Biology Letters
December 2014, Volume 10, Issue 12, Pages 1-4
http://dx.doi.org/10.1098/rsbl.2014.0698
http://archimer.ifremer.fr/doc/00249/36026/

Statistical ecology comes of age

Gimenez Olivier [1, *], Buckland Stephen T. [2], Morgan Byron J. T. [3], Bez Nicolas [4], Bertrand Sophie [4], Choquet Rémi [1], Dray Stephane [5], Etienne Marie-Pierre [6], Fewster Rachel [7], Gosselin Frédéric [8], Merigot Bastien [9], Monestiez Pascal [10], Morales Juan M. [11], Mortier Frédéric [12], Munoz François [13], Ovaskainen Otso [14], Pavoine Sandrine [15, 16], Pradel Roger [1], Schurr Frank M. [17], Thomas Len [2], Thuiller Wilfried [18], Trenkel Verena [19], De Valpine Perry [20], Rexstad Eric [2]

[1] CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
[2] Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews KY16 9LZ, UK
[3] School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NF, UK
[4] IRD, UMR EME 212, Sète, France
[5] Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de 18 Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France
[6] AgroParisTech, UMR MIA 518, Paris, France
[7] Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
[8] Irstea, UR EFNO, Centre de Nogent-sur-Vernisson, 45290 Nogent-sur-Vernisson, France
[9] Université Montpellier 2, UMR EME 212, Sète, France
[10] INRA, BioSP, Avignon, France
[11] Laboratorio Ecotono, CRUB, INIBIOMA-CONICET, Bariloche, Argentina
[12] UPR Bsef, CIRAD, Montpellier, France
[13] UM2, UMR AMAP, Bd de la Lironde, TA A-51/PS2, 34398 Montpellier Cedex 5, France
[14] Department of Biosciences, University of Helsinki, Helsinki, Finland
[15] UMR 7204 CNRS UPMC, Centre for Ecology and Conservation Sciences, Muséum National d'Histoire Naturelle, 55-61 rue Buffon, 75005 Paris, France
[16] Mathematical Ecology Research Group, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
[17] Institute of Landscape and Plant Ecology, University of Hohenheim, 70593 Stuttgart, Germany
[18] Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, Grenoble I, BP 53, 38041 Grenoble Cedex 9, France
[19] Ifremer, Rue de l'Ile d'Yeu, BP 21105, 44311 Nantes Cedex 3, France
[20] Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA

* Corresponding author : Olivier Gimenez, email address : olivier.gimenez@cefe.cnrs.fr

Abstract :

The desire to predict the consequences of global environmental change has been the driver towards more realistic models embracing the variability and uncertainties inherent in ecology. Statistical ecology has gelled over the past decade as a discipline that moves away from describing patterns towards modelling the ecological processes that generate these patterns. Following the fourth International Statistical Ecology Conference (1–4 July 2014) in Montpellier, France, we analyse current trends in statistical ecology. Important advances in the analysis of individual movement, and in the modelling of population dynamics and species distributions, are made possible by the increasing use of hierarchical and hidden process models. Exciting research perspectives include the development of methods to interpret citizen science data and of efficient, flexible computational algorithms for model fitting. Statistical ecology has come of age: it now provides a general and mathematically rigorous framework linking ecological theory and empirical data.

Keywords : citizen science ; hidden Markov model ; hierarchical model ; movement ecology ; software package ; spatially explicit capture–recapture ; species distribution modelling ; state–space model

1. Introduction

Variability is challenging ecology, from genes to individuals, species or ecosystems: quantifying and explaining biological variation is an ever-important goal. Variability arises from both ecological processes and sampling, requiring the modelling of uncertainty, the very nature of statistics (Davidian & Louis 2012; Spiegelhalter 2014).

