On top of that, it would be impractical to try to explain the outcomes of such a modeling procedure in a simple manner; that is, it should not be expected to obtain one single predicted equation that would contain all the information conveyed by the 50 input variables in a simple and accurate manner.

There may be, however, several practical situations in which a large number of variables should be analyzed. For example, Delmotte et al. (2011) used classification and regression trees in order to identify factors that would be able to determine differences between organic farming and conventional farming systems with respect to rice production in a Mediterranean region (France), and they used 45 input variables that were split into four groups of explanatory variables. These statistical methods were effective in assessing long-term variability in rice yield regarding both conventional and organic farming systems. In addition, classification and regression trees, which are nonparametric regression techniques, entail the use of several iterative processes, as is the case with neural network models (for details, see, Schultz and Wieland 1997; Shultz et al. 2000; Kutner et al. 2004), and they may not be regarded as statistical models in the strict sense. Unlike classical linear regression models such as simple and multiple linear regression models, regression trees do not yield a single predicted equation derived from the input data. Therefore, it is possible for agroecologists to utilize a large number of variables in implementing analytical models, but such models may not provide them with a final predicted equation where the

relationship between independent (input) and dependent (output) variables stands out.

Similarly, it would be difficult for agroecologists to combine the effects of organic farming practices on farm margins, soil fertility, and rural employment into one single accurate measure or model (Rigby and Cáceres 2001). In this context, it should be noted that there are several well-known statistical methods (e.g., principal components analysis) that would enable researchers to reduce the dimensionality (the number of variables) of a problem in order to remove both irrelevant and redundant variables from the analysis (see, e.g., Manly 2005). That is, complexity should not be regarded as impracticability, because agroecological data are intrinsically complex and, thus, it should be expected that analyzing them might not represent a trivial task.

Therefore, both fitting complex statistical models (e.g., nonlinear regression models) to intricate data sets such as agroecological data, as well as implementing multivariate hypothesis tests, may well require a large amount of statistical expertise in order to ascertain whether or not such models and tests are being used in an accurate and effective manner. For example, oversimplification of problems and facts, inappropriate parameterization, and inaccurate predictions have all been regarded as major problems for modeling agroecosystems, and all of these problems could not be overcome without accurate use of statistical models. In addition, a constant process of monitoring, assessment, and reassessment is necessary for agroecologists both to characterize a farming system as a sustainable agroecosystem and to recommend that a particular agroecological practice should be maintained (Rigby and Cáceres 2001), and it would be essential for researchers to implement statistical methods in order to make appropriate recommendations.

Moreover, owing to the importance and complexity of agroecosystems, it is essential for agroecologists to bring in statisticians to help them to analyze agroecological data, because statisticians would provide agroecologists with expertise in analyzing complex data sets such as agroecological data in a more accurate manner, and agroecologists would provide statisticians with information on the dynamics of agroecosystems, which would foster collaborative research on the use of statistics in analyzing agroecosystems. For example, although the inappropriateness of conventional agricultural practices for setting up sustainable farming systems has been described and emphasized, marked differences between approaches to attaining sustainability have been described (Rigby and Cáceres 2001), and it would not be possible to characterize any differences between methods as significant without utilizing statistical methods. That is, if researchers were to describe differences between methods as significant or marked without the appropriate utilization of statistical methods, this would sound vague and inconclusive. However, to the best of our knowledge, this has not been highlighted by agroecologists.

ADDITIONAL CONSIDERATIONS

The previous sections showed that statistical and mathematical modeling techniques have been used as important tools in analyzing agroecological data. Therefore, agroecologists could conceivably suppose that a straightforward statistical solution might exist for every practical situation, owing to the relatively large number of papers published by high impact-factor journals in which statistical methods were used for analyzing agroecological issues. In fact, statistical methods have been usefully applied to some published analyses of organic production techniques. Nevertheless, rigorous comparisons between organic and conventional production systems, for example, are not as common as they could be, because it is difficult to meet the requirements of traditional statistical approaches regarding the number of replicates (Drinkwater et al. 1995). Additionally, even though agroecological approaches such as organic farming possess clear potential benefits (Macilwain 2004; Gabriel et al. 2010), long-term comparative analyses need to be conducted to ensure that arguments in favor of the sustainability of agroecosystems are convincing and reliable (Macilwain 2004).