Statistics has long permeated the field of ecology through the contributions of eminent scientists such as Fisher, Haldane and Leslie. However, we detect a recent rise in statistical awareness, manifested in various ways. First, research centres especially devoted to statistical ecology have been created in the USA (Statistical and Applied Mathematical Sciences Institute) and the UK (National Centre for Statistical Ecology). There are also institutes focussed on synthesis (e.g., the National Center for Ecological Analysis and Synthesis and the National Institute for Mathematical and Biological Synthesis, both in the USA). Second, new journals dedicated to methodological advances (not only statistical) have been created and are now having considerable impact (notably Molecular Ecology Resources and Methods in Ecology and Evolution). Third, there are more specialized conferences that provide the opportunity for statisticians to interact with ecologists for mutual benefit. The reasons for this recent rise of statistical ecology are manifold and include the societal demand for scientists to address pressing issues such as global change and the current biodiversity crisis, the need to analyse the massive datasets and the novel data types generated by new technologies, and the popularisation of methods through free statistical packages and the rise in computing power. We view the rise of statistical ecology as a sign that ecological and statistical modelling are coming together with the common goal of understanding complex processes in a formal inferential framework for better predictive capabilities. We acknowledge that not all ecologists agree that ecology lends itself to theorization and prediction (Cooper 2003), or that process-based methods necessarily have higher predictive ability than phenomenological models (Peters, 1991; Breiman 2001). However, past disappointments may simply be due to inappropriate and coarse modelling. If so, progress in both ecological theory and statistical ecology and a better integration of the two should enhance our understanding and our predictive ability of ecological phenomena. In the following, we highlight recent trends in statistical ecology and provide perspectives for the future development of this discipline (see also King 2014).

We analysed the contents of the abstracts of four International Statistical Ecology Conferences (ISECs) held biannually between 2008 and 2014 to provide a picture of recent trends in statistical ecology (Appendix 1). The quantitative results of this analysis show a temporal shift across the different ISECs, from studies focusing on sampling design issues towards predictive studies that aim to integrate the modelling of processes with the analysis of ecological patterns. These results are further synthesized below.

2. Questions being addressed

Assessing species distribution. Species distribution models (SDMs) are now common tools to investigate the main drivers of species range and to forecast potential impacts of environmental changes on biodiversity. Important innovations include the use of point processes to fit SDMs to presence-only data and the mathematical equivalence of MAXENT to generalized linear models (Renner & Warton 2013). SDMs are also being extended to several species to improve the model parameterization for rare species, and to enable the estimation of co-occurrence patterns. Last, the development of hierarchical occupancy models, with their ability to handle spatial dependence and imperfect detection, paves the way for better modelling of the underlying sources of uncertainty (MacKenzie et al. 2006).

Measuring biodiversity (including population dynamics). Biodiversity is multifaceted, involving aspects of species richness, functions, traits and phylogeny. Consequently, the choice of relevant diversity indices is challenging, especially when analysing aspects of functional or phylogenetic diversity and when evaluating the dissimilarities among locations (quadrats, sites, or regions). Moreover, the potential factors driving the dynamics of biodiversity (e.g., competition and environmental filters) need to be disentangled. In the ISECs, estimation of population size, a related topic, has been a major focus, notably through refinements of capture-recapture (CR) methods. There has been an increase in non-invasive methods that use natural identifying characteristics of animals (camera or acoustic traps, genetic markers), with treatment of misidentification error. In parallel, spatially-explicit models have been developed to fully exploit the spatial information in CR data (Royle et al. 2014).

Understanding animal movements. Movement ecology has shifted from phenomenological models of observable patterns to mechanistic models characterizing the underlying processes. In particular, the use of state-space models that account explicitly for the observation process has now become standard (Patterson et al. 2008), and hierarchical models have been developed to model individual movements as functions of behavioural states, past experiences, and environmental heterogeneity (McClintock et al. 2012). While earlier work relied on discrete-time correlated random walks, the use of continuous-time models and the integration of other types of data (e.g., species interactions, population dynamics) are increasing.

Interpreting citizen science data. Data from citizen science programs represent an opportunity to sample large regions and feed long-term monitoring studies. Difficulties arise with recent programs based on web- and smartphone-based technologies that are characterized by the free participation of many laypersons, loose sampling protocols and heterogeneities in the spatiotemporal distribution of observations. These potential sources of bias may be accounted for by the joint modelling of the ecological and observation processes through, e.g., hidden process models (Pagel et al. 2014).

3. Methods

Hidden process modelling. Ecologists have broadly adopted hierarchical, state-space and hidden Markov models to deal with how individuals and populations distribute in space and change over time (Clark 2007). This reflects a move away from modelling spatiotemporal patterns per se and towards modelling the ecological processes that generate those patterns. The timescale of interest might be short, such as for animal behaviour, or medium, such as for migration and demographic processes, or long, such as for changes in species ranges, composition and biodiversity, or for evolutionary processes. By modelling the underlying processes while accounting for observation error and model uncertainty, we seek to gain in predictive ability and hence in the effectiveness of management actions, whether we are managing a commercial fishery, conserving a threatened population, assessing the impact on biodiversity of habitat loss, predicting response of populations to disturbance, or evaluating the effects of climate change on communities.

Coexistence of frequentist and Bayesian frameworks. Bayesian methods are now widely used, largely because they can more easily accommodate realistic ecological models. However two notable trends are emerging: an increasing interest in critically evaluating the performance of Bayesian methods from a frequentist perspective (Little 2011); and the increasing practicality of frequentist tools for hierarchical models previously only amenable to Bayesian methods (e.g., Lele et al. 2007).

Dynamic models. Current research in population dynamics addresses the limits of statistical inference and predictions for nonlinear dynamics (e.g., Hartig & Dormann 2013). Beyond the population, dynamic statistical models are now applied at larger spatial and organizational scales to describe the dynamics of species ranges, communities and ecosystem processes (e.g., Clark et al. 2011). A common feature of these recent statistical models is that they describe how large-scale dynamics arise from underlying principles of demography and/or ecophysiology, aiming to base inference and prediction on processes rather than correlations.

Integrated modelling. Another trend is the popularization of integrated modelling – i.e., combining different data sets in a single, coherent analysis (Newman et al. 2014) – to address a wide variety of ecological questions. Current developments deal with the issues of goodness-of-fit testing, model selection, integration of recent developments in demography (e.g., integral projection models), and testing the assumption that data from different sources can be considered independent. From an ecological viewpoint, integrated modelling now scales from populations up to communities (Péron & Koons 2012).

4. Implementation

Computational algorithms. The development of efficient and flexible computational algorithms for complex models and big datasets ([integrated nested] Laplace approximations, Hamiltonian Monte Carlo and standard Markov chain Monte Carlo algorithms) requires tremendous research efforts, as does their implementation in software packages (e.g., R-

INLA[*], AD Model Builder[†], LaplacesDemon[‡], Stan[§], Nimble[**], OpenBUGS[††], JAGS[‡‡], PyMC[§§], MCMCglmm[***]). When a complete likelihood cannot be easily calculated, methods for estimation based only on simulations and summary statistics (Synthetic likelihood: Wood 2010; Approximate Bayesian Computation: Csilléry et al. 2010) are also receiving attention.

Software development and evaluation. There is a tension between devoting time to developing new methodology, and to enabling other researchers to implement it. Although it is easy to self-publish an R package or a GUI, a culture shift is needed toward more thorough testing and verification of published software. We welcome the initiative of ecological journals to publish software papers, which ensures that publicly-available software is peer-reviewed, and endows software development efforts with much-needed professional recognition.

5. Advice to statistical ecologists

Avoiding statistical machismo[†††]. Given methodological developments and increasing computing power, there is a great temptation to increase model complexity. In some cases this is helpful: previously restrictive assumptions about the observation process can be relaxed; previously intractable ecological mechanisms can be expressed as mathematical

---

[*] http://www.r-inla.org/
[†] http://admb-project.org/
[‡] http://www.bayesian-inference.com/software
[§] http://mc-stan.org/
[**] http://r-nimble.org/
[††] http://www.openbugs.net/w/FrontPage
[‡‡] http://mcmc-jags.sourceforge.net/
[§§] http://pymc-devs.github.io/pymc/
[***] http://cran.r-project.org/web/packages/MCMCglmm/index.html
[†††] http://dynamicecology.wordpress.com/2012/09/11/statistical-machismo/

models and incorporated in estimation. In other cases, however, increasing complication can lead to less robust inference or ecologically insignificant improvements, which nevertheless waste practitioners' time and direct their energies away from less glamorous topics such as improved data collection; there is also often an increased chance of mistakes in implementation. There is a clear need for an evaluation strategy of new, often complex statistical methods to determine the scope of beneficial application for ecology (Hodges 2010). *Beneficial* means that for a given ecological question and dataset, applying the new or modified method provides clearer results and avoids drawing flawed conclusions. Comprehensive model evaluation must include consideration of sample design, covariate selection, goodness-of-fit, and parameter redundancy diagnostics.