In this context, it is important to note that the general usefulness of statistical tests or mathematical modeling approaches is not emphasized in some types of agroecology publications (e.g., Primavesi 1997). Although such works are usually present in low-impact or regional journals, they are normally written in accessible language and may be deemed more applicable than high-impact papers from the standpoint of farmers. This rather serious problem has not yet been approached in a proper manner. Hence, studies on alternative agricultural methods should not just be aimed at encouraging farmers to deploy agroecological techniques, but also at providing agroecologists with robust methods based on rigorous statistical procedures. Moreover, agroecologists should have been endeavoring to clarify some agroecological issues such as scale of farming, sample size, and duration of experiments. However, only a small number of articles have discussed such issues.

Scale of Farming

Scale of agroecological farming is a key issue in agroecological research, since there might well be differences between large-scale agroecosystems and small-scale ones with respect to both ecological and socioeconomic factors that might govern the dynamics of agroecosystems (Dalgaard et al. 2003; Gabriel et al. 2010; Rigby and Cáceres 2001). For instance, Stark et al. (2004) stated that it would be important to determine spatial variability within relatively small cropping farms in order to obtain representative soil samples. Their statement might well be useful for conducting agroecological studies, since these authors used soil samples derived from both conventional and

organic farming systems in order to determine variation in biotic and abiotic soil parameters.

In India as well as in other developing countries, marginal and smallholding farmers have undergone harsh food insecurity conditions, because they usually have no access to credit either from banks or other financial institutions that would provide them with reasonable rates of interest. Thus, such farmers often need to resort to other kinds of borrowing that may well charge them higher rates of interest. Therefore, input costs represent a big issue for small farmers, so that the minimization of these costs is as essential as the maximization of income. In addition, Panneerselvam et al. (2011) highlighted that little information has been published on food security of small organic farmers in India, and the same is true of other developing countries. This all makes smallholding agroecological systems susceptible to a large number of ecological and socioeconomic factors and variables. Hence, such small-scale agroecosystems should have been analyzed in a much more measured and careful way.

Additionally, analyzing the effects of agricultural practices on ecosystems at field, farm, and landscape scales could well require the use of complex statistical techniques. In this context, Gabriel et al. (2010) compared organic farming with conventional agricultural practices in England. Apart from using principal components analysis, they implemented general and generalized linear mixed effect models in both analyzing scale effects of farming practices on biodiversity and identifying scales that could be regarded as most appropriate for maximizing biodiversity that would be beneficial to agroecosystems. The implementation of generalized linear models (see Kutner et al. 2004, for details) is considerably more complex than most general linear models. However, the use of such complex models enabled the authors to conclude that organic farming might well have positive effects on wildlife at both farmland and landscape scales, and that some taxa responded positively to farm management and others responded negatively. Hence, the interactions between farming practices and different spatial scales should be regarded as complex, and complex statistical models might well be useful in analyzing such interactions.

To a large extent, even though several studies have compared organic farming with conventional agricultural systems, little research into comparing the yields of organic and conventional farming systems on a global scale has been done (Seufert et al. 2012).

Small Sample Sizes in Agroecological Research

Sample size is also a key issue for implementing agroecological techniques in an accurate and effective manner. If agroecologists were to use relatively small sample sizes in analyzing agroecological data sets, they might

well face a number of problems. For example, it might not be possible to ascertain whether or not the agroecological data were drawn from a normally distributed population. Moreover, it is a well-known fact that the power-efficiency of any test increases as sample sizes increase. In other words, the larger the sample size, the more accurate the outcomes of statistical tests regarding both parametric and nonparametric hypothesis tests. In this context, it might be suggested that if very small data sets were available, a parametric test would yield an inaccurate outcome (see Sprent 1993, for details), and, thus, a nonparametric statistical test could be taken into consideration. However, even the power-efficiency of nonparametric hypothesis tests can be reduced by very small samples sizes (e.g., ≤ 10).