Going one step further. Many ecological applications are motivated by scientific support for conservation or management decisions. Statistical decision theory has much to offer, both directly in terms of helping rational decision-making, but also in optimizing future data-collection efforts.

6. Conclusions

The dialog between statisticians and ecologists has intensified over recent decades, and ISECs have contributed to this dialog. We encourage even more mixing between statisticians and ecologists, by exhorting the former to go to the field for a sound understanding of the data for relevant modelling (Gimenez et al. 2013) and the latter to embrace courses in mathematics that underpins the reliable application of statistical methods (Barraquand et al. 2014).

In summary, the statistical approaches developed for ecology are maturing toward a statistically rigorous, explanatory and possibly predictive framework for linking theory, data and applications. Exciting research directions are ahead of us that will hopefully help to address pressing issues in the context of global change.

Acknowledgments

We thank the scientific and local organising committees who largely contributed to the success of the ISECs. This is a contribution of the GDR 3645 'Statistical Ecology'.

References

Barraquand F et al. 2014 Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ* **2**, e285 http://dx.doi.org/10.7717/peerj.285.

Breiman L. 2001 Statistical Modeling: The Two Cultures. *Statistical Science* **16**, 199-231.

Clark JS. 2007 *Models for ecological data*. Princeton University Press.

Clark JS, Agarwal P, Bell DM, Flikkema PG, Gelfand A, Nguyen X, Ward E, Yang J. 2011 Inferential ecosystem models, from network data to prediction. *Ecological Applications* **21**, 1523–1536.

Cooper, GJ. 2003 *The Science of the Struggle for Existence: On the Foundations of Ecology*. Cambridge University Press.

Csilléry K, Blum M, Gaggiotti O, François O. 2010 Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, **25**, 410-418.

Davidian M, Louis TA. 2012 What is statistics? *Science*. **336**, 12.

Gimenez O et al. 2013 How can quantitative ecology be attractive to young scientists? Balancing computer/desk work with field work. *An. Cons.* **16**, 134-136.

Hartig F, Dormann, CF. 2013 Does model-free forecasting really outperform the true model? *Proceedings of the National Academy of Sciences* **110**, E3975–E3975.

Hodges J. 2010 Are exercises like this a good use of anybody's time? *Ecology* **91**, 3496-3500.

King R. 2014 Statistical Ecology. *Annual Review of Statistics and Its Application*s **1**, 401-426.

Lele SR, Dennis B, Lutscher F. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters* **10**, 551–563.

Little R. 2011 Calibrated Bayes, for statistics in general, and missing data in particular. *Statistical Science* **26**, 162-174.

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Oxford, UK.

McClintock BT, King R, Thomas L, Matthiopoulos J, McConnell BJ, Morales J. 2012 A General Discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs* **82**, 335-349.

Newman, KB, Buckland, ST, Morgan BJT, King R, Borchers DL, Cole DJ, Besbeas P, Gimenez O, Thomas L. 2014 *Modelling Population Dynamics – Model Formulation, Fitting and Assessment using State-Space Methods*. Springer.

Pagel J, Anderson BJ, O'Hara RB, Cramer W, Fox R, Jeltsch F, Roy DB, Thomas CD, Schurr FM. 2014 Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution* **5**, 751-760.

Patterson TA, Thomas L, Wilcow C, Ovaskainen O, Matthiopoulos J. 2008 State–space models of individual animal movement. *Trends in Ecology and Evolution* **23**, 87-94.

Péron G, Koons DN. 2012 Integrated modeling of communities: parasitism, competition, and demographic synchrony in sympatric ducks. *Ecology* **93**, 2456-2464.

Peters H. 1991 *A Critique for Ecology*. Cambridge University Press.

Renner IW, Warton DI. 2013 Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* **69**, 274–281.

Royle JA, Chandler RB, Sollmann R, Gardner B. 2014 *Spatial Capture-Recapture*. Academic Press.

Spiegelhalter D. 2014 The future lies in uncertainty. Science **345**, 264-265.

Wood SN. 2010 Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104.

Appendix 1.