For instance, Bianconi et al. (2008) described a hypothetical agroecological example that was aimed at comparing crops that were sprayed with a large amount of pesticide with crops that were sprayed with either a small amount of pesticide or no pesticide. These authors showed that the power-efficiency of Dunn's test, a well-known nonparametric multiple comparison test, might well be reduced by very small sample sizes (≤ 10); that is, this nonparametric procedure was incapable of distinguishing between treatments that were very different from one another. Similarly, Seufert et al. (2012) stated that small sample sizes (e.g., 8 and 12) might have affected the accuracy of results derived from comparing organic yields with conventional farming yields in developing countries, because wide confidence intervals resulted from such small sample sizes, which might have reduced the power efficiency of statistical tests.

Furthermore, the combination of relatively small sample sizes and several variables could well generate a multivariate problem that could not be analyzed by means of multivariate statistical approaches due to such small samples.

Duration of Agroecological Experiments

As to the duration of experiments, this is also an important issue in analyzing agroecosystems. For example, 12-month experiments might be regarded as inappropriate for assessing the state of equilibrium of organic farming systems. However, researchers need to be able to depict the initial responses of farmland to agroecological methods in an accurate manner (Ponce et al. 2011), because farmers might not be willing to wait for researchers to show them the outcomes derived from long-running experiments. Hence, it is essential for agroecologists to endeavor to obtain accurate outcomes from experiments conducted over relatively short periods of time (e.g., one-year experiments). This is because a relatively poor farmer would not agree to change their conventional methods to alternative agroecological methods if after a couple of years they could be informed that their smallholding would

not be profitable in the long term. That is, agroecologists need to be able to provide farmers with accurate information regarding the dynamics of farmland, and it is essential to use statistical methods to obtain accurate outcomes and provide farmers with useful information.

Why Not Use Statistical Methods as a Basis for Formalizing Agroecological Techniques?

Complex interactions between ecological and socioeconomic processes govern agricultural production efficiency (Primavesi 1997; Ehlers 1999; Nelson 2004; Best 2008), irrespective of differences in the principles behind the management of different agroecosystems. For example, in Brazil, rural poverty and disparities in farmland tenure models are relevant issues (Wezel et al. 2009) that should be translated into socioeconomic variables in order to model a typical Brazilian agroecosystem effectively and to predict an output variable (e.g., crop biomass) accurately. Similarly, Seufert et al. (2012) mentioned that a key issue for smallholders in developing countries is whether organic yield would be able both to alleviate poverty and to increase food security.

In several parts of the world, there is much more scientific knowledge available on the use of agricultural methods for implementing and managing conventional farming systems than on the use of agroecological procedures. For instance, Ronchi and Nardone (2003) stated that little technical information on the correct management and adoption of genotypes is available to Mediterranean farmers, which might well preclude the employment of innovative agricultural systems. Similarly, Delmotte et al. (2011) asserted that a large amount of scientific knowledge and technical recommendation on the implementation of agricultural methods is devoted to the management of conventional farming systems concerning Mediterranean rice yield. However, a change from conventional to alternative agricultural methods such as organic farming would require a large amount of both knowledge and institutional support in order to help farmers realize the importance of agroecological and socioeconomic requirements (Oelofse et al. 2011).

Largely, the use of appropriate, effective statistical procedures would help agroecologists to improve the current knowledge available on agroecological techniques. However, it should be noted that it may take time for agroecology as a science to be turned into alternative solutions to practical problems, even if adequate statistical methods were utilized whenever possible in analyzing alternative production systems. Nonetheless, the utilization of external industrial inputs could well be reduced if they were proved less effective than their alternative agroecological counterparts were. Therefore, farmers, agroecologists, and statisticians should collaborate fully during the research process. In this context, statistics could substitute for

subjective agroecological analysis. In addition, the fact that higher education and research programs in agroecology have been launched in the United States and Europe (Wezel et al. 2009) could encourage agroecologists to highlight the importance of statistical methods in their papers and, thus, scientific knowledge on agroecological techniques could be increased and enhanced.

FINAL REMARKS

As previously mentioned, there has been much debate over the appropriateness of terms such as “agroecology,” “organic farming,” and “sustainable agriculture.” In this article, highlighting the importance of statistical methods for improving current agroecological methods was more important than restricting ourselves to discussing the meaning of such terms, since all of these terms have been used in describing systems in which every farmland area should be regarded as either an ecosystem or a portion of a larger ecosystem. This does not mean that discussions on the appropriateness of using such terms in different contexts should be considered irrelevant. In fact, this does mean that these kinds of discussion should be based on solid outcomes derived from using statistical methods in analyzing the dynamics of different agroecological farming systems.

As a science and a discipline, agroecology is predicated upon ecological principles, and attaining sustainable development models for agroecosystem management is a multifaceted problem that should be solved by means of multidisciplinary assessments (Dalgaard et al. 2003; Gliessman 2007; Zhu et al. 2012). The importance of ecology, sociology, and economy, for example, has been highlighted in almost every agroecology paper, even though different papers might have laid different amounts of emphasis on such sciences. Hence, in this article, the importance of statistical methods for achieving sustainability in agroecological farming systems was emphasized, because the fact that statistics is important may well be agreed upon by virtually everyone, but the fact that statistical procedures represent a means of improving current agroecological approaches has not been adequately discussed. Therefore, additional attempts to reinforce the idea that statistics may well be essential for agroecologists are both feasible and necessary.

In addition, it seems that agroecology papers that have been published in high-impact peer-reviewed journals (e.g., most of the papers cited in this article), apart from using some kind of statistical method, have taken into consideration socioeconomic aspects of agroecological farming. Even when such papers were devoted to ecological and agronomic relationships between input and output variables, some kinds of socioeconomic factors were either taken into account or mentioned. On the other hand, several

agroecological works that were aimed at assessing socioeconomic and political factors (e.g., Primavesi 1997) have not discussed the effective utilization of statistical methods for analyzing agroecosystems.

In this context, it has been asserted that research into methods for sustainable agriculture has been neglected, since economic and institutional interests have favored research into conventional agro-industrial approaches to food production (Altieri 2002). However, it could be argued that the insularity of some agroecologists might have been hampering the development of agroecology as a formalized and complex science; that is, if agroecologists were to assess and describe the dynamics of agroecosystems by means of effective methods such as statistical methods, then they would be able to devise agroecological methods for growing crops as profitably as possible and, thus, their agroecological projects could be subsidized in a more frequent and effective fashion.

It should be noted that statistical methods for analyzing agroecological data might not be able to help agroecologists to solve all of the current problems concerning crop and animal husbandry. However, such methods could well help agroecologists to assess, tackle, and resolve several agroecological issues in a more reliable and accurate manner, because the less subjective the agroecological approach is, the more reliable its outcomes will be.

REFERENCES

- Altieri, M. A. 1995. *Agroecology: The science of sustainable agriculture*. Boulder, CO: Westview Press.
- Altieri, M. A. 2002. Agroecology: The science of natural resource management for poor farmers in marginal environments. *Agriculture, Ecosystems & Environment* 93: 1–24.
- Best, H. 2008. Organic agriculture and the conventionalization hypothesis: A case study from West Germany. *Agriculture and Human Values* 25: 95–106.
- Bezerra Neto, F., V. C. N. Porto, E. G. Gomes, A. B. Cecílio Filho, and J. N. Moreira. 2012. Assessment of agroeconomic indices in polycultures of lettuce, rocket and carrot through uni- and multivariate approaches in semi-arid Brazil. *Ecological Indicators* 14: 11–17.
- Bianconi, A., J. S. Govone, C. J. Von Zuben, A. C. S. Pião, M. A. Pizano, and L. F. Alberti. 2008. Transformação de dados e implicações da utilização do teste de Kruskal-Wallis em pesquisas agroecológicas. *Pesticidas: Revista de Ecotoxicologia e Meio Ambiente* 18: 27–34.
- Dalgaard, T., N. J. Hutchings, and J. R. Porter. 2003. Agroecology, scaling and interdisciplinarity. *Agriculture, Ecosystems & Environment* 100: 39–51.
- Dalgaard, T., S. Ferrari, and M. Rambonilaza. 2006. Introduction: Features of environmental sustainability in agriculture: some conceptual and operational issues. *International Journal of Agricultural Resources, Governance and Ecology* 5: 107–115.