The R code to perform the analyses described below is provided in the script ISEC-analysis.R as well as the data which are in the ISECData.rda and ISECGraph.rda files.

We performed a text mining analysis and analyzed the lists of the 25 most common words in each ISEC abstract volume (Figure A1). The word *dynam* appeared in the 2014 list, *chang* and *process* in 2012 and 2014 and *time* in 2010, 2012 and 2014 suggesting a growing interest in integrating mechanisms to explain ecological patterns in time. *Predict* reached the top 25 list in 2014 and *chang* in 2012 and 2014, which is in line with a rising concern for global change and with related efforts to predict ecological dynamics under environmental change. The words *distance*, *survey*, *density* and *design* disappeared from the list after ISEC 2008. This reflects the main focus of the first conference on sampling design issues while ISECs 2010, 2012 and 2014 reflected a wider range of interests (e.g., *movement ecolog*y appeared in 2012).

Now focusing on the 16 words that are common to the abstracts of all four ISECs (Figure A1), statistical ecology is without surprise about fitting *model*s to *data* to *estim*ate *parameter*s of ecological relevance. This is achieved by developing *method*s to determine the main *effect*s explaining the *differ*ent patterns in the *distribution*s *of individu*als, *popul*ations and *speci*es. The quantity of interest is predominantly the *abund*ance of *anim*als, considered at different *spatial* scales with a particular attention to the issue of imperfect *detect*ion and adequate *sampl*ing scheme.

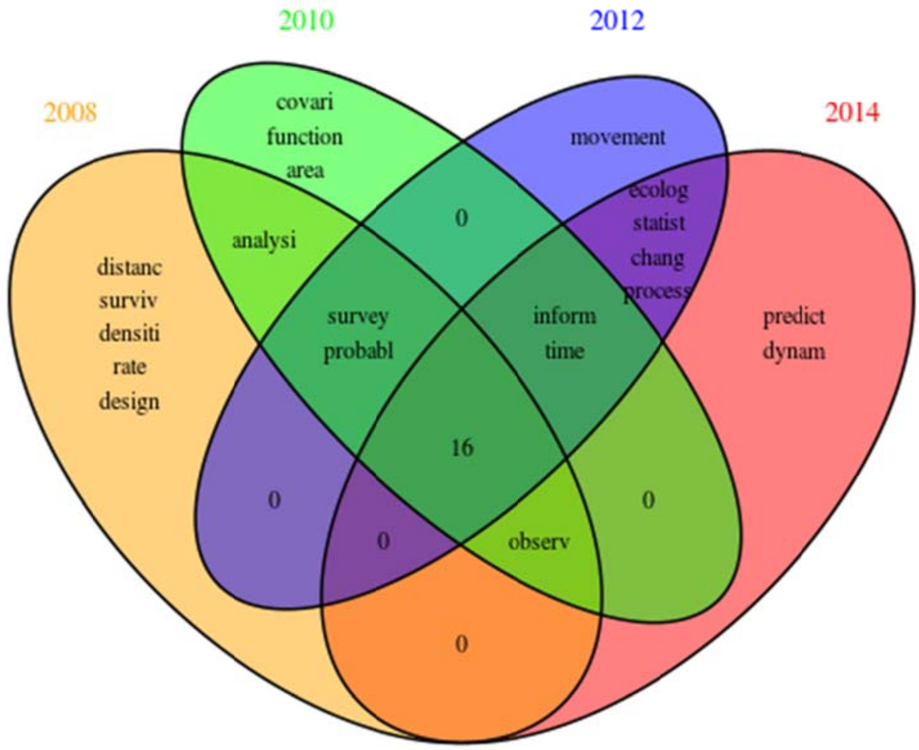

Figure A1. The 25 most frequent words in the abstracts of each ISEC using a Venn diagram. We found 16 words (or at least their root) common to all ISECs: abund, anim, data, detect, differ, distribution, effect, estim, individu, method, model, paramet, popul, sampl, spatial, speci. See Table A1 for the full list of 25 words per year.

Table A1. The 25 most frequent words in the ISEC abstracts (sorted by the number of occurrences). The terms common to all ISEC editions are in bold.

2008	2010	2012	2014
model	**model**	**model**	**model**
estim	**estim**	**data**	**data**
data	**data**	**estim**	**speci**
popul	**speci**	**popul**	**estim**
speci	**popul**	**speci**	**popul**
sampl	**method**	**spatial**	**method**
method	**sampl**	**method**	**distribut**
abund	**abund**	**individu**	**spatial**
survey	**differ**	**sampl**	**sampl**
spatial	**detect**	**paramet**	ecolog
detect	**spatial**	time	**differ**
probabl	survey	**distribut**	**abund**
paramet	observ	survey	**individu**
anim	inform	**abund**	time
individu	**paramet**	**effect**	**detect**
analysi	**anim**	**differ**	**paramet**
distanc	**effect**	**detect**	chang
distribut	**individu**	inform	observ
surviv	analysi	ecolog	predict
observ	probabl	probabl	process

densiti	**distribut**	**anim**	dynam
differ	time	movement	**effect**
rate	covari	statist	**anim**
design	function	chang	statist
effect	area	process	inform

In addition, we performed a multivariate analysis (non-symmetric correspondence analysis) of the 50 most common words found in the abstracts of the four conferences (Figure A2). Figure A2a shows major trends of semantic variation among the abstracts. The first major trend (abscissa) contrasts the studies focusing on sampling design issues (e.g., high positive scores of *transect*, *design*, *sample*, *detec*) and the studies focusing on characterizing processes and resulting patterns (negative scores). The second axis contrasts the field of population studies based on capture-recapture approaches vs. approaches investigating community dynamics, habitat modelling and species distributions (e.g., high positive scores of *communiti*, *speci*, *habitat*, *distribut*). These two axes represent 10.04% of the overall variation among abstracts. Figure A2b shows the 90% convex hulls of each ISEC conference based on the scores of their abstracts. We found a significant variation (randomization test, $p < 0.001$) with an overall trajectory toward lower scores on the first axis (more process-oriented works) and toward more emphasis on community dynamics, habitat modelling and species distributions on the second axis.

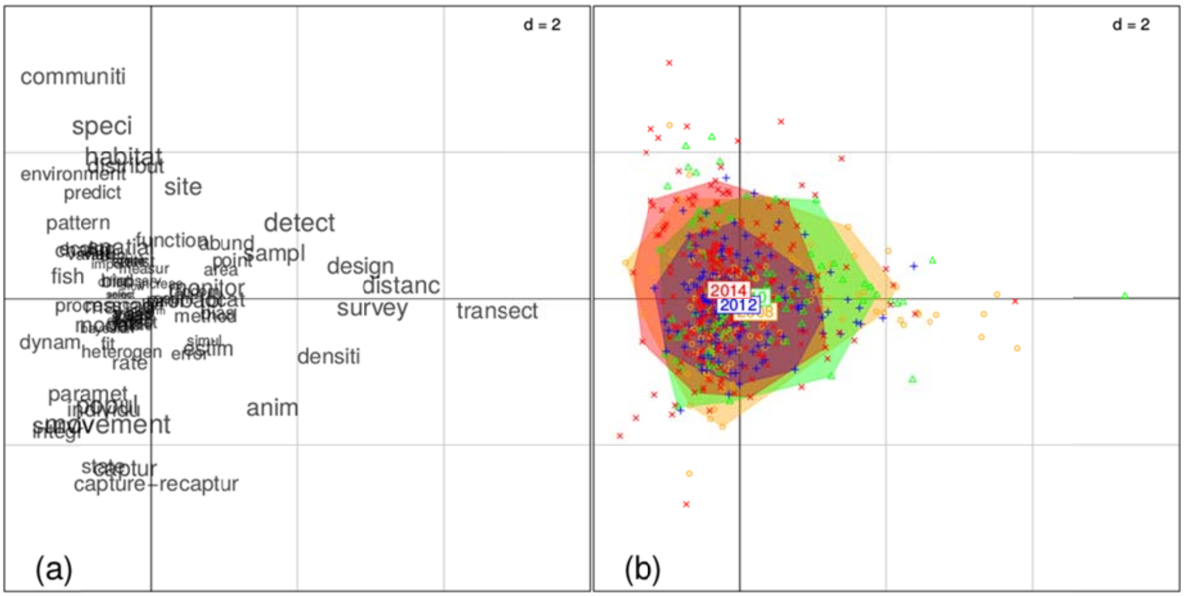

Figure A2. Semantic variations found among the abstracts of the four ISECs identified by a non-symmetric correspondence analysis of the word-by-abstract table. The resulting first factorial map shown here illustrates major trends of semantic variation among the abstracts. (a) The higher scoring words are those most contributing to semantic variation in the factorial map. The size of labels is proportional to the contribution of each word to the first two axes. (b) Distinguishing the 90% convex hulls of abstracts for the four ISECs underlines thematic variations through time.