- Dantsis, T., A. Loumou, and C. Giourga. 2009. Organic agriculture's approach towards sustainability; its relationship with the agro-industrial complex, a case study in Central Macedonia, Greece. *Journal of Agricultural and Environmental Ethics* 22: 197–216.
- Debeljak, M., G. R. Squire, D. Kocev, C. Hawes, M. W. Young, and S. Džeroski. 2011. Analysis of time series data on agroecosystem vegetation using predictive clustering trees. *Ecological Modelling* 222: 2524–2529.
- Delmotte, S., P. Titonell, J. C. Mouret, R. Hammond, and S. Lopez-Ridaura. 2011. On farm assessment of rice yield variability and productivity gaps between organic and conventional cropping systems under Mediterranean climate. *European Journal of Agronomy* 35: 223–236.
- Di Felice, V., R. Mancinelli, R. Proulx, and E. Campiglia. 2012. A multivariate analysis for evaluating the environmental and economical aspects of agroecosystem sustainability in central Italy. *Journal of Environmental Management* 98: 119–126.
- Dorigo, W. A., R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* 9: 165–193.
- Drinkwater, L. E., and S. S. Snapp. 2007. Nutrients in agroecosystems: Rethinking the management paradigm. *Advances in Agronomy* 92: 163–186.
- Drinkwater, L. E., D. K. Letourneau, F. Workneh, A. H. C. van Bruggen, and C. Shennan. 1995. Fundamental differences between conventional and organic tomato agroecosystems in California. *Ecological Applications* 5: 1098–1112.
- Ehlers, E. 1999. *Agricultura sustentável: Origens e perspectivas de um novo paradigma* [Sustainable agriculture: Origins and perspectives of a new paradigm]. Guaíba: Livraria e Editora Agropecuária.
- Gabriel, D., S. M. Sait, J. A. Hodgson, U. Schmutz, W. E. Kunin, and T. G. Benton. 2010. Scale matters: The impact of organic farming on biodiversity at different spatial scales. *Ecology Letters* 13: 858–869.
- Giles, J. 2004. Is organic food better for us? *Nature* 428: 796–797.
- Gliessman, S. R. 2007. *Agroecology—The ecology of sustainable food systems*, 2nd ed. Boca Raton, FL: CRC/Taylor&Francis.
- Hashemi, S. M., and C. A. Damalas. 2011. Farmers' perceptions of pesticide efficacy: Reflections on the importance of pest management practices adoption. *Journal of Sustainable Agriculture* 35: 69–85.
- Kizos, T., G. Veikontis, and J. I. Marín-Guirao. 2011. Comparison of organic and integrated farming systems: The case of sultana table grapes in Korinthos, Greece. *Journal of Sustainable Agriculture* 35: 27–47.
- Kutner, M. H., C. J. Nachtsheim, J. Neter. 2004. *Applied Linear Regression Models*, 4th edition. New York: Irwin.
- Levin, G. 2007. Relationships between Danish organic farming and landscape composition. *Agriculture, Ecosystems & Environment* 120: 330–344.
- Macilwain, C. 2004. Is organic farming better for the environment? *Nature* 428: 797–798.