We also addressed the structure of the research fellow communities participating in ISECs. The co-authorship network was built and analyzed using a stochastic block model (Figure A3) to identify groups of authors. Based on the Integrated Completed Likelihood criterion, 15 groups of authors were detected. The isolated contributors are grouped in cluster 9, which is one of the most important in numbers. The other groups may be named communities since they were characterized by high within probability connectivity. Cluster 1 (in dark blue) was found to be a central hub of in the ISEC community. A meta-community, formed by clusters 1, 7, 10, 12 and 15 was identified and mainly contained the initial ISEC contributors. The 9 remaining clusters exhibited high level of within-connections but poor between-connections. This may indicate a need for more communications and exchanges between communities and disciplines within statistical ecology.

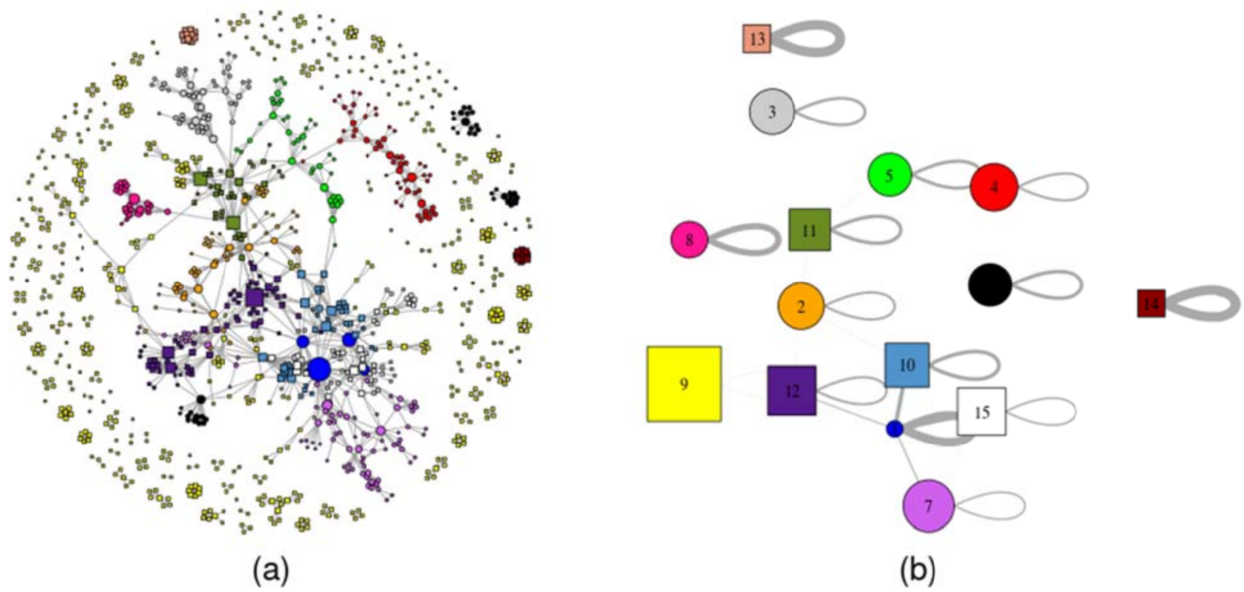

Figure A3. Analysis of the ISEC coauthorship network. The first graph (a) presents the network of copublications, based on the talks given during the four ISEC editions. Each vertex represents a contributor. The colors indicate the clusters detected by the stochastic block model analysis. The size of a vertex is proportional (in log scale) to the number of coauthors. The second figure (b) sums up the properties of each cluster. The width of the edges depends on the probabilities of connection between or within clusters, while the size of the vertices relies on the assignment marginal probabilities.