- Manly, B. F. J. 2005. *Multivariate statistical methods: A primer*, 3rd ed. Boca Raton, FL: Chapman & Hall.
- Monokrousos, N., E. M. Papatheodorou, and G. P. Stamou. 2008. The response of soil biochemical variables to organic and conventional cultivation of *Asparagus* sp. *Soil Biology & Biochemistry* 40: 198–206.
- Nautiyal, C. S., P. S. Chauhan, and C. R. Bhatia. 2010. Changes in soil physico-chemical properties and microbial functional diversity due to 14 years of conversion of grassland to organic agriculture in semi-arid agroecosystem. *Soil & Tillage Research* 109: 55–60.
- Nelson, L. 2004. What is organic farming? *Nature* 428: 796.
- Nkala, P., N. Mango, and P. Zikhali. 2011. Conservation agriculture and livelihoods of smallholder farmers in central Mozambique. *Journal of Sustainable Agriculture* 35: 757–779.
- Oelofse, M., H. Høgh-Jensen, L. S. Abreu, G. F. Almeida, A. El-Araby, Q. Y. Hui, T. Sultan, and A. de Neergaard. 2011. Organic farm conventionalisation and farmer practices in China, Brazil and Egypt. *Agronomy for Sustainable Development* 31:689–698.
- Panneerselvam, P., J. E. Hermansen, and N. Halberg. 2011. Food security of small holding farmers: Comparing organic and conventional systems in India. *Journal of Sustainable Agriculture* 35: 48–68.
- Ponce, C., C. Bravo, D. G. de León, M. Magaña, and J. C. Alonso. 2011. Effects of organic farming on plant and arthropod communities: A case study in Mediterranean dryland cereal. *Agriculture, Ecosystems & Environment* 141: 193–201.
- Primavesi, A. 1997. *Agroecologia: Ecosfera, tecnosfera e agricultura* [Agroecology: Ecosphere, technosphere and agriculture]. São Paulo: Livraria Nobel S.A.
- Rigby, D., and D. Cáceres. 2001. Organic farming and the sustainability of agricultural systems. *Agricultural Systems* 68: 21–40.
- Ronchi, B., and A. Nardone. 2003. Contribution of organic farming to increase sustainability of Mediterranean small ruminants livestock systems. *Livestock Production Science* 80(1–2): 17–31.
- Schultz, A., and R. Wieland. 1997. The use of neural networks in agroecological modelling. *Computers and Electronics in Agriculture* 18(2–3): 73–90.
- Schultz, A., R. Wieland, and G. Lutze. 2000. Neural networks in agroecological modelling—stylish application or helpful tool. *Computers and Electronics in Agriculture* 29: 73–97.
- Seufert, V., N. Ramankutty, and J. A. Foley. 2012. Comparing the yields of organic and conventional agriculture. *Nature* 485: 229–232.
- Sheahan, C. M., D. B. Bray, M. G. Bhat, and K. Jayachandran. 2012. Ecological, economic, and organizational dimensions of organic farming in Miami-Dade County. *Journal of Sustainable Agriculture* 36: 83–105.
- Sprent, P. 1993. *Applied nonparametric statistical methods*, 2nd ed. London: Chapman & Hall.
- Stark, C. H. E., L. M. Condrón, A. Stewart, H. J. Di, and M. O’Callaghan. 2004. Small-scale spatial variability of selected soil biological properties. *Soil Biology & Biochemistry* 36: 601–608.

- Tu, C., J. B. Ristaino, and S. Hu. 2006. Soil microbial biomass and activity in organic tomato farming systems: Effects of organic inputs and straw mulching. *Soil Biology and Biochemistry* 38: 247–255.
- Wezel, A., and V. Soldat. 2009. A quantitative and qualitative historical analysis of the scientific discipline of agroecology. *International Journal of Agricultural Sustainability* 7: 3–18.
- Wezel, A., S. Bellon, T. Doré, C. Francis, D. Vallod, and C. David. 2009. Agroecology as a science, a movement and a practice. A review. *Agronomy for Sustainable Development* 29: 503–515.
- Zhang, W. J., F. B. Jiang, and J. F. Ou. 2011. Global pesticide consumption and pollution: with China as a focus. *Proceedings of the International Academy of Ecology and Environmental Sciences* 1: 125–144.
- Zhu, W., S. Wang, and C. D. Caldwell. 2012. Pathways of assessing agroecosystem health and agroecosystem management. *Acta Ecologica Sinica* 32: 9–17.